Entity Recommendation for Everyday Digital Tasks

GIULIO JACUCCI, Finnish Centre for Artificial Intelligence, Department of Computer Science, University of Helsinki

PEDRAM DAEE, Department of Computer Science, Aalto University

TUNG VUONG, Department of Computer Science, University of Helsinki

SALVATORE ANDOLINA, University of Palermo

KHALIL KLOUCHE, Department of Computer Science, University of Helsinki

MATS SJÖBERG, CSC—IT Center for Science

TUUKKA RUOTSALO, Department of Computer Science,University of Helsinki and University of Copenhagen

SAMUEL KASKI, Finnish Centre for Artificial Intelligence, Department of Computer Science, Aalto University and University of Manchester

Recommender systems can support everyday digital tasks by retrieving and recommending useful information contextually. This is becoming increasingly relevant in services and operating systems. Previous research often focuses on specific recommendation tasks with data captured from interactions with an individual application. The quality of recommendations is also often evaluated addressing only computational measures of accuracy, without investigating the usefulness of recommendations in realistic tasks. The aim of this work is to synthesize the research in this area through a novel approach by (1) demonstrating comprehensive digital activity monitoring, (2) introducing entity-based computing and interaction, and (3) investigating the previously overlooked usefulness of entity recommendations and their actual impact on user behavior in real tasks. The methodology exploits context from screen frames recorded every 2 seconds to recommend information entities related to the current task. We embodied this methodology in an interactive system and investigated the relevance and influence of the recommended entities in a study with participants resuming their real-world tasks after a 14-day monitoring phase. Results show that the recommendations allowed participants to find more relevant entities than in a control without the system. In addition, the recommended entities were

---

also used in the actual tasks. In the discussion, we reflect on a research agenda for entity recommendation in context, revisiting comprehensive monitoring to include the physical world, considering entities as actionable recommendations, capturing drifting intent and routines, and considering explainability and transparency of recommendations, ethics, and ownership of data.

CCS Concepts: • **Information systems → Users and interactive retrieval**; **Recommender systems**; • **Human-centered computing → Human computer interaction (HCI)**;

Additional Key Words and Phrases: Proactive search, user intent modeling

1 INTRODUCTION

Recommender systems are increasingly becoming an integral part of a person's everyday digital life by continuously monitoring and updating user models to recommend information that may be useful without requiring explicit user actions [20, 36, 91]. Some recommender systems have taken advantage of app-specific entities and data structures (such as movie recommenders), while other recommenders operate across apps by using generic textual input (like predictive keyboards) but they tend to recommend shallow textual structures. This work is targeting both cross-app input and cross-app rich entities. Research in this area has produced over the past few decades examples of implementation of similar systems, but remains unclear to which extent these are effective in supporting people in realistic situations. Moreover, research should investigate their ability to infer tasks and address their complexity to recommend the right information at the right time.

We present an approach for entity recommendation in every digital task, based on learning from continuous image captures of the screen. Users are recommended information entities such as people, applications, documents, or topics based on their current task. We evaluate the approach in a study demonstrating the relevance and usefulness of recommendations.

Our aim is to define *entity recommendation* in everyday digital tasks informed by previous contributions, providing an approach and open challenges to consolidate and revise this area with a research agenda. We propose for the approach the following core principles: comprehensive digital activity monitoring and entity-oriented computing and interaction. To validate this approach, we present the implementation of a system based on these principles and investigate entity recommendation based on real-world *everyday digital task* data. More importantly, we conduct an evaluation to measure the performance beyond relevance by quantifying the usefulness and *influence* of the recommended entities on user behavior. We use the term *entity recommendation* as the central task of a recommender system is providing personalized recommendations based on a user's personal preferences and historical behavior [57, 96]. With the term "*everyday*" we denote that the research is interested to move toward realistic studies. The hypothesis is that monitoring the information entities displayed on the screen reveals information relevant to tasks and can be useful in predicting user *entities that are interesting and useful* to the task the user is performing.

Evaluation approaches in this area have focused on the algorithmic effectiveness of information retrieval and recommendation methods [41], overlooking the influence of the methods on everyday digital tasks [45] and with less appreciation for **human−computer interaction** (**HCI**) aspects such as realistic task evaluation [44, 45, 65].

To operationalize such research, we need a method that is general enough to perform user modeling on a variety of everyday digital tasks. Instead of more traditional modeling approaches (e.g., [55, 74]) based on the logs of interaction events (e.g., the opening of a file, the browsing of a web page, or the execution of an application), which have limited access to semantic information apart from a file's metadata, we propose an entity-centric strategy [4, 53]. The recent use of entity approaches in search stems from semantic computing and from efforts to create Internet-wide knowledge bases and graphs to support semantic search. Entities are semantic data objects with a collection of properties corresponding to real-world objects they represent, and they can be linked to other entities [15, 71]. This allows for easier extraction of meaning from the user's activity and can thus be used to provide users with less ambiguous, actionable information.

The solution we employ uses digital activity monitoring capable of extracting the texts shown on people's computer screens. We use this rich source of context to extract relevant entities such as documents, applications, people, and topics. To compute a model of relevant information to the user, we implement an online machine learning method to learn about users' interest across all entities by proactively monitoring users' digital activity inside the screen and/or interactively monitoring it through explicit feedback. The method exploits the information in thousands of screen frames collected from the user's screen to detect the context and to recommend novel entities related to the current user task. The method is implemented in a system named EntityBot.

To understand the effectiveness and usefulness of recommended entities in everyday digital tasks, we conducted a study using participants' real-life data and tasks. EntityBot was installed on users' laptops for 2 weeks to conduct unsupervised learning of the entities' representation and their relationship during actual work tasks. After this, users participated in an experimental session resuming previous work tasks. EntityBot is set up as a separate screen where recommended entities are visualized during users' work. Users may open recommended documents, applications, and contacts or give feedback by selecting an entity that, in turn, performs an update on the model and recommends new entities. Figure 1 illustrates this system setup. In the experiments, we use a control condition to investigate the benefit of the EntityBot system and verify, without visualization and user feedback, to which extent the recommended entities are relevant and not already present. The user study demonstrates the viability of the approach by showing the benefit provided by EntityBot in terms of discovering novel and relevant entities. More importantly, the results indicate that the recommended entities were found to be useful and influenced the user tasks, leading to improved user experience in completing the task. Our contributions include the following:

— *an approach* informed by previous work for Entity Recommendation for Everyday Digital Tasks characterized by comprehensive digital activity monitoring and entity-oriented computing and interaction;
— *a study* in realistic everyday digital tasks, answering the overlooked question of whether the recommended entities are useful and impact user behavior.

In the discussion, we propose elements of a research agenda in terms of open challenges uncovered by this approach and study. In the next section, we provide an overview of previous work and we introduce the two principles at the core of our approach.

2 RELATED WORK AND BACKGROUND PRINCIPLES

Recommendation systems are increasingly affecting users' everyday lives due to their ability to help them find relevant information in the fast-expanding digital universe. They monitor contextual signals to create and update a user profile, which is then used to provide users with information tailored to their context (e.g., [39, 63]). Contextual signals may include page visits [104], click-through data [19], or pre-search context [55], or a combination of behavioral signals [96].
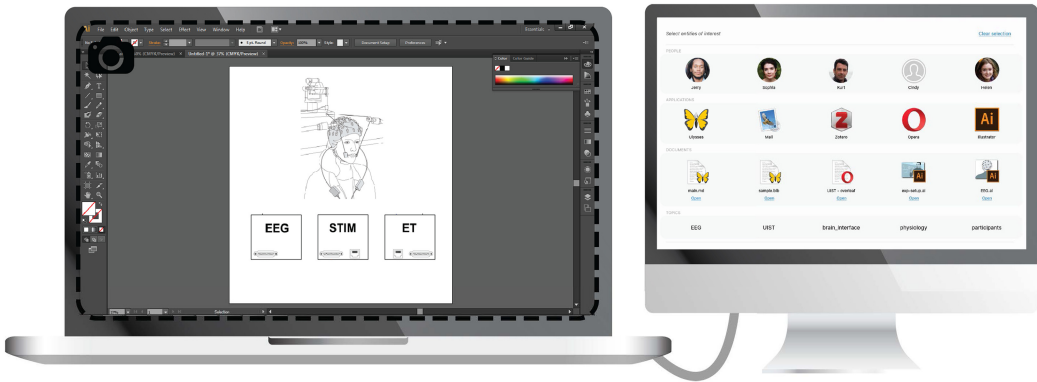
Fig. 1. An example where a user is drawing a picture for a scientific paper on the laptop (left). The EntityBot continuously extracts information from the user's screen, in this case the picture (e.g., containing entities EEG, STIM, ET, Adobe Illustrator, ...), and discovers the user's evolving interest. It recommends entities in real-time and displays them on a separate screen on the right (e.g., the related manuscript, other related vector graphic files, or co-authors to contact and seek feedback on the figure) that could help with the task.

The evaluation of whether the provided information is actually useful and can influence users' behavior in real-world tasks has remained less studied [44]. In particular, a large body of research has targeted algorithmic approaches that aim to retrieve or recommend relevant items to individual users based on their context away from realistic task evaluation. [17, 19, 28, 39, 56, 62, 104].

Recommender systems do not require users to perform a specific action but leverage their context and past interactions to anticipate their needs [55, 94] and provide them with information that is likely to be relevant in an automatic way. Our work focuses on such systems, more specifically on recommender systems that operate *just-in-time* [77], meaning that aim at delivering information at that moment when the user needs it, but without the user explicitly requesting it.

Such systems extract the context from the user's data to form a representation of the current user's possible interests and predict contextually relevant information. Previous works have considered mainly specific tasks or specific user data for context (e.g., only email or browser data), or focused more on conventional recommendation tasks predicting the value of immediate search results after a query issued by a user by using specific behavioral signals. Entity recommendation, based on analyzing the task context, can be seen as a mixed-initiative system [2]. This feature is shared by some of the systems we review here and by recent search bots [5, 100] and conversational search agents, raising the question of the impact of such recommendations that are not initiated by the user on the user's subsequent behavior. Table 1 summarizes several attempts to create recommendation systems or information retrieval based on context for a variety of tasks varying from conventional web search [29, 55, 59, 94] to document writing and meeting preparation [28, 56, 109]. They mainly differ in their ability to capture the context, in the type of information they provide for the users, and in the extent to which they allow users to explicitly affect their models through interactive feedback. This prior work mainly focused on how relevant document recommendations could be provided within tasks that were either simulated [29, 56], or otherwise limited in scope [77, 109], thus failing to provide a clear picture of the more interesting problem of understanding how recommendations *influence* real-world tasks. Evaluations are typically conducted using simulation studies on pre-existing data sets with the goal of measuring whether a novel algorithm improves effectiveness or efficiency over existing ones. Unfortunately, while such approaches have led to major algorithmic improvements, they often provide only limited insights on the influence of the methods on everyday digital tasks [45].

Over the years, works have observed that advances in algorithmic effectiveness may sometimes provide only little practical added value and have promoted evaluations based on usefulness and more realistic tasks [17, 61, 74, 76]. Currently, the perception, use, and experience of recommendations are important topics [87] as well as going beyond accuracy in the evaluation, for example, in the case of recommender systems in the health domain [97]. We introduce a method for entity recommendation that would be able to support people in a wide variety of everyday digital tasks; in turn, this would permit studying the influence and usefulness of the recommendations on everyday digital tasks with more rigor.

In this section, we first review Section 2.1 insights originating from earlier related systems with a special focus on those supporting primary tasks. After this analysis (summarized in Table 1) and informed by how previous work addressed how to monitor users and what types of items should be recommended, we introduce two principles on which our approach is based. We introduce comprehensive digital activity monitoring Section 2.2, briefly introducing how digital activity monitoring has evolved and proposing to model users more holistically by capturing everything that is displayed on a computer screen. We finally propose entity-based interaction Section 2.3 that generalizes how previous systems considered different types of items to be recommended to enable entities to be actionable and allow interaction for the user.

2.1 Related Recommendation Systems and Studies

Many existing studies in information-seeking research promote a view of the search activity as something belonging to a wider high-level task [46]. As a result, a large body of research has started investigating novel ways to model the context of users and use such models to infer users' interests and search intents. This has led, for example, to the emergence of paradigms such as search personalization, where the user's task or search context is modeled [12, 93, 96] using previous queries and page visits [104], or click-through data [19], or by modeling the immediate pre-search context [55], or the task that motivates the information need [51, 66].

Proactive search [29] (or *anticipatory search* [60]) is a natural extension of search personalization [96] with the explicit query step removed. Here, the user's context is continuously being monitored and the user model is being updated to anticipate the upcoming information need. An early example of this paradigm is the *Remembrance Agent* [76] which monitors a user's personal data (e.g., emails and text documents) and continuously displays a list of documents related to what the user is doing now. Here, document relevance is simply estimated based on the frequency of common words in the currently active text, and there is no long-term modeling. *Letizia* [59] is a similar early example that provides automatic recommendations during web browsing. A more modern application can be found in current smartphones (e.g., in the form of Google Assistant), which tries to model not only short-term search intents but also long-term interests and habits based on several months of collected data [36]. Here, user-specific context classes (e.g., tasks, interests, or habits) are identified from their search history. Another approach is to extract patterns related to the time of the day [92](e.g., that a certain task is usually done in the morning) and use these to anticipate resources the user will need at that time.

Researchers have also studied recommendations offered to users while they are performing specific tasks. For example, in [17], contextual text and image queries are performed based on the text written by the user in a word processing application, and, in [62], reference recommendations are shown in a similar scenario. Other examples include showing personal documents related to the current email being read or written by the user [28], or during authoring of a PowerPoint presentation [61]. In [56], a more generic approach to capturing the search intent from the primary task context is proposed. However, the experimental part mainly studies the writing task. Other

Table 1. Comparison of Related Retrieval and Recommendation Systems

	Recommended Entities	Context Extraction	Interactive Feedback	Just-in-time	Activity supported	Evaluation of usefulness
Letizia [59]	Web documents	Browsing behavior	No	Yes	Web search	No
Remembrance Agent [76]	Documents	Emails and written notes	No	Yes	Writing a newspaper-style article [77]	Subjective scores
Query-free News Search [40]	Web documents	Caption data	No	Yes	News search	No
Elliot and Jose [29]	Documents	Browser	On keywords	Yes	Multi-session search	No
Koskela et al. [56]	Web documents	Text from text editor	On keywords	Yes	Document writing	Number of selected recommendations
Watson [17]	Web documents and images	Text from text editor and browser	No	Yes	Document writing and web search	Potential usefulness of first recommendation (subjective scores)
SidePoint [61]	Text snippets and images	Text from presentation authoring software	No	Yes	Presentation writing	Qualitative feedback on the system (lab study)
CAPERS [109]	Emails	Calendar	No	Yes	Meeting preparation in an enterprise	Number of clicking and hovering actions (field experiment)
Reply With [98]	Attachable items in emails	Email conversation	No	No	Assist users with composing emails	No
IQ [28]	Documents, people, topics	Emails	On overall quality of results	Yes	Reading or composing emails	No
CAAD [74]	Documents, applications, email addresses	Apps that make native OS calls	Update clusters of context structures	Yes	Computer-based information work	Perceived usefulness from questionnaires and interviews (field experiment)
Vuong et al. [102]	Documents and keywords	Any app	No	No	Task detection and proactive retrieval within a closed set of tasks	No
EntityBot (our system)	Documents, people, applications, and topics	Any app	One-click explicit feedback on entities (Optional)	Yes	Everyday digital tasks	Influence of recommendation on tasks plus interviews (field data collection followed by a lab phase)

examples include recommending personal documents related to the current email being read or written by the user.

In contrast to these earlier approaches, we propose a more general approach based on screen recording, which is not restricted to a specific task or application. For example, unlike the

Remembrance Agent [76], which watches whatever is typed or read in a word processor to recommend documents, our approach uses both system logs and textual context from any application to recommend entities. Similar to [29] and [56] that support specific tasks and allow users to affect the recommendation of documents through interactive feedback, we enable user feedback on recommended entities, but we do so to support any digital task. Compared to other approaches that monitor the context more comprehensively through logs of interaction events [74, 83], we couple system logs with **screen monitoring (SM)** to also capture textual context that might otherwise be missed. In addition to dramatically expanding the scope of its usage, our approach also facilitates collecting data across most of the computer activities and thus enables a more complete modeling of the entire HCI.

Contrary to prior work, we demonstrate a system that is used to study the relevance and usefulness of entity recommendations in supporting a variety of heterogeneous everyday digital tasks, with a particular focus on understanding the actual influence that recommended entities have on tasks.

Prior studies have pointed out the need to more rigorously evaluate the usefulness of recommended information [17, 77], as relevance does not necessarily correlate with usefulness. In [17], to determine whether the sources returned by Watson were useful in the context of a particular task, the authors asked six participants to send them a copy of the last paper they wrote, fed the paper to Watson, and returned the first list of recommendations retrieved by the system to the participants for subjective assessment of usefulness. Similarly, researchers have conducted qualitative evaluations of usefulness: in Remembrance Agent, through subjective scores given by participants [77], and in SidePoint, through qualitative feedback on the system [61]. Other more quantitative approaches have considered interaction events such as the number of clicks on recommended items as proxies of usefulness [56, 109]. Although these studies reveal potential benefits and challenges in more realistic settings, they typically rely on simulated work tasks, thus failing to provide a realistic picture of the actual influence of proactive recommendations in real-world situations.

The study that comes closest to our vision is presented in [74] with the evaluation of the CAAD system. CAAD automatically generates task representations (as context structures) from the logs of low-level interaction events. Contrary to most systems, CAAD captures context from most applications (i.e., those that make a native OS call). The researchers conducted an evaluation of the system in the participants' working environment over the course of three days. In addition, the researchers used questionnaires of perceived usefulness and interviews to gather qualitative feedback on the system. One limitation of the study was that some participants used applications that did not make any native OS calls and were therefore invisible to CAAD. Although the study used real-world tasks, it shared most of the limitations of other studies, including the missed opportunity to measure the influence of recommendations on tasks and the overall focus on the perceived accuracy of suggestions. All these studies also missed the opportunity to recommend semantically rich items. This is a factor that affects the overall actionability of recommendations and, consequently, their perceived usefulness.

In contrast, we study how entity recommendations can be useful in everyday digital tasks by measuring their actual influence on people's tasks both quantitatively and qualitatively through an in-the-wild data collection followed by a lab phase where people resumed their real-world tasks. Contrary to other studies, we also employ a system that can provide actionable entity recommendations through an entity-centric approach and the use of interactive feedback. As typical everyday digital tasks require users to switch across several applications, our system is not confined to specific applications. By employing screen recording, our system is able to model task-related context comprehensively, avoiding compatibility issues experienced in prior research [74].

2.2 Comprehensive Digital Activity Monitoring

Sensing, collecting, and storing data on people's activities have been the focus of a large body of research [9, 22, 31–33, 68]. Both physical and digital activities have been monitored to study challenges related to proactive delivery of information, such as notifications [68], interruptibility [22], and interaction [9]. More recently, research on digital activity monitoring and the prediction of user behavior has typically focused on large-scale tracking (e.g., based on what people are sharing on social media) [107, 110]. This mass-monitoring approach has important drawbacks, including loss of privacy and lack of trust in the system [21]. In contrast, other lines of research have started to investigate technologies that empower individuals to monitor their personal data, thus placing the collection and analysis of data into the hands of the individuals themselves [24, 89]. The work described in this article completes the picture by providing the modeling and user interaction for facilitating this vision.

Most of the approaches mentioned so far have focused on monitoring specific applications or other limited information sources. However, recent work [101, 102] has explored using screen monitoring SM, which captures the entire visual content of the computer screen for task recognition. Already early research found that techniques, such as latent semantic analysis [25], with a simple bag-of-words data representation, can be effective in detecting users' tasks and were helpful for context-aware recommendations. This tracking "inside the screen" paradigm has the benefit of being more general, as any visually communicated information can potentially be captured and utilized for building a richer task model. An approach similar in spirit, but more limited, is described in [38], where seen text snippets are associated with files opened at the same time.

Our work builds on the same idea of long-term SM, but we utilize an online machine learning approach that learns about user search intents and information interest in real-time based on SM and, if available, explicit feedback. Compared to previous user-modeling solutions, such as [36], our approach through the modeling algorithm is fully personalized (trained independently for each user), does not need to model the user's actions (e.g., searching on the web or clicking an item in an application), and is computationally inexpensive.

2.3 Entity-Based Interactions in Information Retrieval

In IR, entities are references to real-world objects or concepts (e.g., persons, places, movies, topics, or products). In web searches, most emitted queries pivot around a specific entity [73]. This can be seen as a generalization of the keyword concept used by previous systems, but with clearer affordances in the computer interface (e.g., clicking on a people entity could initiate an action to contact that person).

Entity-based queries in existing search engines result not only in a relevant entry (e.g., from Wikipedia) but in a knowledge graph with relevant information and related entities, providing quick links to further the exploration. Miliaraki et al. [67] studied the behavior of users of Yahoo Spark, a system that recommends related entities alongside Yahoo Search results; the users take advantage of the system to engage in exploratory entity search by discovering information through successive clicks on recommended entities. Recent research work explores novel interaction techniques through direct manipulation of displayed entities [3, 52, 54, 80]. These systems display results as interactive objects that can be used as a query or part of a query in a new search. The benefits of these entity-based approaches come mainly from their reliance on users having to interact with the information they recognize (e.g., recommended or retrieved entities) over users having to recall information (e.g., when typing a query). When providing both options, entity-based systems have been shown to substantially reduce the need for typing, which makes them especially useful for touch devices [52].

In this article, we propose using entities as the basic unit for representing the user's evolving interests. Unlike other approaches for search intent visualization and manipulation mainly based on keywords [82], we adopt an entity-centric approach. Entities allow the computation of a sophisticated user model and, as discussed above, more actionable recommendations for the user. Entities represent familiar real-world objects and concepts and could, as interactive objects, provide the right affordances to interactively feedback on representations of their interest.

3 ENTITYBOT: IMPLEMENTING OUR APPROACH IN A SYSTEM

To study entity recommendation in everyday digital tasks, we implement the EntityBot system. Below we describe the design goals that led to its main features.

3.1 Design Goals

—*Entity-centric approach.* Everyday digital tasks may require users to perform several actions, such as opening a document, contacting a person, or searching for resources associated with given topics [86]. Therefore, the items recommended by the EntityBot system were not limited to documents but included various kinds of information entities—such as people, applications, documents, and topics—that can be used to represent the task but also to perform specific actions related to the tasks. Moreover, to make the recommended entities more actionable, the **user interface (UI)** included hyperlinks that permit direct access to the entities.

—*Task-related context.* Everyday digital tasks often require access to information that is distributed across several applications and services [13]. To study entity recommendation in everyday digital tasks, we designed EntityBot with the capability to access task-related information across application boundaries. Unlike prior work with access to partial data, which are only obtainable through predefined applications or services [42], EntityBot was designed to function without the need for special application-dependent customization.

—*Interactive feedback and learning.* Complex search tasks are highly interactive [48, 64], but current recommendation systems typically do not include support for interactivity [17, 59, 61, 76, 109]. Allowing users to interactively affect the provided recommendations [84] could improve the overall quality of recommendations [79, 84], but people may not always be willing to provide this kind of input [34, 103] if the mechanism is too orthogonal to their current task. Therefore, we designed the EntityBot to provide users with the possibility to affect the recommendation through interactive relevance feedback. While EntityBot can properly work with implicit relevance feedback alone, its UI design includes easy mechanisms to provide explicit relevance feedback on recommended entities when needed.

EntityBot recommends entities beyond application boundaries. From this point of view, it is similar to Siri suggestions [7], which recommend people and applications mostly based on the frequency of usage and routines. Most previous works (see Table 1) recommend information within an application (e.g., email, web browsing). In Table 1 the closest is CAAD [74], which extracts file names, applications, and email addresses through **operating system (OS)** calls. EntityBot, in addition, operates **entity extraction (EE)** in processing on all text available in a given active window. More importantly, most systems including Siri suggestions do not provide possibilities for explicit user feedback. The practical implementation of the system consists of three main components. A digital activity monitoring module extracts entities across application boundaries. An online machine learning method learns about user interests in real-time based on SM data and, if available, explicit feedback. A UI presents the list of recommended entities. In this section, we describe these three main components.

3.2 Digital Activity Monitoring Module

Collecting real-life data and tasks is a prerequisite for making relevant recommendations. We aimed for a methodology that is able to unobtrusively collect all possible digital activities on a user's computer. Capturing the text read by users on the computer screen offers great potential to capture important information, including the visually communicated input and output (i.e., visual content that is generated and presented to a user on a computer screen). Traditional approaches to capturing text are based on Accessibility **Application programming interface (APIs)** [78, 99]. These APIs can identify the visible text in the UI [106]. Although reliability may differ across platforms, this is a viable approach that should be tested. We have opted for screen capture as it offers further opportunities to process visual content and gives the research full control over the processing of the data. Accessibility and other system-level APIs are generally insufficient to process content and analyze the interaction, as previous work has required additional data or sensors [10]. Whereas SM is capable of capturing all possible information across application boundaries. It is also able to capture users' ongoing task-related context that can be used to infer their potential intents and to recommend relevant information supporting users' tasks. Apart from audio, it captures all user inputs and presentation of content that occurs on the computer screen, providing a rich documentation of the user's interaction. Due to comprehensive logging mechanisms and support for a variety of applications, we chose to use SM in the experiments. The digital activity monitoring system is composed of four components: SM, **optical character recognition (OCR)** system, EE system, and OS logger.

— *SM* captures screenshots of active windows at 2-second intervals or, alternatively, captures the text read by the users on the screen. SM is developed into two versions: a Mac OS version and an MS Windows version. We utilized the Core Graphics framework to implement the Mac OS version and the Desktop App UI to implement the MS Windows version. Both perform an identical function that saves the screenshots of active windows as images.

— *OCR system* detects and extracts text from the screenshots. We utilized Tesseract 4.0,[1] which is a commonly used and very accurate OCR implementation.

— *EE system* detects and extracts available entities from the OCR-processed screenshots. We utilized the IBM Bluemix Natural Language Understanding API[2] to extract two types of entities that included people's names and keywords.

— *OS logger* collects information associated with the screenshots recorded, such as names of active applications, titles of active windows, available URLs of web pages, or available file paths of documents that are stored on the computer. In addition, the OS logger also collects timestamps of when the screenshots are captured.

All OCR-processed screenshots, extracted entities, and collected OS information were encrypted and stored as log files on the laptops for further access in the later phase. The digital activity monitoring system had a pause button that allowed participants to temporarily pause the monitoring when they did not want to share some of their private activities.

3.3 Learning and Recommendation

The learning method receives the logged data from the digital activity monitoring (Section 3.2) and prepares recommendations for display on the UI (Section 3.4). The user then may provide explicit feedback on the recommended entities, which helps the method to update the recommendations. In our solution, the relationship between entities is established mainly based on their co-occurrence

---

[1]https://github.com/tesseract-ocr/tesseract/wiki.
[2]https://www.ibm.com/watson/services/natural-language-understanding/.

and partially through temporal closeness. In the following, we first discuss the modeling challenges in learning the user interests and recommendation of entities and then explain the details of the implemented approach.

3.3.1 Overview. We define the context as a vector that represents all digital activities on a user's computer at each time step. The set of digital activities in our setting includes OCR-processed screenshots, recorded OS information, and extracted entities from EE systems. The user's continuous activity creates a large number of context vectors. Given this definition of context, the user information interest is modeled by a function that maps all entities (i.e., elements of the context vector) and all context vectors to continuous relevance values. Learning this function is challenging because the number of entities and potential context vectors is very large (here tens of thousands), while the learning signal from the user is rarely explicit. Any method attempting to learn this mapping just based on explicit feedback alone would have difficulty due to the dimension of the problem and the small sample size [26].

To amend the limited feedback, we assumed that the recent contexts (i.e., the user's latest digital activities) are relevant to the user's current intention. Even with this additional learning signal, the learning method still has to overcome the inherent noise in the logged context vectors, whether it is the noise in the OCR system or false entity detection by the EE system. Furthermore, the model needs to solve a high-dimensional learning task in real-time for interactive use. To overcome these challenges, we propose an online machine learning method for computing recommendations, which uses the logged context vectors to build a lower-dimensional representation of user interest. Learning the information interest function in this dimension helps to reduce the noise in the context vectors and overcome the problem of limited feedback. We connect the lower-dimensional interest on entities to contexts by making an intuitive linearity assumption between the relevance of entities and contexts. Finally, we define appropriate priors on parameters and likelihood functions (for different learning signals) to learn the user interest given recent contexts and the user's explicit feedback. The learned interest is then used for entity recommendation. The mathematical details are provided in the following subsections.

3.3.2 Exploiting Logged Contexts for User Interest Representation on Entities. We consider the unigram model and store the logged contexts in the matrix $X \in \mathbb{R}^{|E| \times |C|}$, where the element (i, j) describes the tf-idf weighting of the entity i in context j, E and C are the sets of entities and observed contexts, and $|.|$ denotes the set size. As mentioned, E and C are very large and X is noisy; still the data in X contains valuable information about how the different entities and contexts are correlated. We are interested to find a representation for entities, such that co-occurring entities get similar representations and at the same time reduce the dimension. To this end, we perform **truncated singular value decomposition (truncated SVD)** on X to get the projection matrix $W_K \in R^{|C| \times K}$ that enables us to project entities into a latent K-dimensional space. This dimensionality reduction in context space is justified as the logged contexts naturally contain redundant context vectors. The user interest is defined as a linear model in this latent space,

$$r^E = XW_K\theta, \tag{1}$$

where $r^E \in \mathbb{R}^{|E|}$ is the vector containing relevances of all entities (we use r_i^E to refer to the *ith* element) and θ is the K-dimensional latent user interest which will be learned in 3.3.4.

3.3.3 Connecting Entities and Contexts. The user interest θ, as introduced in Equation (1), only maps the entities to their relevance values. We follow the keywords-documents connection idea in [23] to connect the interest to the relevance of contexts by assuming that the relevance of a

context is a weighted sum of the relevance of entities that have appeared in it; in other words,

$$r_j^C = \sum_{i=1}^{|E|} p_{(i|j)} r_i^E, \tag{2}$$

where r_j^C refers to the relevance of the jth context (with some abuse of notation), and $p_{(i|j)}$ is the likelihood of the ith entity being present in the jth context. This likelihood is not available, but it can be approximated based on the logged contexts (i.e., X). We normalize the columns of X so that elements of each context vector sum up to one and denote the resulting matrix as $\hat{X}$. Using this approximation and writing Equation (2) in a vector format gives $r^C = \hat{X}^\top r^E$. Finally, by using Equation (1) we can directly connect the user interest to contexts

$$r^C = \hat{X}^\top X W_K \theta. \tag{3}$$

3.3.4 Learning User Interest. In the online phase of the study, the user can provide explicit feedback on the recommended entities through the UI. As mentioned, we additionally make the assumption that the recent "contexts", which the user has worked on, are relevant to the user's current intentions. This way, the interest function is updated based on both the similarity between contexts and the temporal information. To incorporate these types of learning signals, and for modeling convenience, we assume that the relevance is a sample from a Gaussian distribution with mean value r_i^E (if it is on the ith entity) or r_j^C (if it is on the jth context) as defined in Equations (1) and (3). The noise of these distributions should be different for entities and contexts, since the feedback on them is very different.

These feedback likelihoods are connected through the shared user interest θ. By assuming a Multivariate Gaussian prior on θ, we can complete the Bayesian inference loop and compute the posterior of θ after receiving explicit feedback and recent contexts. The posterior has a closed form solution and is employed to estimate the expected relevance of entities and contexts (Equations (1) and (3)), which, respectively, are used to rank all entities (of different types: people, keywords, and applications) and contexts (with their corresponding linked documents) to be recommended to the user. Details of the posterior inference are provided in the appendix.

3.3.5 Computational Complexity. The main computational bottlenecks of the system are the SVD calculation and the projection of all logged contexts and entities to the latent space. These computations can be done offline, already before the beginning of the online phase of user interaction.[3] We used the Gensim Python library [75] for fast and memory-efficient computation of truncated SVD. This offline computation took only a few minutes for all the participants of the study.

In the online phase, the computational cost comes from the posterior calculation, which is cubic on the dimension of the latent space $O(K^3)$ (see the Appendix). For K in order of hundreds the computation is immediate (we used $K = 100$ in the experiments). The source code is available at https://github.com/HIIT/Entity-Recommendation-for-Everyday-Digital-Tasks.

3.4 User Interface and Entity Interaction

Figure 2 illustrates EntityBot's UI. It implements three specific features: (1) showing the entities being recommended by the system, (2) allowing the selection of entities of interest by the user (explicit feedback), and (3) allowing direct action on entities when relevant. In the following, we describe how each of these features was implemented in our experimental setup.

---

[3]SVD can also be updated incrementally in real-time after receiving each context. Considering the short duration of the online phase of the experiment, we decided to build the latent space only once before the start of the experiment.
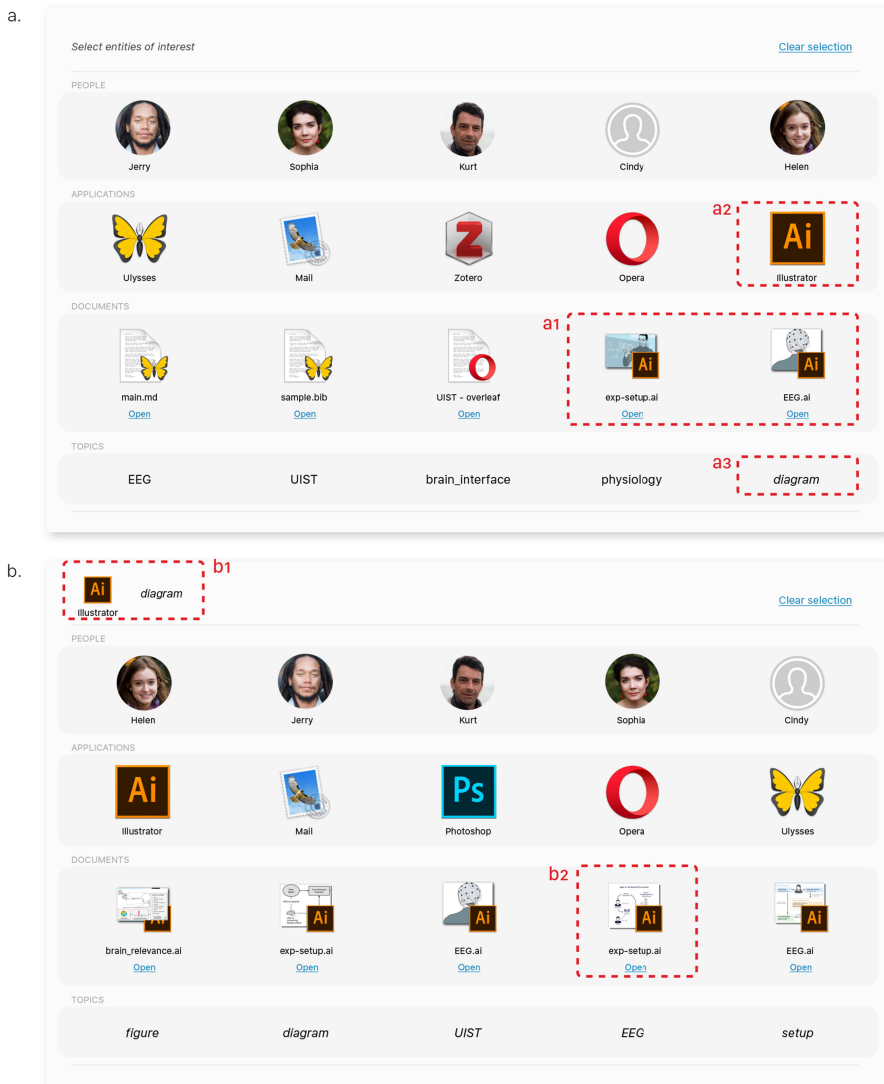
Fig. 2. Two states of EntityBot's UI. Recommended entities are displayed within four rows, here with five items each: people, applications, documents, and topics. The user can select entities of interest by clicking on them, which updates the recommendations. Example: In (a), the user sees entities related to her current work. She notices figures she has made for one of her papers (a1). She clicks on "Illustrator" (an application for editing vector graphics) (a2), then on the topic "diagram" (a3). (b) As a result, the entities of interest are displayed in the top area (b1) and the system updates the recommendations accordingly with the user's selection. In the documents row, she selects an illustration (b2) that she will modify for use in her new paper.

3.4.1 Showing the Entities Being Recommended by the System. Recommended entities are displayed within four rows of five items, with one row per entity type—that is, people, applications, documents, and topics (as keywords). People are identified by their name under a photo-based icon when available and a standard anonymous silhouette when not. Applications are identified by their name under a standard icon or logo of the application or service. Documents are identified by

their name under an icon based on a preview of their content, with a small icon of the application used to read or edit it. Each document has a hyperlink below the icon that facilitates access to the actual content. We implemented a script for the hyperlink that upon user click automatically opens the document using the path of the file, URL of the webpage, and unique ID of the email. Finally, topics are identified as a single keyword. In each row, the recommended entities are ranked horizontally from left to right. Since the main purpose is to show a small variety of the most relevant entities, the ranking is not visually emphasized. As users perform their tasks, the system progressively updates the recommendations. These changes are reflected on the UI as entities eventually shift places and new entities replace the old ones in each row. In the prototyping phase, since entities are displayed on an orthogonal grid, some users tried to derive meaning from the vertical alignment of entities across rows. To prevent that, the grouping of recommended entities by type in each row has been emphasized with a gray rectangle that acts as a container.

3.4.2 Allowing the Selection of Entities of Interest by the User (Explicit Feedback). When the users are interested in a specific entity among the recommendations, they must be able to express their interest in a way that informs the system so that the recommendations update accordingly. To that end, every recommended entity displayed on the UI can be selected with a click. As a result, the selected entity, or entity of interest, appears in the area at the top, and the overall recommendations (in every row) are updated, taking the selection into account (i.e., positive feedback on the selected entity is sent to the system). More entities can then be selected and added to the entities of interest at the top of the screen, providing an explicit way to influence the recommendations. Entities of interest can be removed from the selection by clicking the cross that appears at their upper-right-hand corner when the mouse cursor hovers over their icon. Removal of an entity of interest from the selection sends neutral feedback on the selected entity to the system, which updates the recommendations accordingly. The entire selection of entities of interest can be reset by clicking the "Clear selection" button on the right.

3.4.3 Allowing Direct Action on Entities when Relevant. An important feature of the system is that it makes the recommendations actionable. While work on translating recommended people and keywords into potential actions is ongoing, the present version simply allows one to open recommended applications and documents directly. Figure 3 illustrates the UI through an example scenario.

3.4.4 Example Scenario. Figure 3 illustrates the UI through an example scenario. In the scenario, Alice is evaluating a Master Thesis about interactions in virtual reality. The work is quite interesting, and she is almost done with the first pass of annotations. A notification pops up in the corner of her display: it is an email from the university administration stating the budget she has submitted for a conference trip in Hong Kong next month has been approved (Figure 3.1). This makes her think she better hurry if she wants to be able to choose which flights and hotel will better suit her needs. Prioritizing the new task, she interrupts the evaluation and opens the travel portal. Now, entities such as Hong Kong, extracted from the email notification and travel portal (Figure 3.2), are part of the recent context vectors, which are projected into a lower dimensional space and used to update the interest model. The updated interest model re-ranks all entities and prioritizes entities related to an earlier Alice task about the conference paper she has been working on. After a few seconds from the beginning of the new task, the entity recommendation display gets updated with entities of the budget document Alice sent last week, the concerned administration person, as well as Bob, a colleague and co-author of the same conference paper (Figure 3.3). By glancing at the screen, Alice sees the name of Bob and she recalls he said he will be attending the same conference. She realizes Bob could provide useful advice for what concerns which hotel to
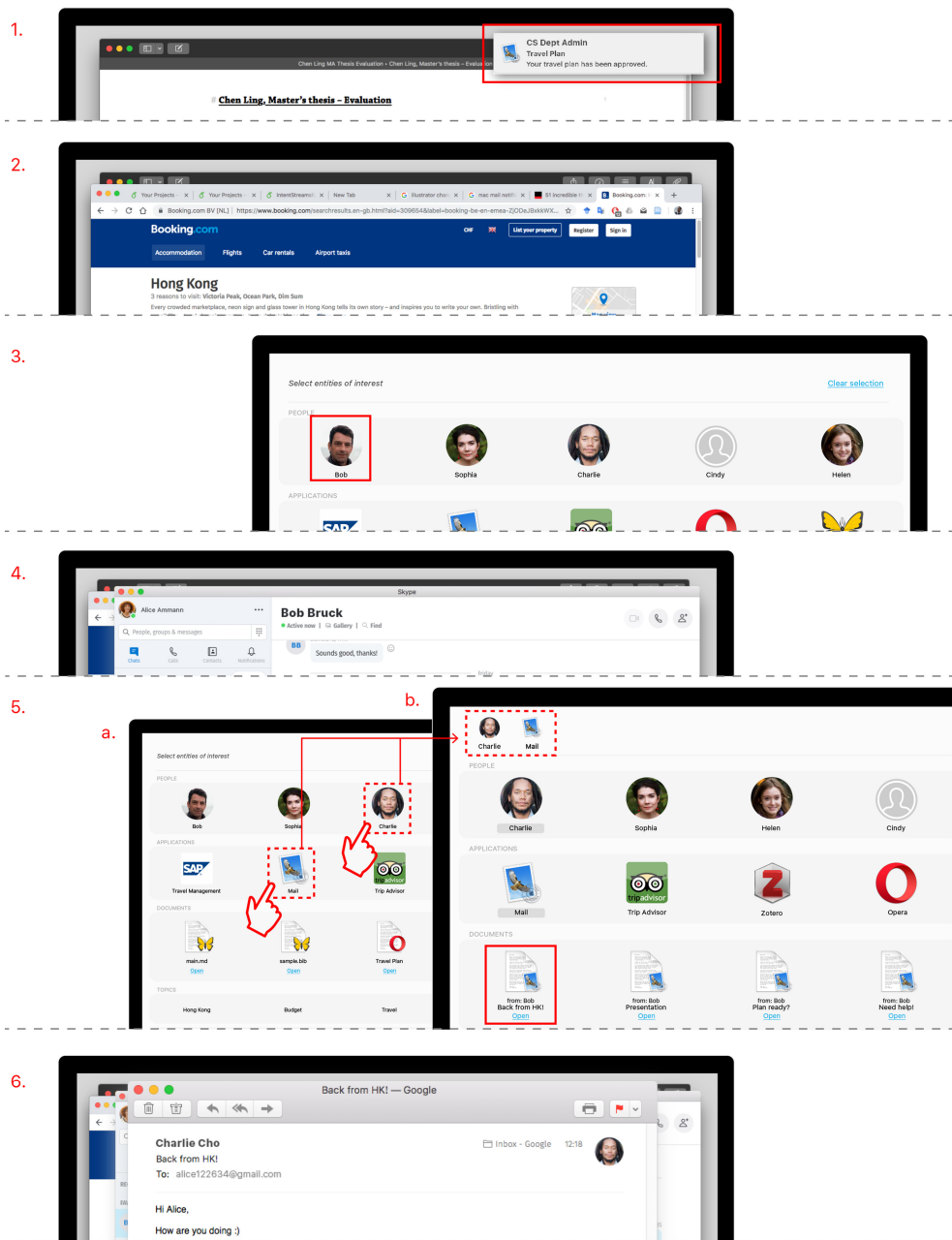
Fig. 3. An example scenario: A notification (1) reminds Alice she should book flights and accommodation for a conference in Hong Kong (2). Glancing at the recommended entities (3), she notices Bob, a co-author of the paper to be presented at the conference, and recalls he said he will be attending the same event. She texts him to learn where he will be staying (4). After bookings have been made, she notices in the recommendations her friend Charlie. She recalls Charlie went to Hong Kong last year and sent her an email about it. Selecting the entities "Charlie" and "Mail" (5a), she immediately recovers the email among the updated recommendations (5b). Its content gives her ideas of what to do there (6).

stay, as he had probably already planned his trip. She opens a new direct messaging conversation (Figure 3.4). The entity recommendation screen now shows the Hong Kong-related trip advisor links she has found in her past explorations, but the system also recommends her friend Charlie as this entity could also be associated through the updated interest model. That recommendation makes her recall Charlie had once sent an email describing his own trip to Hong Kong. Looking for that email, she selects Charlie and the email client, which tells the system to focus on items related to these entities in the next following recommendations (Figure 3.5a). With the provided feedback, the system manages to dig up Charlie's old emails (Figure 3.5b), which included some personal recommendations that have become useful at last.

4   USER STUDY

The purpose of the experiment was to evaluate the quality of the recommended entities, usefulness and influence of EntityBot with respect to the tasks, and users' subjective experience with EntityBot.

4.1   Experimental Design

The study followed a within-subject design with two system configurations:

— *Experimental (E)*: condition with the user model and the recommendations visible for the user. The input to the system is the context (user digital activity) and input provided via explicit interactions in EntityBot.
— *Control (C)*: condition with the user model, but recommendations not visible for the user. The input to the system is the context. In this condition, the recommendations did not affect what was presented to the user or how the user interacted and thus the condition functions as a control for the interactive presentation and feedback of the recommended entities.

In both conditions, participants could use any application running on their laptops, as they would do normally, with the only difference being the availability of interactive recommendations in the experimental condition. The control condition allowed us to quantify the *relevance* of recommended information resulting from the user model without incorporating the user model into the interactive system. The experimental condition allowed us to quantify the *relevance and influence* of the full system. The two conditions were counterbalanced by changing the order in which the participants were subjected to each condition.

For both system configurations, we consider the *User Screen (S)* as a natural baseline to quantify the information appearing on participants own screens, i.e., the context only. This baseline determines the information that the user has already accessed based on the applications available on his or her personal computer. The specifications of the conditions are shown in Table 2.

4.2   Research Questions

We defined the following research questions to understand the differences between the experimental and control conditions, and what appears in participants' own screens:

RQ1. Does the entity recommendation provide relevant information beyond what the user can find without it?
RQ2. Does the entity recommendation *influence* the user's information behavior?

4.3   Participants

We recruited participants by sending a recruiting email message to relevant mailing lists at the University of Helsinki in Finland. We provided a questionnaire in the recruiting message to collect

Table 2. Configurations of Each Compared Condition

	User Screen (S)	Control (C)	Experimental (E)
Context (entities extracted from screen data)	X	X	X
User model	-	X	X
Recommended entities visible	-	-	X
Explicit user feedback on recommended entities	-	-	X

The features are additive, that means the features of the user screen are included in
the control, and the features of the control condition are included in the
experimental condition.

information about the participants' backgrounds, the activities that they do, and the amount of
time they spent on their laptops in the past several weeks. Only respondents who used a laptop
as the main device for performing their everyday digital activities were considered eligible for the
study. Having a high educational background was another eligibility criterion, as it was assumed
that people satisfying this criterion would more likely use their laptops for everyday digital
activities. Overall, 14 respondents were eligible and volunteered to participate in the study. One
participant quit after three days due to a private reason, leaving the final number of participants at
13. Of these, six participants had bachelor's degrees, and seven participants had master's degrees.
Five men and eight women with an average age of 25 years (SD = 5) participated. In return for their
efforts, participants were compensated with 150 euros. Participants signed an informed consent
form that explained the study's procedure, data management, and data usage policy. They were
provided with an easy mechanism to interrupt/resume the monitoring at any time for tasks that
they did not feel comfortable tracing. Participants were also informed that they were allowed to
withdraw from the experiment at any time and all of their data would be removed from the server.

4.4   Ethics

We are aware of the privacy implications of using SM data for research and have taken active steps
to protect the participants. In particular, to safeguard participant privacy during the experiment,
the participants' logged data were encrypted, stored locally on participants' laptops during moni-
toring, and never exposed to anyone except the participants. Upon completing the experiment, we
only archived interaction logs during the lab study for evaluation, but we removed all personally
identifiable information and any user-level metadata. We followed the ethical guidelines and
principles of data anonymization and minimization at every stage of the data processing. The
logs were archived and stored in a server protected by authentication mechanisms and firewalls.
This work received ethical approval from the University of Helsinki in Finland. The experiment
consisted of two phases (two weeks of digital activity monitoring and a lab study), which are
illustrated in Figure 4 and described in more detail in the following subsections.

4.5   Phase 1: Data Collection

The purpose of the data collection phase was to collect a set of entities that the users had accessed
before the lab study. The entities were extracted after visiting web pages in the browser or access-
ing applications and documents stored locally on personal computers. The collected data were
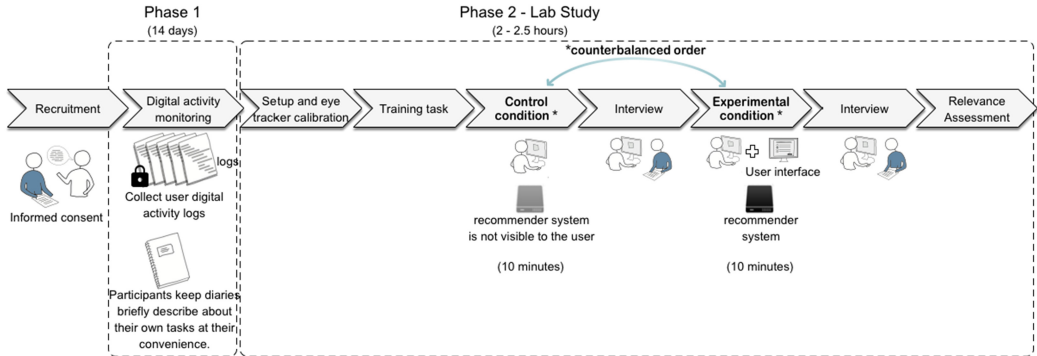used for recommendation.

Fig. 4. Two phases of the experimental procedure: (1) two-week digital activity monitoring in which logging software was installed on the participants' laptops to continuously collect digital activity logs which were encrypted and stored; and (2) lab study included system setup, a training task, two tasks selected and assigned under control and experimental conditions, interviews to collect direct comments and impression from participants following 10-minute task performance, and lastly relevance assessment on the recommended entities.

4.5.1 Two-week Digital Activity Monitoring. In the digital activity monitoring phase, the logging software was installed on participants' laptops and set to run continuously in the background thread for 14 days. Participants were advised to use their laptops as usual and avoid pausing the software unless it was necessary during this period. Participants were asked to keep a diary describing their daily digital activities. We provided a mini booklet with a diary template including four fields: a brief statement about a digital task, related keywords describing the task, people involved in the task, and the estimated duration of the task in a day. Participants were asked to write the diary using a pen and the provided booklet whenever they felt comfortable to avoid interference with their tasks. Several example digital tasks were demonstrated to ensure that participants understood the requirements of the diary, in particular the definition of a task. We define a real-world task that can be viewed as a collection of concrete steps striving for a certain goal. A broader real-world task can involve processing and accessing information through disparate sources such as word processing documents, web pages, emails, and any available files stored on a personal laptop. A task can be a wider ongoing project, thesis work, software implementation, or digital communication with other people. Alternatively, a task can be composed of leisure activities, such as making travel plans that include collecting vacation ideas from various websites, checking maps, accessing related notes, and making flight itineraries. To ensure that the diary was composed appropriately, the participants were informed that they would receive a weekly reminder about filling it in via a short text message. Following the discussion, participants also booked their relevant time slots for Phase 2.

4.5.2 Automatic Data Pre-processing. The data collected from each participant contained all Web activity logs and interaction history for all applications that were recorded during the two-week digital activity monitoring. The data comprises a stream of screenshots, where each screenshot is associated with a document location, an application name, and a timestamp. The document locations, including file paths, web page URLs, and unique IDs to emails, were obtained from the collected OS log information. In addition to the names of local applications, domain names of the recorded web pages were also extracted and considered as separate applications. OCR-processed documents were automatically pre-processed using a script. Contents of the documents were

tokenized by utilizing gensim library[4] and lowercasing terms. Keywords and people entities in the documents were not tokenized but only lowercased, with an underscore separating terms in multi-term entities. In addition, the nltk library's English stopwords[5] were utilized to remove stop words.

4.5.3 Data Collection Results. In total, the data collection consists of average 140,035 ($SD = 134,024$) screenshots per participant. This represents a history of average 78 hours ($SD = 73$) of computer usage per participant. We focused on unique screenshots by discarding duplicate screenshots or constant screen capturing of non-informative changes on the screen. This resulted in average 7,466 ($SD = 4,463$) screenshots per participant. The OCR process and EE resulted in an average of 1,204 ($SD = 555$) documents, 16,341 ($SD = 7,141$) keywords, 1,804 ($SD = 757$) people, 108 ($SD = 45$) applications, and 9,151 ($SD = 3,351$) non-entity terms per participant.

4.6 Phase 2: Lab Study

Participants came back to our lab for the Phase 2 of the experiment. We asked participants to review their diaries and select two tasks that they performed during the monitoring phase. In particular, we asked participants to pick two tasks that they felt were similar in category, with the same level of complexity, and comparable in duration. Participants were asked to write descriptions of the two selected tasks in a note. Then, the experimenter randomly assigned the two tasks to the two experimental conditions. To counteract fatigue and other carryover effects, we counterbalanced the order in which the participants were subjected to each experimental condition. Table 3 summarizes the tasks that were selected by the participants for the lab study for both conditions.

After selecting the tasks, the participants were briefed about the procedure of the experiment: they were asked to resume the tasks on their own laptop, engage in a short interview, and assess the relevance of the entities.

4.6.1 Experimental Setup. Prior to starting the experiment, the experimenter set up the participant's laptop to connect to a secondary display which was an integrated 22" monitor of the SMI RED eye tracker.[6] The second screen was turned on and the eye tracker calibrated prior to the experimental condition. The EntityBot's UI was set to run on the secondary monitor, and the participants performed their tasks on their own laptops. This experimental setup was designed in such a way that it was easy for the participants to observe that potentially relevant entities have been recommended, while at the same time it was easy to ignore the recommendations if the participants did not need support from EntityBot. Participants were told that they had the freedom to use or ignore the secondary monitor according to their needs. In addition, the OBS Studio screen recorder[7] was installed on the participant's laptop to record the screens of both monitors. The lab setup is shown in Figure 5.

The participants then conducted two tasks, one for each system condition, preceded by a training task.

4.6.2 Training Task. The purpose of the training session was to allow the participants to familiarize with the recommendation system. We trained participants on how to operate the EntityBot. Participants were allowed as much time as needed to get familiar with the system. Training sessions typically lasted around five minutes.

Table 3. Task Types, their Associated Examples, and the Number of Tasks that Were Selected and
Performed in the Lab Study for Both Conditions

Task Type	Examples	C	E
Data processing	Processing, analyzing, plotting, and testing data using spreadsheets, statistics software, text editors.	1	1
Thesis work	Materials gathering, thesis writing using various applications, such as file storage, repository, visualization tools, word processing, and a variety of websites.	3	2
Literature review	Reading articles, writing reviews pertaining to a specific topic using pdf readers, web browsers, and word processing.	1	1
Studying	Searching information pertaining to exercises, writing a report, language studying by accessing abundant resources from online tutorials and using many tools, such as word processing, pdf reader, file explorer, etc.	3	1
Personal information management	Reviewing notes, managing files and emails pertaining to a specific topic using note-taking tools, PowerPoint, file explorer.	1	1
Programming	Project implementation, Web development, figure sketching using matlab, other integrated development environments, and various web frameworks.	2	1
Course preparation	Material gathering and lecture slides preparation using web browsers, powerpoint, word processing, pdf readers.	1	1
Travel planning	Hotel booking, bus/flight/train ticket reservations, searches using map interfaces.	0	2
Social life	Managing group work, arranging shifts at work, communicating with others via instant messaging, email, online timetables, word processing.	1	3

C = Control, E = Experimental.

4.6.3 Main Tasks (10+10 Minutes). After the training phase, the participant executed the main tasks; one with the experimental condition and another one with the control condition. After each task, we went through the video recordings of the two screens (external screen only in the experimental condition), and asked the participants to explain what they were doing during the task. For the experimental condition, we put a special focus on understanding the participants' interests when interacting with the system.

4.6.4 Interview. After each task, we conducted a semi-structured interview to capture participants' direct comments and impressions of the experience when using the system. In particular, the interviews conducted after the second task included questions on the experiences in the two different conditions.

4.6.5 Relevance Assessment. Finally, we collected relevance assessments on the information that had been presented during the task. Assessing all task-related relevant information would
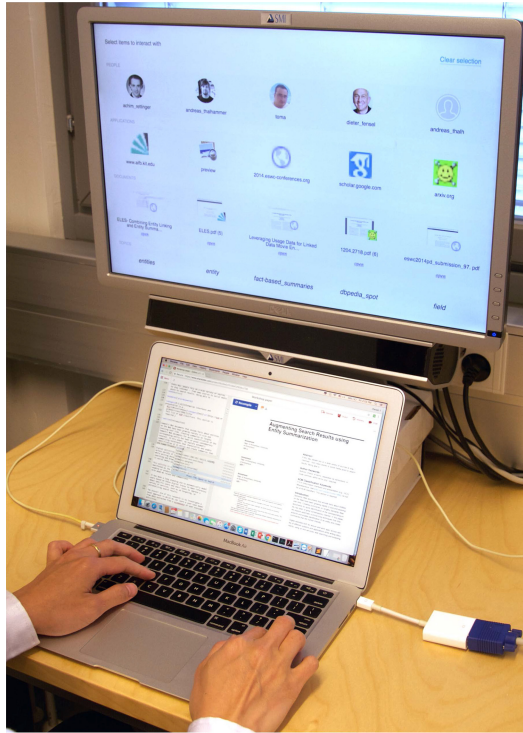
Fig. 5. The experimental setup. Participants used their own laptops to perform the tasks. An external monitor was set up to connect to the laptops showing the EntityBot's UI. SMI eye tracking device was installed and mounted onto the external monitor to track participants' eye gaze behavior during the tasks. In the figure, a participant is producing a manuscript for a journal submission while EntityBot continuously suggests relevant entities such as related articles and authors.

mean assessing the entire internet and all information stored on the user's personal computer, which is, of course, not practically possible. To address this challenge and diminish user effort and fatigue from manual assessment on a large number of entities, we used a standard pooling strategy, widespread in information retrieval research, that permits to obtain reliable results despite the possibility of having incomplete and imperfect relevance information [16].

We asked the participants to assess the relevance of three sets of entities: (1) a random sample chosen from all entities appeared on the participants' own screen, including the four aforementioned entity types; (2) a random sample chosen from the same number of entities appeared on the screen, from the top recommendations; and (3) the top five recommendations (the entities that would be selected for the UI). These entities were extracted at every recommendation point, which occurred at 10-second intervals.

At the end of the task, entities extracted at all recommendation points were merged, sorted alphabetically, and presented in spreadsheet file format for user convenience in providing relevance assessment. We asked participants to provide subjective relevance judgments on the entities according to how they were relevant to a given task. We demonstrated several examples of task-relevant entities to ensure participants understand the requirement. The participants rated the entities on a scale from 0 to 3 (0: not relevant, 1: low relevance, 2: medium relevance, 3: high relevance).

Table 4. Summary of Considered Measures and Results in Different Experimental Conditions

	Measure	Results
Relevance	Recall (in C, E, and S).	Figures 6 and 7
	Average precision and nDCG of recommended entities (in C and E).	Table 5
	Semi-structured interview about the quality of entities (in E).	Section 5.5
Influence	Utilization of recommended information (in C and E).	Figures 9 and 10
	Attention on the recommendation screen (in E).	Figure 8
	Explicit interaction with the recommendation screen (in E).	Table 5
	Semi-structured interview about the influence of entities (in E).	Section 5.6

4.7 Measures

Objective and subjective measures were defined to operationalize *relevance* and *influence* of information in different experimental conditions (see Table 4).

4.7.1 Recommendation Relevance (RQ1). Recommendation relevance was measured through **recall**, **precision**, and **normalized discounted cumulative gain (nDCG)** [47]. Recall was computed using the set of information appearing on a participant's screen, and in the conditions in which the recommendations were considered, the recommendations were added to this set. The recall is then computed as the fraction of relevant entities compared to all relevant entities discovered in any condition.

Precision and nDGC were computed for the top five ranked entities of each type in control and experimental configurations.[8] Precision shows the ratio of relevant entities in the top five recommendations. nDGC is a standard measure used to evaluate the ranking performance (in the range 0 to 1; 1 representing the perfect ranking algorithm) of the user model given the assessment scores. All measures were computed against a ground truth of relevant entities, which was obtained via relevance assessment by each user after the task. Please note that, as the recall is computed based on imperfect relevance information, caution should be used when interpreting its value in absolute terms (see Section 4.6.5). Instead, recall can be interpreted in relation to control conditions: the higher the recall, the higher fraction of relevant information provided to the user.

4.7.2 Influence of Recommended Information (RQ2). The influence was investigated through three types of measures targeting different aspects of human interaction with the entity recommendation system. **Utilization of recommended information** was measured as the fraction of relevant entities occurring in the user's screen subsequently after they were recommended. Only unique entities were counted. The rationale was that in order for the system to assist the user in performing the task more effectively, the system should recommend entities that are actually used by the user in the task. **Attention on recommended information** was measured as the total duration of the gaze fixations of the participants on the recommendation screen. **Explicit interaction** was measured as the amount of feedback (clicks) on the recommended entities and the number of opened recommended documents.

4.7.3 Subjective Experience of Relevance and Influence (RQ1 and RQ2). **The subjective experience of relevance** and **the subjective experience of influence** were investigated by using a semi-structured interview. The interview explored aspects related to the participants' impression that the system had influenced the task, including the quality of the experience of use, and the overall experience of relevance of the items displayed. Participants' answers were transcribed and

---

[8]Precision and nDGC are not computed for the user screen (S) condition due to not having a natural ranking measure.

underwent a thematic analysis. Transcripts were reviewed, and recurring themes were identified and organized into a codebook. The codes were then applied to the corpus of data [35].

4.8 Statistical Testing Procedure

We applied paired-sample t-tests to determine whether there are significant differences in performance among experimental, control conditions, and the condition that considers the information appearing on the screen. To test the significance levels, we used precision, recall, nDCG, utilization, and feedback as dependent variables and conditions as independent variables. Cohen's d values for t-tests were also computed to measure the effect sizes between the conditions. All normality tests were also conducted with Shapiro–Wilk. RStudio software v1.1.4 was used for the calculation of statistical significance with 0.05 as the p-value threshold. Please note that caution should be used when reading p-values. It has been argued that p-values in HCI research should be coupled with other kinds of visualizations and analysis, including qualitative results [27]. The experiment we present is seeking an indication of how entity recommendation brings benefits in realistic tasks. As the experiment does not have a hypothesis testing confirmatory character, a multiple comparison correction was not applied [11]. The reader is encouraged to read p-values in light of Figures 6–10, and the other qualitative findings for a correct interpretation of the results.

5 RESULTS

The average number of screens captured and recorded was 47 ($SD = 13$) in the control condition, and 43 ($SD = 13$) in the experimental condition. On average, the number of documents and applications opened during the tasks were, respectively, 11 ($SD = 6$) and 5 ($SD = 3$) in the control condition, whereas in the experimental condition participants opened 10 ($SD = 6$) documents and 6 ($SD = 3$) applications. There was an average of 411 ($SD = 216$) keyword entities and 34 ($SD = 23$) people entities that occurred during the task in the control condition, whereas an average of 465 ($SD = 36$) keyword entities and 40 ($SD = 36$) people entities were found on the screens captured from the primary monitor during the task in the experimental condition.

5.1 Recommendation Relevance

Figure 6 illustrates the recall of relevant items appearing on the user screen and in the user model in both control and experimental conditions during the lab study. For documents, applications, and people entities the user model has been able to find more relevant items compared to the user screen in both the control and experimental conditions (see Table 5 for a summary of statistics). In particular, there was a significant difference in the recall of applications in both control ($p = 0.002$, $d = 1.59$) and experimental ($p = 0.001$, $d = 3.89$) conditions, compared to the corresponding user screen conditions. Shapiro–Wilk tests indicated that our data were normally distributed ($W > 0.93$, $p > 0.3$). For the keyword entity, however, the user screen recall was higher than the user model conditions. This result reflects the intuition that our entity-centric approach may not work well with generic terms, such as keywords, which are typical of prior work [102].

Furthermore, the user model in the experimental condition has consistently achieved higher recall for all entity types, compared to the control condition. As shown in Figure 7, the user models in both conditions start from roughly the same level of recall. Performance was low at the beginning, but, after one minute the system in the control and experimental conditions performed well with high recall. The reason was because users did not manage to open many relevant documents for the task at first since they were asked to close all the windows prior to the test. After two minutes, the recall consistently improved and was higher than the user screen condition over time. The results indicate that although a majority of task-relevant entities can be retrieved within one minute after the user starts to interact with the computer, the best performance of the system is achieved
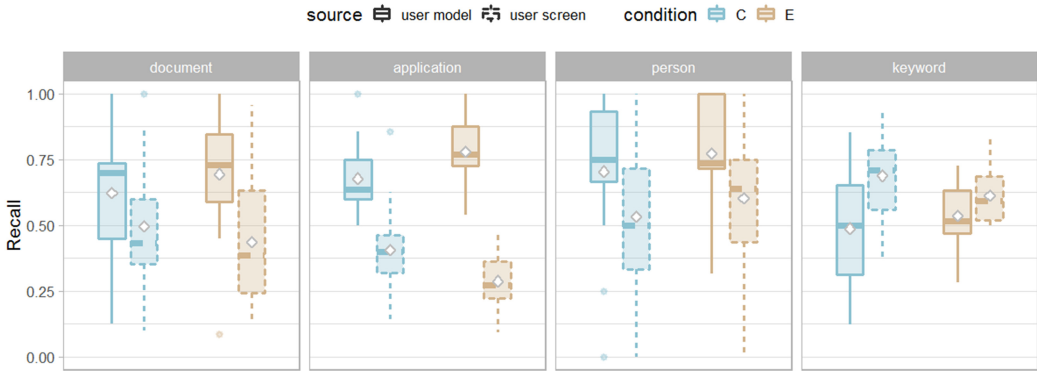
Fig. 6. Recall. Paired comparison between user screen and user model in both control and experimental conditions.
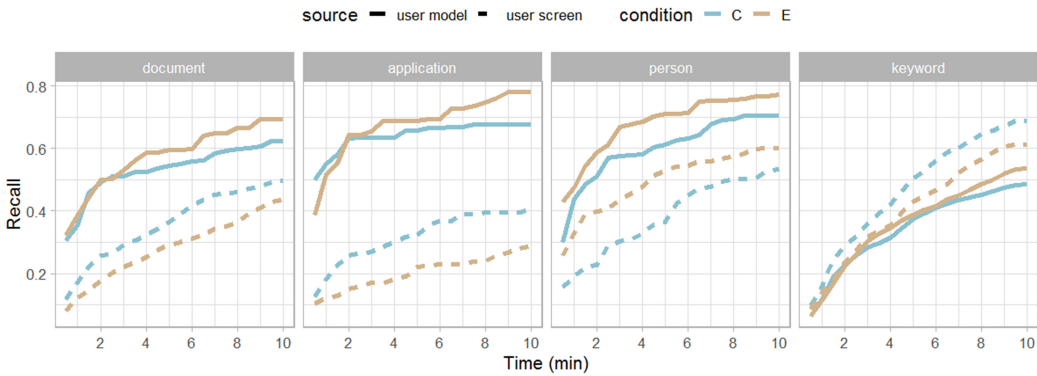

Fig. 7. Recall over time for user screen and user model in both control (C) and experimental (E) conditions.

only after two minutes of digital activity monitoring. However, the improvement is faster in the experimental condition, indicating the positive effect of user explicit feedback on recommended entities. This effect is also evident by comparing the precision and nDCG of the top five predictions of the user model in Table 5 (Precision and nDCG columns).

5.2 Use of Recommended Information

Figures 9 and 10 provide a quantitative indication of how the information provided by EntityBot was utilized in the user tasks. More specifically, the percentage of recommended document and application entities that participants used in their task only after they were recommended by EntityBot was significantly higher in the experimental condition (in which the entities were actually shown) than in the control condition (in which the system was running in the background) with $p = 0.04$, $d = 1.1$ and $p = 0.04$, $d = 0.88$, respectively (see Table 5 for statistics). Shapiro–Wilk tests also showed normality in the data with $W > 0.87$, $p > 0.07$. This result indicates that the recommended entities were used in the task because participants saw them on the EntityBot's UI, and not by chance. However, for people and keyword entities, we did not see any significant difference between control and experimental conditions. This result suggests that specific actions on these entities were not needed for performing the tasks. Nevertheless, the people entity was

Table 5. Results: (1) Recall of Relevant Items Appearing on the User Screen and in the User Model in Both Control and Experimental Conditions during the Lab Study

	Recall			Precision	nDCG	Utilization	Feedback
	User Model	User Screen	Comparison	User Model	User Model	User Model	User Model
Applications							
Experimental	0.78 (0.15)	0.29 (0.10)	$p < 0.001$	0.55 (0.23)	0.69 (0.01)	0.46 (0.44)	1.62 (2.40)
Control	0.68 (0.15)	0.41 (0.19)	$p = 0.002$	0.48 (0.22)	0.68 (0.02)	0.17 (0.17)	
Comparison	$p = 0.04$	$p = 0.10$		$p = 0.48$	$p = 0.83$	$p = 0.04$	
Documents							
Experimental	0.69 (0.26)	0.44 (0.25)	$p = 0.09$	0.73 (0.27)	0.67 (0.04)	0.26 (0.28)	1.46 (2.03)
Control	0.62 (0.25)	0.50 (0.25)	$p = 0.36$	0.65 (0.30)	0.58 (0.03)	0.05 (0.10)	
Comparison	$p = 0.42$	$p = 0.52$		$p = 0.51$	$p = 0.30$	$p = 0.04$	
People							
Experimental	0.77 (0.21)	0.60 (0.28)	$p = 0.13$	0.39 (0.34)	0.49 (0.01)	0.04 (0.06)	1.62 (2.57)
Control	0.70 (0.30)	0.53 (0.31)	$p = 0.28$	0.23 (0.21)	0.40 (0.02)	0.11 (0.28)	
Comparison	$p = 0.48$	$p = 0.41$		$p = 0.10$	$p = 0.36$	$p = 0.38$	
Keywords							
Experimental	0.54 (0.12)	0.61 (0.11)	$p = 0.25$	0.46 (0.28)	0.52 (0.05)	0.01 (0.01)	0.23 (0.60)
Control	0.49 (0.22)	0.69 (0.16)	$p = 0.07$	0.42 (0.26)	0.48 (0.07)	0.01 (0.01)	
Comparison	$p = 0.45$	$p = 0.19$		$p = 0.74$	$p = 0.62$	$p = 0.73$	

Apart from keywords, the user model was consistently able to find more relevant items than the users would find on their own. The difference was significant for the entity *application*. Additionally, on average, the user model found more relevant items in the experimental condition. The difference was significant for the entity *application*; (2) Precision; (3) nDCG; (4) Utilization; and (5) the average number of feedback clicks on entities. *SD* values are provided in parentheses.

Fig. 8. Heatmap visualization of the eye gaze fixations on EntityBot.

an important memory cue and feedback to direct the recommendations to relevant and useful documents and applications (Table 5, Feedback column).

5.3 Interaction with Recommended Information

During the attendance to EntityBot, on average, the participants provided explicit feedback on 4.92 entities; on average once every 16 seconds during the attendance to EntityBot. Most of that feedback was targeted at applications, people, and documents (Table 5, column feedback). Participants used the hyperlinks provided by EntityBot to open, on average, 18.6% (SD = 26.7%) of the documents supporting them performing the task, and all these documents were assessed highly relevant

with a score of 3. A comparison with Section 5.2 suggests that most of the times the utilization of recommended documents happened through the use of hyperlinks (M = 72.7%, SD = 34.4).

5.4 Attention on Recommended Information

On average, participants spent 1.3 out of 10 minutes of the lab study attending to EntityBot. This indicates that EntityBot was used, but the participants still tended to focus more on the main task with 87% of the task duration. As expected, most of the task-related activities were performed with the participants' own laptops. A heatmap visualization of the gaze fixations to EntityBot is shown in Figure 8. In summary, these results, together with those of Section 5.3, indicate that EntityBot was used more than 10% of the total task duration for implicitly perceiving information, explicitly providing feedback to direct the recommendation process and accessing important documents related to the task.

5.5 Subjective Experience of Relevance

Overall, the participants reported a positive experience with the EntityBot system (N = 9): *"I enjoyed it... I got surprised that it recommended things that I just intended for" (P01).* All participants (N = 13) found the entities recommended by the system relevant:

> *They were right on the spot (P04).*

> *They were surprisingly relevant, various kinds of applications and various kinds of materials... I had for example this task manager, various emails, various files... they were really relevant for the subject... so it was quite good in that sense (P07).*

> *They were mostly very good. I think there was maybe a couple of them... they weren't relevant, but mostly it had picked up the right side (P03).*

However, some participants (N = 2) specified that it took a while for them to get relevant results: *"At first they didn't have much to do with what I was doing... but the more feedback I gave to the application I felt, like, relevance got a lot higher each time, so in the end, they were pretty good" (P11).*

Furthermore, 10 participants mentioned that EntityBot helped them recall specific entities useful for their task:

> *It reminded me Github. Although I would have opened it while I was working, it didn't really came to my mind... I could forget to commit my code, so it is useful (P05).*

> *I didn't remember that article that I got in the recommendation until I was reminded by the interface and then I was, like, 'Yay, that is actually something that's gonna help me further with the subject' (P12).*

5.6 Subjective Experience of Influence

The majority of participants (N = 10) felt that EntityBot affected the task:

> *It made it easier to find the book, and faster (P11).*

> *It had a positive effect. I was thinking to remember the file name, and what I did was to just watch what the system was recommending me, and I just immediately found the name of the file. So it saved my time, otherwise I would have to find the location where I put that file. So it just simply did it for me (P01).*

> *Yes, it did. For example, I ended up opening one of the files that was suggested, so that was something that directed quite a lot the task (P04).*

A smaller proportion of participants ($N = 3$), however, did not have the same impression that EntityBot had affected the task, either because the task was easy and did not require much support or because of the limited time they had for the task: *"No. I knew what I had to do so I didn't need the support" (P10)*; *"Just within 10 minutes it didn't really affect, apart from when I opened my 'Github Desktop'" (P05)*; *"10 minutes is very short for writing a thesis" (P10)*.

Most participants reported that EntityBot was useful at some point during the task ($N = 11$), typically by making it easier to access information or to recall relevant information related to the task. More specifically, six participants found EntityBot useful in enabling faster information access:

> *It was smart enough to recognize that when I use Overleaf it would give me the book I'm using as material... and I could open the book, the link right away from there... That was a big help because if it wasn't there I would have had to go to the University's library website and then search the book and then go... [...] and then open it... but now I could just click it right there (P11).*

> *I was looking for a file [thesis doc] that is related to the one I have [opened in my laptop], a separate one that has comments on it, and I found it there [on the EntityBot system], and it was exactly what I was looking for... without... I was not writing anything, like, that I am looking for a file, it was just in my mind... but somehow it appeared (P07).*

EntityBot was found easy to use by all participants. The interface was found to be *"straightforward"*, *"simplistic"*, and *"obvious"*. Participants particularly appreciated the proactiveness of the system and how it effortlessly provided useful information:

> *I didn't have to do anything, it just suggests [things] for me there. So it's not yet another search-ware or anything that I need to act on it. It's just there, I don't need to do anything with it but when I look at it, it provides me relevant solutions so... easy (P07).*

As the interviews indicated, the majority of the participants ($N = 11$) noted a difference between the experimental conditions and preferred working with the support of EntityBot. Notably, four participants explained how the system was considered a companion that helped carry out the task with new ideas and made it easier to stay on track:

> *The system gives me ideas to think out of the box... it's like talking with somebody, having a discussion that it could widen my ways of thinking how the task should be done (P02).*

> *It's like having a notebook next to you with notes on what you should be focusing on, like a list of things you should be doing (P12).*

However, some participants ($N = 4$) reported an increased effort due to the switch of attention required by the multiple screens setup: *"My attention was a bit divided because of the two screens..." (P12)*; *"I switched my attention sometime to the tool were I saw the various suggestions... it could be even not so good if I had all time switched attention" (P03)*.

The majority of participants felt in control of EntityBot ($N = 11$) because they had the feeling that the system was reacting to their digital activity. In particular, six participants pointed out that the feeling of control was mainly due to the capability of directing the system through the *feedback mechanism*:

> *When I checked some boxes [gave feedback] or took them off [removed the feedback], it [the EntityBot's content] changed, so, I kind of felt like it was me who affected how it worked, so I felt I was in control (P03).*

Nevertheless, two participants pointed out that they would feel more in control if they could remove unuseful entities from the set of recommendations: *"I think I would feel more in control if I could actually take off the recommendations, like, the tasks [keywords] or the pages [documents] I already did, like, [those] I don't have to use anymore" (P11).* Also, one participant didn't feel in control: *"No, since it's only giving me something that I can choose" (P10).*

Some participants had specific wishes or suggestions for how to improve the system. More specifically, three participants mentioned that they would prefer to have the system running on their laptops: *"If it could have been somehow embedded into the same screen... the suggestions would have been directly there where I was working" (P04).* Finally, three participants missed the capability to conduct explicit searches, at least for initializing the system: *"In the beginning it's a bit hard because you cannot set any recommendations... it would be useful to set for example some keywords right away" (P12).*

6 DISCUSSION AND STUDY FINDINGS

6.1 *RQ1: Does the Entity Recommendation Provide Relevant Information Beyond What the User Can Find Without It?*

Although the findings are promising, they do not permit us to provide a definitive answer to this question. We observed that the entity recommendation is able to find noticeably more relevant entities than those found on the user screen during the actual task in both the experimental and control configurations (Figure 6 and Recall in Table 5). The improvement is consistently greater when the user is able to interact with the system (experimental condition) compared to the control condition (see precision and nDCG in Table 5). The benefits of the interaction and user feedback are also supported by the qualitative results from the interviews, which showed that the feedback mechanism has an important role in improving the quality of recommended entities for the majority of participants.

We saw participants engage in tasks ranging from thesis writing, data processing, and coding, to travel planning and other social tasks. Within such tasks, participants were involved in various activities, and their interests frequently changed. As demonstrated in other recent work, systematic and deep processing of content can provide opportunities to personalize access to information [75]. Results on the relevance and influence of recommendations, together with qualitative findings, show that the system effectively captured participants' rapidly evolving interests and provided them with recommendations that positively influenced their tasks. However, we learned that a method based on digital activity monitoring also has drawbacks. For example, active windows may include information that is always visible regardless of the particular task, such as web bookmarks. This information may confuse the model.

Overall, we believe that while our quantitative data provided us with interesting insights, which were supported by qualitative data, longer sessions and more data would have provided better opportunities to answer this research question more conclusively.

6.2 *RQ2: Does the Entity Recommendation Influence the User's Information Behavior?*

Yes. In the experimental condition, participants had access to recommendations that were subsequently used to perform the tasks (Figures 9 and 10). Interaction results (see Section 5.3 and Figure 8) show that participants looked at the recommendations and used the hyperlinks provided by the system to access documents useful for their tasks. In particular, from Figure 10 we can see that the improvement in utilization comes already within 2 minutes and continues to grow improving 20%–30% for the document and application entities.
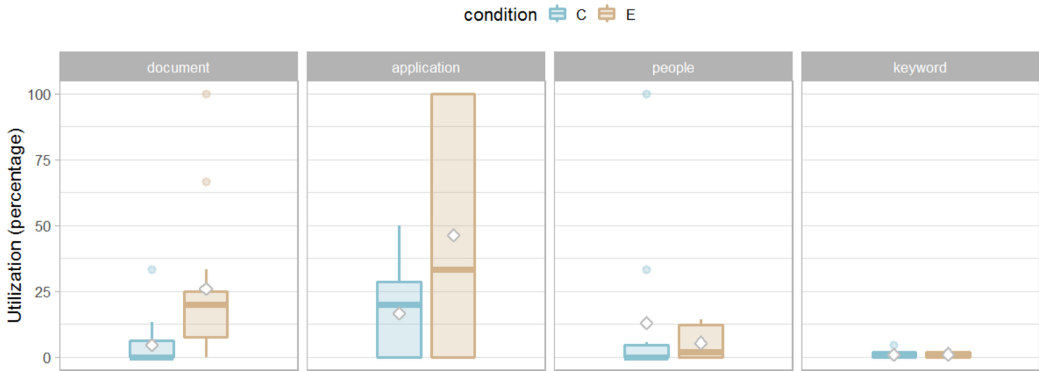
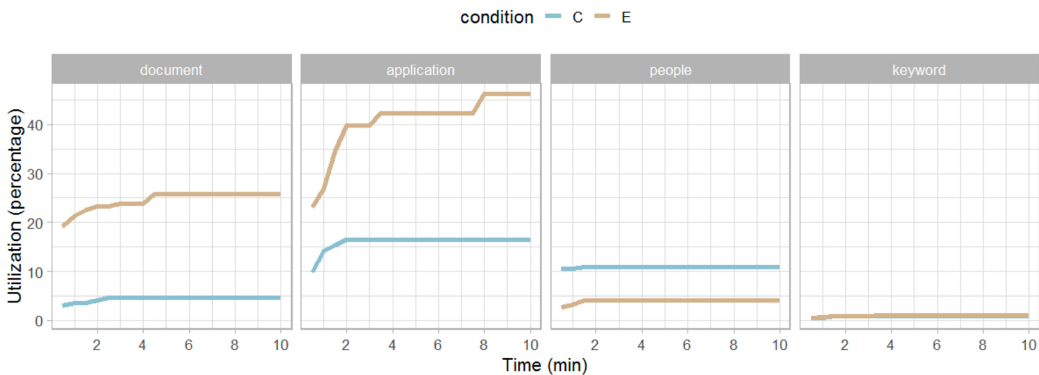Fig. 9. Utilization of recommendations on the task.

Fig. 10. Utilization of recommendations on the task over time.

Qualitative feedback from interviews further confirmed that the recommendations positively affected the task by reminding participants about specific activities or pieces of information that were relevant to their task. The system was able to recommend useful, forgotten, or unknown resources. One challenge with complex everyday tasks is recalling useful resources that we bookmark or store for future reference. As part of their tasks, people typically organize a wide range of resources that they think could become useful to allow easier retrieval at a later stage. For example, people bookmark web pages, organize emails in folders, or save documents to their hard drives. However, these useful resources are often forgotten until well after their usefulness has passed [50]. With our system, we remove this "out-of-sight out-of-mind" problem by automatically recommending important entities that users need for performing their tasks at an appropriate time.

Results on the utilization of recommended information on the task (Table 5) show that 26% of the documents and 46% of the applications recommended in the experimental condition were subsequently used to accomplish the task. The comparison with the control condition, yielding respectively, values of 5% and 17%, demonstrates that in the experimental condition, on average, 21% of documents (26% − 5% = 21%) and 29% of applications (46% − 17% = 29%) used to perform the task would not have been used without the support of our system. Qualitative findings from interviews confirm the intuition given by the quantitative results, with ten participants reporting cases in which the system helped recall useful entities related to the task.

Additionally, the entity recommendations provided a source of inspiration, as indicated by interviews. This suggests that the task-related context structures extracted by the system contained

insightful information that users had not previously noticed. Participants generally reported an improved user experience when performing tasks with the support of EntityBot. Participants perceived the system as a companion that provided useful insights on how to perform the task. By reminding participants about the various entities related to the task, the system permitted the reconstruction of the typical activities they performed during their task. Furthermore, the system allowed faster access to information. During their tasks, participants needed to access various documents, such as PDFs, emails, code snippets, or websites. Forgetting the name or location of a file, the applications used to communicate with team members, or the subject of an important email could make it difficult to retrieve the needed information. With our system, participants could easily retrieve important documents they needed for their tasks, either without any explicit input (e.g., by just opening a different relevant document) or by selecting key entities in the system. This allowed the participants to save time with a consequent improvement of the perceived experience. The benefits provided by the system came with moderate costs in terms of division of attention, as reported by some participants. Overall, the participants reported a better user experience when the implemented system was available.

One key finding related to an entity-centric approach is *enabling effortless access to information*. In the interviews, participants reported cases in which the needed entities were correctly predicted at the right time, based solely on the implicit interactions of the users, permitting immediate access. Other times, the entities could be easily retrieved after giving feedback to entities already recommended by the system. Prior research points out that, although searching seems to offer obvious benefits over other methods, in many cases people still prefer to find their personal information by navigating through folders [95].

A second finding concerned the use of *entities as feedback*. Results show that the explicit feedback was consistently used during the 10-minute experimental tasks, with an average of 4.92 selected entities. Qualitative findings suggest that this mechanism was one of the main factors contributing to the overall positive experience with the system. While the good acceptance and use of explicit feedback align with the idea that interactivity is an important factor to be supported in search activities that are part of wider primary tasks [48], it contrasts with other studies where feedback mechanisms were not well received [56, 105]. One possible explanation is that using entities as interactive objects could have played a main role in fostering a more active use of feedback. This intuition was confirmed by the low usage of the keyword entity, which was the information item that most resembled the information used in other systems.

A third finding consists of recommending *actionable entities*. Results from the experimental condition show participants used the provided hyperlinks to open documents directly from EntityBot (Section 5.3). Making the entities actionable supported faster access to information, which positively affected the user experience as discussed above. However, it is interesting to note that the recommendations were useful regardless of whether people clicked on them. For example, from the interviews, we learned that a participant engaged in a task that required naming files with a certain convention. The participant explained that to retrieve the information he would have had to refer to a prior project, which in turn would have required remembering the location of the files with the desired naming convention used in that project. The recommendation of a file with the proper naming convention, before he could even start searching for it, saved P1 much time and did not require opening the file; the mere sight of it solved the problem.

6.3 Limitations

Evaluating recommender systems is a challenging task. In this work, we aimed to study how entity recommendations could support people in real-world situations. We combined an in-the-wild data collection phase with a lab study phase where people resumed their real-world tasks. The

experimental procedure permitted a good degree of ecological validity together with the control of confounding factors during the actual intervention in the lab. On the other hand, our work suffers from the limitations of both in-the-wild and controlled studies. While the lab setting permitted control over many factors (e.g., the delivery mode of recommendations through the same secondary screen), the fact of being outside the user's natural work environment may have affected the naturalness of user behavior.

In this article, we did not compare the recommendation methods with other approaches such as using accessibility data [99]. So, we do not know if a method that uses only specific data sets such as logs would be better or worse than our method but only in one type of task, as the focus of the research was to study the actual influence that comprehensive entity recommendations have on everyday digital tasks.

Our approach did not consider data besides text (e.g., images), which may have provided additional information beyond the text captured from screen recordings. This is also a limitation but may also open new opportunities to utilize image data and computer vision techniques to capture even more precise user activity. On the other hand, although an image might have no textual information within it, we leveraged other meta-information surrounding the file, such as its title, application name, and URL in case that the image is on a webpage. These textual contents may contain useful pieces of information that provide much of the user's context for the image. Furthermore, although our study does not suffer from the limitations of using simulated tasks, permitting the observation of real-world situations, it misses the advantages of controlling the type of tasks participants performed.

The EntityBot system designed for our study also presented some limitations that future work could address. While our approach based on screen content allowed us to build rich user models from a single data source with little or no human supervision, and to provide useful entity recommendations, the method also has drawbacks. For example, our approach does not handle the entity resolution problem, which could lead to the recommendation of multiple entities referring to the same real-world entity. Furthermore, our modeling solution, at its core, uses linear models to tackle the three main challenges in the considered setup (i.e., high dimension and inherent noise in the digital activity monitoring data, limited explicit interaction with the user, and real-time performance necessary for interactive use). Although linear models are the preferred solution for the mentioned challenges, more complex models (e.g., deep neural networks) may be able to learn more complex user interests. Moreover, in this work, we considered the expected relevance of entities as the measure for recommendations.

Lastly, the data collection in the digital activity monitoring phase involved participants who turned the monitoring off for some part of their activities. Further, studies of the kind of data the participants conceal on purpose would help set the privacy boundaries expected by users in a more automatic way.

7 CONCLUSIONS

Although entity recommendations are emerging in commercial platforms and services, research in this area is scattered and often based on specific applications, focusing on the quality of ranking or relevance away from realistic use. Our aim was to revive the area of entity recommendations for everyday digital tasks. We contributed an approach informed by previous work and updated by two core principles: comprehensive digital activity monitoring and entity-based interaction. We implemented our approach in a concrete system (the EntityBot) and validated it through a user study with realistic data. The study contributes to investigating the usefulness of entity recommendation in realistic everyday digital tasks beyond relevance, which has been generally overlooked. Next,

we reflect on the core principles of our approach and propose themes as elements of a research agenda for future work.

7.1 Comprehensive Digital Activity Monitoring

The study demonstrated that an entity-centric approach based on digital activity monitoring can effectively support users working with heterogeneous applications in their everyday digital tasks. Digital activity monitoring, in particular a screen recorder built on top of the applications' UI layer, permitted us to extract context across application boundaries, overcoming one of the main limitations of prior work. Future work could explore how contextual information about the task and text similarity could be used to address the entity resolution problem. Considering the notion of comprehensiveness, what we were able to capture was comprehensive considering the task as carried out on a laptop, while a task might have actions carried out on another device such as a smartphone or in the physical environment such as talking to someone or interacting with real-world objects. In the following, we consider additional types of data for the monitoring module.

— *Mobile devices.* Information arising from monitoring mobile devices is different from data and information confined to desktop and laptop computers. While being mobile, the information on the screen is more often likely to be linked to real-world interactions and contribute with complementary information. Researchers should investigate in which way this type of data contributes to task and user models and to the usefulness of recommendations.

— *Audio.* Recent developments in recognizing entities from soundscapes and speech provide yet another rich and complementary source of information. For example, we have developed a search bot capable of listening to conversations [6], detecting entities to be used as query context to retrieve possibly useful related entities.

— *Vision.* Recent vision-based retrieval systems support entity detection in images, providing a source of information about the immediate circumstances around the user in the physical world.

These are opportunities to further research and push entity recommendations beyond the digital realm accounting for interactions and entities in the physical world.

7.2 Entity-Based Computing and Interaction

The entity-based approach in particular, as implemented in the UI, had several benefits including effective access to entities (people, applications, and documents) and easy collection of relevance feedback.

Enabling effortless access to information. One of the advantages of automatically predicting and presenting users with potentially useful information in advance is that when users decide to search for it, the information might be readily available without users needing to formulate queries [85]. This was often observed in our study. Because people typically need to retrieve information in support of a primary task, they prefer to do it in the most automatic ways that take less attention away from their primary tasks. While searching or managing information in folder structures requires effort [14], novel approaches should support users in concentrating on creative tasks while automating the rest. With our approach, recommendations were automatically computed in the background without requiring users to formulate queries, thus requiring less mental effort. Additionally, our entity-centric approach provided visual cues that iteratively guided users toward the desired information by following familiar elements, similar to the events in folder navigation scenarios. However, our method shares the limitations of other systems for search and

recommendation. Recommendations do not always arrive in the same order, making them less consistent than navigating through well-known structures.

Entity-centric approach, feedback, actions, and feeling of control. Entity selection was found to be a simple and effective way to specify one's own evolving interest when performing a task in order to retrieve the desired information.

Results show that the explicit feedback was consistently used during the 10-minute experimental tasks, with an average of 4.92 selected entities. Qualitative findings suggest that this mechanism was one of the main factors contributing to the overall positive experience with the system. While the good acceptance and use of explicit feedback aligns with the idea that interactivity is an important factor to be supported in search activities that are part of wider primary tasks [48], it contrasts with other studies where feedback mechanisms were not well received [56, 105]. One possible explanation is that using entities as interactive objects could have played a main role in fostering a more active use of feedback. This intuition was confirmed by the low usage of the keyword entity, which was the information item that most resembled the information used in other systems.

Overall, our quantitative and qualitative results suggest that giving feedback to entities could be an effective way to find the desired information while reducing the feeling of not being in control, a problem often faced when using information retrieval, and in particular, recommender systems [5, 90]. However, as suggested by some participants, future implementations should consider adding support for negative feedback and explicit search to provide users with even more control.

Providing actionable insights. Traditional keyword recommendations may not necessarily indicate possible directions for performing a task, therefore providing users with little practical value. Entity recommendation, on the other hand, provides more actionable information items, because the item type already suggests the actions that could be done on the item. For example, seeing a recommendation of a person entity already signifies the actions that can be associated with that person, such as calling, texting, and emailing. By providing hyperlinks that permit instant access to related resources, our system makes the insights provided by the entity recommendation even more actionable. Results from the experimental condition demonstrated how participants used entities to open documents directly from EntityBot, making the entities actionable by supporting faster access to information. The entities were however useful regardless of whether people clicked on them as they provided useful cues and information that were used affecting the information behavior. This is in line with prior research on proactive information cards [88] and good abandonment in aggregated search [58] showing that the retrieved content does not need to be clicked to be useful. Although in this study we quantitatively measured the influence of entity recommendation, our method was not able to account for the situation of good abandonment. Devising a measure that more comprehensively takes into account the various facets of influence is a problem that requires future research.

8 FUTURE WORK ON ENTITY RECOMMENDATION

Our contributions so far have included synthesizing an approach to previous scattered work in this area, advancing the field toward more comprehensive monitoring, entity-based approaches, realistic evaluation, actual usefulness, and impact on behavior. In the following, we highlight themes that can contribute to a research agenda in this area.

Revisiting digital activity monitoring encompassing the physical world. As discussed above, capturing all that appears on the screen of a desktop or laptop computer is more comprehensive than monitoring within one application or collecting only certain types of data. Considering mobile devices for digital monitoring will yield more or novel information since users might continue a task using sequential devices or use different devices for different tasks [49]. A variety of questions arise, not only on the difference and information value of data sets originated

from mobile devices versus computers, but also on how the resulting recommendations should be computed and displayed. For example, different recommendations could be displayed depending on which device is used. Considering additional sensors (such as location, accelerometers, nearby devices, or other sensor information), the data will include contextual information that could be useful in improving the quality of recommendations. Finally, considering audio and visual sensors, the monitoring can more clearly include information of entities in the physical world, opening yet more avenues of research.

Revisiting entity interaction toward actionable and resourceful entities. The usefulness of a recommended entity can be tied to the system's ability to provide affordances for action by linking to the functionality of a specific application. As an example, recommending a person might provide an easy handle to call, email, chat, or author search. Machine learning research can be extended further by investigating the recommendations of entities in relationship by extending the modeling of entities, users, and their tasks. UI and evaluation in realistic settings should investigate how to include such new functionality.

Explainability and transparency of recommendations. Qualitative results suggest that enabling interactive feedback on entities may increase trust and perceived control of the system. However, a feeling of control and trust should be guaranteed even when relying on implicit feedback alone. A recent important theme in recommender systems is their explainability and transparency [108]. In the case of entity recommendations, the user might find it useful to have explanations because they affect decisions and selection [69]. An important aspect to consider is the possibility to design ways in which the user models can be explicitly presented to users [8], in this case possibly representing tasks and allowing users to give feedback or edit it directly [81], which gives them more control beyond mere accountability. Researchers should investigate the trade-off of providing more information for explainability or functionality for editing user models versus the cognitive effort and increasing complexity of the resulting UI.

Intent drift and routines. Our approach was evaluated using a two-week monitoring period. In this context, it is more difficult to observe the evolution of tasks, their life cycles, and routines. We think this research theme, connected to the temporality of observation, is a formidable avenue of research. More sophisticated machine learning methods can be developed to consider the decay of the relevance of data. Furthermore, longer observations can provide opportunities to learn routines by understanding the recurring relationships of entities. All of this requires developing further models of users and tasks.

Implicit interaction. A variety of implicit signals have been used in modeling users in information retrieval, including behavioral data such as click-through data. A well-researched signal is eye tracking, which has shown potential as implicit relevance feedback in documents and image search [18, 70]. Capturing screen content and other contextual information, including speech as discussed above, is a type of implicit data collection. Recently, neurophysiological measures beyond behavioral ones have emerged as an information source for informing relevance or interest models [37]. These types of signals can potentially provide effortless relevance feedback without distracting the user as they do not require the user's explicit action or attention [30, 43]. While these studies show that the signals are noisy, they also demonstrate the feasibility and potential of using neurophysiological measures, not only for implicit relevance feedback but also for other psychophysiological states, including positive or negative valence. This could provide yet another research opportunity; namely, considering the affective states of users in recommendations.

Ethical aspects and ownership of data. This research poses important ethical issues both in conducting user studies and in developing technologies that can potentially be exploited against the user. Our study adhered to strict ethical standards and received ethical approval. However, we also see the opportunity raised by this work to support the recent MyData movement and related

research to empower individuals by improving their right to self-determination regarding their personal data and providing tools to manage it [1, 72]. The research agenda could consider ways in which users could be provided alternative recommendations to commercial ones or a way to assess the suitability or quality of recommendations.

APPENDIX

The following likelihood functions model the three possible learning signals from the user – namely, newly generated context, explicit feedback on a document, which is linked to a context, and explicit feedback on an entity:

$$f^{n_C} \sim \mathrm{N}(x_n^C \theta, \sigma_{n_C}^2),$$
$$f_j^{e_C} \sim \mathrm{N}(x_j^C \theta, \sigma_{e_C}^2),$$
$$f_i^{e_E} \sim \mathrm{N}(x_i^E \theta, \sigma_{e_E}^2).$$

Here, the $f^{\cdot\cdot}$ are the learning signals (feedback values), the $x.$s are the projected feature vectors, and the σ^2 denote the feedback noises. We distinguish entities and contexts by letters E and C, and the new observing context and user's explicit feedback by letters n and e, respectively. Following Equations (1) and (3), x_j^C is the jth row of $\hat{X}^\top X W_K$, x_i^E is the ith row of $X W_K$, and $x_n^C = \hat{v}^\top X W_K$, where $\hat{v}$ is the normalized (sums up to one) feature vector of the new context (see 3.3.3). To complete the probabilistic model, we put a Gaussian prior distribution on θ as $\theta \sim \mathrm{N}(0, \sigma_\theta^2 \mathbb{I})$. The posterior of this Bayesian regression has a closed form solution as

$$p(\theta \mid M_{n_C}, F_{n_C}, M_{e_C}, F_{e_C}, M_{e_E}, F_{e_E}) = \mathrm{N}(\theta \mid \mu, \Sigma), where$$
$$\Sigma^{-1} = \sigma_\theta^{-2} \mathbb{I} + \sigma_{n_C}^{-2} M_{n_C}^\top M_{n_C} + \sigma_{e_C}^{-2} M_{e_C}^\top M_{e_C} + \sigma_{e_E}^{-2} M_{e_E}^\top M_{e_E},$$
$$\mu = \Sigma(\sigma_{n_C}^{-2} M_{n_C}^\top F_{n_C} + \sigma_{e_C}^{-2} M_{e_C}^\top F_{e_C} + \sigma_{e_E}^{-2} M_{e_E}^\top F_{e_E}).$$

Here, the $F.$ are the vectors of observed learning signals and the $M.$ are the design matrices (that is, feature vectors corresponding to the learning signals) from different sources.

To tune the hyperparameters of the model, we performed a sensitivity analysis on a pilot user in multiple tasks that the participant recorded during the data collection phase. The performance was not sensible to the value of variance parameters. These parameters were fixed as $\sigma_{e_E}^2 = \sigma_{e_C}^2 = 0.01$, $\sigma_{n_C}^2 = 0.05$, $\sigma_\theta^2 = 0.1$. The dimension of the latent space was set as $K = 100$. Positive learning signal was coded as value 1 and we considered the observed contexts in approximately the last minute of the interaction as the new contexts that are relevant to the user interest.

PRIOR PUBLICATION STATEMENT

This contribution is original and has not been previously published. All text, system and relative code, user study and analysis contained in this article is unpublished and there are no concurrent submissions.

REFERENCES

[1] MyData. 2019. MyData Declaration 2019. Retrieved March 10, 2020 from https://mydata.org/declaration/.
[2] J. E. Allen, Curry I. Guinn, and Eric Horvtz. 1999. Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications* 14, 5 (1999), 14–23.
[3] Salvatore Andolina, Khalil Klouche, Jaakko Peltonen, Mohammad Hoque, Tuukka Ruotsalo, Diogo Cabral, Arto Klami, Dorota Głowacka, Patrik Floréen, and Giulio Jacucci. 2015. IntentStreams: Smart parallel search streams for branching exploratory search. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, New York, NY, 300–305. DOI:https://doi.org/10.1145/2678025.2701401

[4] Salvatore Andolina, Khalil Klouche, Tuukka Ruotsalo, Patrik Floréen, and Giulio Jacucci. 2018. Querytogether: En-
     abling entity-centric exploration in multi-device collaborative search. *Information Processing & Management* 54, 6
     (2018), 1182–1202.

[5] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and
     Giulio Jacucci. 2018. Investigating proactive search support in conversations. In *Proceedings of the 2018 Designing
     Interactive Systems Conference.* ACM, New York, NY, 1295–1307. DOI : https://doi.org/10.1145/3196709.3196734

[6] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and
     Giulio Jacucci. 2018. SearchBot: Supporting voice conversations with proactive search. In *Proceedings of the Compan-
     ion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing.* 9–12.

[7] Apple. 2021. About Siri Suggestions on iPhone. Retrieved from https://support.apple.com/en-gb/guide/iphone/
     iph6f94af287/ios.

[8] Krisztian Balog, Filip Radlinski, and Shushan Arakelyan. 2019. Transparent, scrutable and explainable user models
     for personalized recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and
     Development in Information Retrieval.* 265–274.

[9] Nikola Banovic, Christina Brant, Jennifer Mankoff, and Anind Dey. 2014. ProactiveTasks: The short of mobile device
     use sessions. In *Proceedings of the 16th International Conference on Human–Computer Interaction with Mobile Devices
     & Services.* ACM, New York, NY, 243–252. DOI : https://doi.org/10.1145/2628363.2628380

[10] Lingfeng Bao, Deheng Ye, Zhenchang Xing, Xin Xia, and Xinyu Wang. 2015. Activityspace: a remembrance frame-
     work to support interapplication information needs. In *Proceedings of the 2015 30th IEEE/ACM International Confer-
     ence on Automated Software Engineering.* IEEE, 864–869.

[11] Ralf Bender and Stefan Lange. 2001. Adjusting for multiple testing–when and how? *Journal of Clinical Epidemiology*
     54, 4 (2001), 343–349.

[12] Paul N. Bennett, Ryen W. White, Wei Chu, Susan T. Dumais, Peter Bailey, Fedor Borisyuk, and Xiaoyuan Cui. 2012.
     Modeling the impact of short- and long-term behavior on search personalization. In *Proceedings of the 35th Interna-
     tional ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, New York, NY, 185–194.
     DOI : https://doi.org/10.1145/2348283.2348312

[13] Ofer Bergman, Ruth Beyth-Marom, and Rafi Nachmias. 2006. The project fragmentation problem in personal infor-
     mation management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, New
     York, NY, 271–274. DOI : https://doi.org/10.1145/1124772.1124813

[14] Ofer Bergman and Steve Whittaker. 2016. *The Science of Managing Our Digital Stuff* (1st ed.). The MIT Press.

[15] Roi Blanco, Berkant Barla Cambazoglu, Peter Mika, and Nicolas Torzec. 2013. Entity recommendations in web search.
     In *Proceedings of the International Semantic Web Conference.* Springer, 33–48.

[16] Chris Buckley and Ellen M. Voorhees. 2004. Retrieval evaluation with incomplete information. In *Proceedings of the
     27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* 25–32.

[17] Jay Budzik, Kristian J. Hammond, and Larry Birnbaum. 2001. Information access in context. *Knowledge-Based Systems*
     14, 1–2 (2001), 37–53. DOI : https://doi.org/10.1016/S0950-7051(00)00105-2

[18] Georg Buscher, Andreas Dengel, Ralf Biedert, and Ludger V. Elst. 2012. Attentive documents: Eye tracking as implicit
     feedback for information retrieval and beyond. *ACM Transactions on Interactive Intelligent Systems* 1, 2 (2012), 1–30.

[19] Huanhuan Cao, Daxin Jiang, Jian Pei, Qi He, Zhen Liao, Enhong Chen, and Hang Li. 2008. Context-aware query sug-
     gestion by mining click-through and session data. In *Proceedings of the 14th ACM SIGKDD International Conference on
     Knowledge Discovery and Data Mining.* ACM, New York, NY, 875–883. DOI : https://doi.org/10.1145/1401890.1401995

[20] Shiyu Chang, Yang Zhang, Jiliang Tang, Dawei Yin, Yi Chang, Mark A. Hasegawa-Johnson, and Thomas S. Huang.
     2017. Streaming recommender systems. In *Proceedings of the 26th International Conference on World Wide Web.* 381–
     389.

[21] Amir Chaudhry, Jon Crowcroft, Heidi Howard, Anil Madhavapeddy, Richard Mortier, Hamed Haddadi, and Derek
     McAuley. 2015. Personal data: thinking inside the box. In *Proceedings of the 5th Decennial Aarhus Conference on
     Critical Alternatives.* Aarhus University Press, 29–32. DOI : https://doi.org/10.7146/aahcc.v1i1.21312

[22] Sunny Consolvo, David W. McDonald, Tammy Toscos, Mike Y. Chen, Jon Froehlich, Beverly Harrison, Predrag Klas-
     nja, Anthony LaMarca, Louis LeGrand, Ryan Libby, Ian Smith, and James A. Landay. 2008. Activity sensing in the
     wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*
     ACM, New York, NY, 1797–1806. DOI : https://doi.org/10.1145/1357054.1357335

[23] Pedram Daee, Joel Pyykkö, Dorota Glowacka, and Samuel Kaski. 2016. Interactive intent modeling from multiple
     feedback domains. In *Proceedings of the 21st International Conference on Intelligent User Interfaces.* ACM, New York,
     NY, 71–75. DOI : https://doi.org/10.1145/2856767.2856803

[24] Yves-Alexandre de Montjoye, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland. 2014. openPDS: protecting
     the privacy of metadata through SafeAnswers. *PloS One* 9, 7 (2014), e98790. DOI : https://doi.org/10.1371/journal.pone.
     0098790

[25] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41, 6 (1990), 391–407. DOI : https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

[26] David Donoho and Jared Tanner. 2009. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences* 367, 1906 (2009), 4273–4293. DOI : https://doi.org/10.1098/rsta.2009.0152

[27] Pierre Dragicevic. 2016. Fair statistical communication in HCI. In *Modern statistical methods for HCI*. Springer, 291–330.

[28] Susan Dumais, Edward Cutrell, Raman Sarin, and Eric Horvitz. 2004. Implicit queries (IQ) for contextualized search. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 594–594. DOI : https://doi.org/10.1145/1008992.1009137

[29] Desmond Elliott and Joemon M. Jose. 2009. A proactive personalised retrieval system. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. ACM, New York, NY, 1935–1938. DOI : https://doi.org/10.1145/1645953.1646269

[30] Manuel J. A. Eugster, Tuukka Ruotsalo, Michiel M. Spapé, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. 2016. Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. *Scientific Reports* 6, 1(2016), 38580.

[31] Denzil Ferreira, Anind K. Dey, and Vassilis Kostakos. 2011. Understanding human-smartphone concerns: A study of battery life. In *Pervasive Computing*, Kent Lyons, Jeffrey Hightower, and Elaine M. Huang (Eds.). Springer, Berlin, 19–33.

[32] James Fogarty, Carolyn Au, and Scott E. Hudson. 2006. Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*. New York, NY, 91–100. DOI : https://doi.org/10.1145/1166253.1166269

[33] Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, and James A. Landay. 2007. MyExperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services*. ACM, New York, NY, 57–70. DOI : https://doi.org/10.1145/1247660.1247670

[34] Jeremy Goecks and Jude Shavlik. 2000. Learning users' interests by unobtrusively observing their normal behavior. In *Proceedings of the 5th International Conference on Intelligent User Interfaces*. ACM, New York, NY, 129–132. DOI : https://doi.org/10.1145/325737.325806

[35] Greg Guest. 2014. *Public health research methods*. Sage Publications.

[36] Ramanathan Guha, Vineet Gupta, Vivek Raghunathan, and Ramakrishnan Srikant. 2015. User modeling for a personal assistant. In *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*. ACM, New York, NY, 275–284. DOI : https://doi.org/10.1145/2684822.2685309

[37] Jacek Gwizdka, Yashar Moshfeghi, and Max L. Wilson. 2019. Introduction to the special issue on neuro-information science. *Journal of the Association for Information Science and Technology* 70, 9 (2019), 911–916.

[38] Karl Gyllstrom and Craig Soules. 2008. Seeing is retrieving: building information context from what the user sees. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*. ACM, New York, NY, 189–198. DOI : https://doi.org/10.1145/1378773.1378798

[39] Negar Hariri, Bamshad Mobasher, and Robin Burke. 2014. Context adaptation in interactive recommender systems. In *Proceedings of the 8th ACM Conference on Recommender Systems*. ACM, New York, NY, 41–48. DOI : https://doi.org/10.1145/2645710.2645753

[40] Monika Henzinger, Bay-Wei Chang, Brian Milch, and Sergey Brin. 2003. Query-free news search. In *Proceedings of the 12th International Conference on World Wide Web*. ACM, New York, NY, 1–10. DOI : https://doi.org/10.1145/775152.775154

[41] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22, 1 (Jan. 2004), 5–53. DOI : https://doi.org/10.1145/963770.963772

[42] Shamsi T. Iqbal and Brian P. Bailey. 2011. Oasis: a framework for linking notification delivery to the perceptual structure of goal-directed tasks. *ACM Transactions on Computer-Human Interaction* 17, 4 (Dec. 2011), 28 pages, Article 15. DOI : https://doi.org/10.1145/1879831.1879833

[43] Giulio Jacucci, Oswald Barral, Pedram Daee, Markus Wenzel, Baris Serim, Tuukka Ruotsalo, Patrik Pluchino, Jonathan Freeman, Luciano Gamberini, Samuel Kaski, et al. 2019. Integrating neurophysiologic relevance feedback in intent modeling for information retrieval. *Journal of the Association for Information Science and Technology* 70, 9 (2019), 917–930.

[44] Dietmar Jannach, Paul Resnick, Alexander Tuzhilin, and Markus Zanker. 2016. Recommender systems-beyond matrix completion. *Communications of the ACM* 59, 11 (2016), 94–102. DOI:https://doi.org/10.1145/2891406

[45] Dietmar Jannach, Oren Sar Shalom, and Joseph A. Konstan. 2019. Towards more impactful recommender systems research. *CEUR Workshop Proceedings* 2462 (2019), 15–17.

[46] Kalervo Järvelin and Peter Ingwersen. 2004. Information seeking research needs extension towards tasks and technology. *Information Research: An International Electronic Journal* 10, 1 (2004), n1.

[47] Kalervo Järvelin and Jaana Kekäläinen. 2000. IR evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, New York, NY, 41–48. DOI:https://doi.org/10.1145/345508.345545

[48] Kalervo Järvelin, Pertti Vakkari, Paavo Arvola, Feza Baskaya, Anni Järvelin, Jaana Kekäläinen, Heikki Keskustalo, Sanna Kumpulainen, Miamaria Saastamoinen, Reijo Savolainen, and Eero Sormunen. 2015. Task-based information interaction evaluation: the viewpoint of program theory. *ACM Transactions on Informations Systems* 33, 1 (March 2015), 30 pages, Article 3. DOI:https://doi.org/10.1145/2699660

[49] Tero Jokela, Jarno Ojala, and Thomas Olsson. 2015. A diary study on combining multiple information devices in everyday activities and tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* 3903–3912.

[50] William Jones, Susan Dumais, and Harry Bruce. 2002. Once found, what then? A study of "keeping" behaviors in the personal use of Web information. *Proceedings of the American Society for Information Science and Technology* 39, 1 (2002), 391–402. DOI:https://doi.org/10.1002/meet.1450390143 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/meet.1450390143.

[51] Evangelos Kanoulas, Ben Carterette, Paul D. Clough, and Mark Sanderson. 2011. Evaluating multi-query sessions. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, New York, NY, 1053–1062. DOI:https://doi.org/10.1145/2009916.2010056

[52] Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. 2015. Designing for exploratory search on touch devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, New York, NY, 4189–4198. DOI:https://doi.org/10.1145/2702123.2702489

[53] Khalil Klouche, Tuukka Ruotsalo, and Giulio Jacucci. 2018. From hyperlinks to hypercues: Entity-based affordances for fluid information exploration. In *Proceedings of the 2018 designing interactive systems conference.* 401–411.

[54] Khalil Klouche, Tuukka Ruotsalo, Luana Micallef, Salvatore Andolina, and Giulio Jacucci. 2017. Visual re-ranking for multi-aspect information retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval.* ACM, New York, NY, 57–66. DOI:https://doi.org/10.1145/3020165.3020174

[55] Weize Kong, Rui Li, Jie Luo, Aston Zhang, Yi Chang, and James Allan. 2015. Predicting search intent based on pre-search context. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, New York, NY, 503–512. DOI:https://doi.org/10.1145/2766462.2767757

[56] Markus Koskela, Petri Luukkonen, Tuukka Ruotsalo, Mats Sjöberg, and Patrik Floréen. 2018. Proactive information retrieval by capturing search intent from primary task context. *ACM Transactions on Interactive Intelligent Systems* 8, 3, Article 20 (July 2018), 25 pages. DOI:https://doi.org/10.1145/3150975

[57] Michael G. Lamming and William M. Newman. 1992. Activity-based information retrieval: technology in support of personal memory. In *Proceedings of the IFIP 12th World Computer Congress on Personal Computers and Intelligent Systems - Information Processing'92.* Vol. 3. North-Holland Publishing Co., 68–81.

[58] Jane Li, Scott Huffman, and Akihito Tokuda. 2009. Good abandonment in mobile and pc internet search. In *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, New York, NY, 43–50. DOI:https://doi.org/10.1145/1571941.1571951

[59] Henry Lieberman. 1995. Letizia: an agent that assists web browsing. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence.* Vol. 1. Morgan Kaufmann Publishers Inc., San Francisco, CA, 924–929. Retrieved from http://dl.acm.org/citation.cfm?id=1625855.1625975.

[60] Daniel J. Liebling, Paul N. Bennett, and Ryen W. White. 2012. Anticipatory search: using context to initiate search. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval.* ACM, New York, NY, 1035–1036. DOI:https://doi.org/10.1145/2348283.2348456

[61] Yefeng Liu, Darren Edge, and Koji Yatani. 2013. SidePoint: A peripheral knowledge panel for presentation slide authoring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, 681–684. DOI:https://doi.org/10.1145/2470654.2470750

[62] Avishay Livne, Vivek Gokuladas, Jaime Teevan, Susan T. Dumais, and Eytan Adar. 2014. CiteSight: Supporting contextual citation recommendation using differential search. In *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval.* ACM, New York, NY, 807–816. DOI:https://doi.org/10.1145/2600428.2609585

[63] Tariq Mahmood and Francesco Ricci. 2007. Learning and adaptivity in interactive recommender systems. In *Proceedings of the 9th International Conference on Electronic Commerce*. ACM, New York, NY, 75–84. DOI : https://doi.org/10.1145/1282100.1282114

[64] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Communications of the ACM* 49, 4 (April 2006), 41–46. DOI : https://doi.org/10.1145/1121949.1121979

[65] Sean M. McNee, John Riedl, and Joseph A. Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *Proceedings of the CHI'06 Extended Abstracts on Human Factors in Computing Systems*. ACM, New York, NY, 1097–1101. DOI : https://doi.org/10.1145/1125451.1125659

[66] Massimo Melucci. 2012. *Contextual Search*. Now Publishers Inc., Hanover, MA.

[67] Iris Miliaraki, Roi Blanco, and Mounia Lalmas. 2015. From "Selena Gomez" to "Marlon Brando": Understanding explorative entity search. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 765–775. DOI : https://doi.org/10.1145/2736277.2741284

[68] Tadashi Okoshi, Julian Ramos, Hiroki Nozaki, Jin Nakazawa, Anind K. Dey, and Hideyuki Tokuda. 2015. Reducing users' perceived mental effort due to interruptive notifications in multi-device mobile environments. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, NY, 475–486. DOI : https://doi.org/10.1145/2750858.2807517

[69] Valeria Orso, Tuukka Ruotsalo, Jukka Leino, Luciano Gamberini, and Giulio Jacucci. 2017. Overlaying social information: the effects on users' search and information-selection behavior. *Information Processing & Management* 53, 6 (2017), 1269–1286.

[70] Georgios Th Papadopoulos, Konstantinos C. Apostolakis, and Petros Daras. 2013. Gaze-based relevance feedback for realizing region-based image retrieval. *IEEE Transactions on Multimedia* 16, 2 (2013), 440–454.

[71] Marius Pasca. 2004. Acquisition of categorized named entities for web search. In *Proceedings of the t13th ACM International Conference on Information and Knowledge Management*. 137–145.

[72] Antti Poikola, Kai Kuikkaniemi, and Harri Honko. 2015. Mydata a nordic model for human-centered personal data management and processing. Retrieved from http://urn.fi/URN:ISBN:978-952-243-455-5.

[73] Jeffrey Pound, Peter Mika, and Hugo Zaragoza. 2010. Ad-hoc object retrieval in the web of data. In *Proceedings of the 19th International Conference on World Wide Web*. ACM, New York, NY, 771–780. DOI : https://doi.org/10.1145/1772690.1772769

[74] Tye Rattenbury and John Canny. 2007. CAAD: an automatic task support system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 687–696. DOI : https://doi.org/10.1145/1240624.1240731

[75] Radim Řehůřek and Petr Sojka. 2010. Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, 45–50. DOI : http://is.muni.cz/publication/884893/en.

[76] Bradley Rhodes and Thad Starner. 1996. Remembrance Agent: A continuously running automated information retrieval system. In *Proceedings of the 1st International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology*. 487–495.

[77] B. J. Rhodes and P. Maes. 2000. Just-in-time information retrieval agents. *IBM Systems Journal* 39, 3-4 (July 2000), 685–704. DOI : https://doi.org/10.1147/sj.393.0685

[78] Anne Spencer Ross, Xiaoyi Zhang, James Fogarty, and Jacob O. Wobbrock. 2020. An epidemiology-inspired large-scale analysis of android app accessibility. *ACM Transactions on Accessible Computing* 13, 1 (2020), 1–36.

[79] Yong Rui, T. S. Huang, M. Ortega, and S. Mehrotra. 1998. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology* 8, 5 (Sept. 1998), 644–655. DOI : https://doi.org/10.1109/76.718510

[80] Tuukka Ruotsalo, Khalil Klouche, Diogo Cabral, Salvatore Andolina, and Giulio Jacucci. 2016. Flexible entity search on surfaces. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*. ACM, New York, NY, 175–179. DOI : https://doi.org/10.1145/3012709.3012732

[81] Tuukka Ruotsalo, Jaakko Peltonen, Manuel J. A. Eugster, Dorota Głowacka, Patrik Floréen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. 2018. Interactive intent modeling for exploratory search. *ACM Transactions on Information Systems* 36, 4 (2018), 1–46.

[82] Tuukka Ruotsalo, Jaakko Peltonen, Manuel J. A. Eugster, Dorota Glowacka, Patrik Floréen, Petri Myllymäki, Giulio Jacucci, and Samuel Kaski. 2018. Interactive intent modeling for exploratory search. *ACM Transactions on Information Systems* 36, 4, Article 44 (Oct. 2018), 46 pages. DOI : https://doi.org/10.1145/3231593

[83] Tara Safavi, Adam Fourney, Robert Sim, Marcin Juraszek, Shane Williams, Ned Friend, Danai Koutra, and Paul N. Bennett. 2020. Toward activity discovery in the personal web. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. ACM, New York, NY, 492–500. DOI : https://doi.org/10.1145/3336191.3371828

[84] Gerard Salton and Chris Buckley. 1990. Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science* 41, 4 (1990), 288–297. DOI: https://doi.org/10.1002/(SICI)1097-4571(199006)41: 4<288::AID-ASI8>3.0.CO;2-H

[85] Procheta Sen, Debasis Ganguly, and Gareth Jones. 2018. Procrastination is the thief of time: evaluating the effectiveness of proactive search systems. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, New York, NY, 1157–1160. DOI: https://doi.org/10.1145/3209978.3210114

[86] Chirag Shah. 2018. Information fostering - being proactive with information seeking and retrieval: perspective paper. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. ACM, New York, NY, 62–71. DOI: https://doi.org/10.1145/3176349.3176389

[87] Donghee Shin. 2020. How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior* 109 (2020), 106344.

[88] Milad Shokouhi and Qi Guo. 2015. From queries to cards: Re-ranking proactive card recommendations based on reactive search history. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 695–704. DOI: https://doi.org/10.1145/2766462.2767705

[89] Mats Sjöberg, Hung-Han Chen, Patrik Floréen, Markus Koskela, Kai Kuikkaniemi, Tuukka Lehtiniemi, and Jaakko Peltonen. 2017. Digital me: Controlling and making sense of my digital footprint. In *Proceedings of the Symbiotic Interaction*. L. Gamberini, A. Spagnolli, G. Jacucci, B. Blankertz, and J. Freeman (EDs.), Lecture Notes in Computer Science, Vol. 9961, Springer, 155–167. DOI: https://doi.org/10.1007/978-3-319-57753-1_14

[90] Timothy Sohn, Kevin A. Li, William G. Griswold, and James D. Hollan. 2008. A diary study of mobile information needs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 433–442. DOI: https://doi.org/10.1145/1357054.1357125

[91] Yang Song and Qi Guo. 2016. Query-less: Predicting task repetition for nextgen proactive search and recommendation engines. In *Proceedings of the 25th International Conference on World Wide Web*. 543–553.

[92] Yang Song and Qi Guo. 2016. Query-Less: Predicting task repetition for nextgen proactive search and recommendation engines. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, 543–553. DOI: https://doi.org/10.1145/2872427.2883020

[93] David Sontag, Kevyn Collins-Thompson, Paul N. Bennett, Ryen W. White, Susan Dumais, and Bodo Billerbeck. 2012. Probabilistic models for personalizing web search. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*. ACM, New York, NY, 433–442. DOI: https://doi.org/10.1145/2124295.2124348

[94] Jaime Teevan. 2008. How people recall, recognize, and reuse search results. *ACM Transactions on Information Systems* 26, 4 (Oct. 2008), 27 pages, Article 19. DOI: https://doi.org/10.1145/1402256.1402258

[95] Jaime Teevan, Christine Alvarado, Mark S. Ackerman, and David R. Karger. 2004. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 415–422. DOI: https://doi.org/10.1145/985692.985745

[96] Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2010. Potential for personalization. *ACM Transactions on Computer-Human Interaction* 17, 1, Article 4 (April 2010), 31 pages. DOI: https://doi.org/10.1145/1721831.1721835

[97] Thi Ngoc Trang Tran, Alexander Felfernig, Christoph Trattner, and Andreas Holzinger. 2020. Recommender systems in the healthcare domain: state-of-the-art and research issues. *Journal of Intelligent Information Systems* (2020), 1–31.

[98] Christophe Van Gysel, Bhaskar Mitra, Matteo Venanzi, Roy Rosemarin, Grzegorz Kukla, Piotr Grudzien, and Nicola Cancedda. 2017. Reply with: Proactive recommendation of email attachments. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, New York, NY, 327–336. DOI: https://doi.org/10.1145/3132847.3132979

[99] Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W. White. 2019. Bridging screen readers and voice assistants for enhanced eyes-free web search. In *Proceedings of the World Wide Web Conference*. ACM, New York, NY, 3590–3594. DOI: https://doi.org/10.1145/3308558.3314136

[100] Tung Vuong, Salvatore Andolina, Giulio Jacucci, and Tuukka Ruotsalo. 2021. Spoken conversational context improves query auto-completion in web search. *ACM Transaction on Information Systems* 39, 3, Article 31 (April 2021), 32 pages. DOI: https://doi.org/10.1145/3447875

[101] Tung Vuong, Giulio Jacucci, and Tuukka Ruotsalo. 2017. Proactive information retrieval via screen surveillance. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 1313–1316. DOI: https://doi.org/10.1145/3077136.3084151

[102] Tung Vuong, Giulio Jacucci, and Tuukka Ruotsalo. 2017. Watching inside the screen: Digital activity monitoring for task recognition and proactive information retrieval. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3, Article 109 (Sept. 2017), 23 pages. DOI: https://doi.org/10.1145/3130974

[103] Ryen White, Ian Ruthven, and Joemon M. Jose. 2002. The use of implicit evidence for relevance feedback in web retrieval. In *Proceedings of the 24th BCS-IRSG European Colloquium on IR Research: Advances in Information Retrieval*. Springer-Verlag, Berlin, 93–109. Retrieved from http://dl.acm.org/citation.cfm?id=645319.757469.

[104] Ryen W. White, Paul N. Bennett, and Susan T. Dumais. 2010. Predicting short-term interests using activity-based search context. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*. ACM, New York, NY, 1009–1018. DOI : https://doi.org/10.1145/1871437.1871565

[105] Ryen W. White and Gary Marchionini. 2007. Examining the effectiveness of real-time query expansion. *Information Processing & Management* 43, 3 (May 2007), 685–704. DOI : https://doi.org/10.1016/j.ipm.2006.06.005

[106] Shunguo Yan and P.G . Ramachandran. 2019. The current status of accessibility in mobile apps. *ACM Transactions on Accessible Computing* 12, 1 (2019), 1–31.

[107] Dingqi Yang, Daqing Zhang, Longbiao Chen, and Bingqing Qu. 2015. NationTelescope: Monitoring and visualizing large-scale collective behavior in LBSNs. *Journal of Network and Computer Applications* 55 (2015), 170–180. DOI : https://doi.org/10.1016/j.jnca.2015.05.010

[108] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. 2014. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *Proceedings of the 37th International ACM SIGIR Conference on Research & development in Information Retrieval*. 83–92.

[109] Qian Zhao, Paul N. Bennett, Adam Fourney, Anne Loomis Thompson, Shane Williams, Adam D. Troy, and Susan T. Dumais. 2018. Calendar-aware proactive email recommendation. In *Proceedings of the 41st International ACM SIGIR Conference on Research &Development in Information Retrieval*. ACM, New York, NY, 655–664. DOI : https://doi.org/10.1145/3209978.3210001

[110] Zack Zhu, Ulf Blanke, Alberto Calatroni, and Gerhard Tröster. 2013. Human activity recognition using social media data. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*. ACM, New York, NY . DOI : https://doi.org/10.1145/2541831.2541852