

---

# Naturalistic Digital Task Modeling for Personal Information Assistance via Continuous Screen Monitoring

**Tung Vuong**

Dept. of Computer Science,  
University of Helsinki  
vuong@cs.helsinki.fi

**Giulio Jacucci**

Dept. of Computer Science,  
University of Helsinki  
giulio.jacucci@helsinki.fi

**Tuukka Ruotsalo**

Dept. of Computer Science,  
University of Helsinki  
tuukka.ruotsalo@helsinki.fi

---

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

ACM.

*UbiComp/ISWC'18 Adjunct.*, October 8–12, 2018, Singapore, Singapore

ACM 978-1-4503-5966-5/18/10.

<https://doi.org/10.1145/3267305.3274130>

**Abstract**

Despite the growth of personal digital information use, both in scale and application diversity, conventional user models are still reliant on limited user input data to improve a variety of services for specific applications and tasks. This trend toward increased application diversity renders it difficult for a system to generate inferences about a user's evolving interests and naturalistic tasks in real-life settings. This workshop paper introduces a novel approach, aimed at training a user model to recognize real-life tasks on the basis of naturalistic user behavioral data via continuous screen monitoring. The resulting task model could be used in real-life settings for personal information assistance, which proactively retrieves useful documents and resources for the user, on a personal computer, with respect to the task context and information demand.

**Author Keywords**

Digital task modeling; personal information assistance; screen monitoring

**ACM Classification Keywords**

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; See [<http://acm.org/about/class/1998/>]: for full list of ACM classifiers. This section is required.

## Introduction

Personal digital activities, for both leisure and work, may be characterized as non-deterministic, where some steps of the activity may be done on various platforms or applications. For instance, while writing documents, using online services (e.g., to contact co-authors), or searching online searchable archives (e.g., to check facts, terminologies, or locate related works) have become integral tools for the writing work process. This activity is communicated via a computer screen and manifested as the applications we use, the content we examine or produce, and the entities (e.g., other people) with whom we relate and communicate.

It has been a long-standing aim to automatically model users, based on these digital traces, to generate inferences about specific users and their tasks. The key factor driving this process is the use of historical observations regarding user behavior to improve a variety of personal assistant services, such as personalizing and ranking search results [14], or recommending entities of interest [16].

Conventionally, user models are constructed from predefined interaction logs, such as queries, the links or menu options clicked, and items browsed, and they were specific to certain tasks and applications, such as news browsing [2], personal information management tools [5], or search rankings [14]. However, while the logs can, to some extent, capture user behavior, but they can not capture the unpredictable nature of real-life activities, such as task switching, application switching, and task sessions that are split into several chronological parts.

In real life, everyday digital activities are components of naturalistic tasks that are personalized and performed by users, by means of digital information resources and tools, such as documents, applications, scripts, graphics, messages, and other computing services. In their simplest form,

naturalistic tasks can be regarded as concrete processes, engaged in and meaningfully labeled by users, the associated activities of which encompass a broad variety of applications and share a common topical context. For example, a task could be one part of a wider ongoing project, related to many documents that were read and written by the user via distinct applications, or through digital communication with other people.

Naturalistic digital tasks and their real-life contexts are, however, hard to reconstruct via laboratory studies or log analysis. We are beginning to see a paradigm shift within user modeling and information retrieval wherein research is, increasingly, leaving the safety of controlled studies or the constraints of predetermined log analyses, in favor of moving into "the wild" [11]. This relatively recent research orientation allows for exploration of the multiple modes of understanding technological interventions in the real-life situations where humans naturally use them.

The overall aim of this research is to propose an approach whereby we can train a user task model by exposing it to naturalistic user behavioral data that could be extracted from continuous screen monitoring. Our approach employs a computer vision software that monitors every digital activity occurring inside the screen. The resulting data is then fed into Latent Semantic Analysis (LSA) [3] which is an unsupervised topic modeling for detecting tasks in which the user was engaged during screen monitoring. The model could be used to provide personal information assistance by recovering tasks and related documents in real-time. The challenge, relative to previous approaches, is that our approach does not limit the learning data to structured information from specific systems or separate user input, but attempts to learn from any screen frame that has been displayed to a user, from any application.

## Previous Work

In earlier work, we used interactive intent modeling to assist users throughout their search tasks by combining intent modeling and visual interfaces [9, 10]. However, the intent model required the users to provide explicit feedback and did not support proactivity feature. Our most recent work investigated a proactive search agent that could support natural spoken conversations by augmenting the conversations with additional information [1]. Both have shown the effectiveness of search systems in assisting users performing their tasks, but they were lab-based studies and tasks were designed to answer specific research questions.

The problem of task recognition has recently attracted a great deal attention throughout web search community [6, 8]. Recognizing tasks is important to various competencies; it permits a system to determine user satisfaction with search results [11], predict future intents [9, 10], and provide proactive assistance [7, 1]. In traditional approaches, the main user signals used to recognize tasks typically rely on search engine query logs [12]. Query logs are often used for within-session learning, to infer the short-term preferences of a particular user [4]. However, web search sessions are merely small subsets of a broader task. Naturalistic tasks, in real life, may be interleaved with several applications that can be used side-by-side and there is no time limit for such tasks.

In summary, prior studies focused on recognizing and supporting tasks that were predetermined for specific studies or to answering specific research questions. Conventionally, the experimenter designed such studies to capture tasks in relatively short sessions that varied from seconds to hours [13]. Some attempted to model long-term user preferences but collected data reliant on limited user input, including clicks and queries [6, 8]. Our approach extends beyond

lab-based and controlled environments, in which tasks are recognized on the basis of naturalistic user behavioral data, by analyzing a 24/7 continuous recording of information changes on the computer screen.

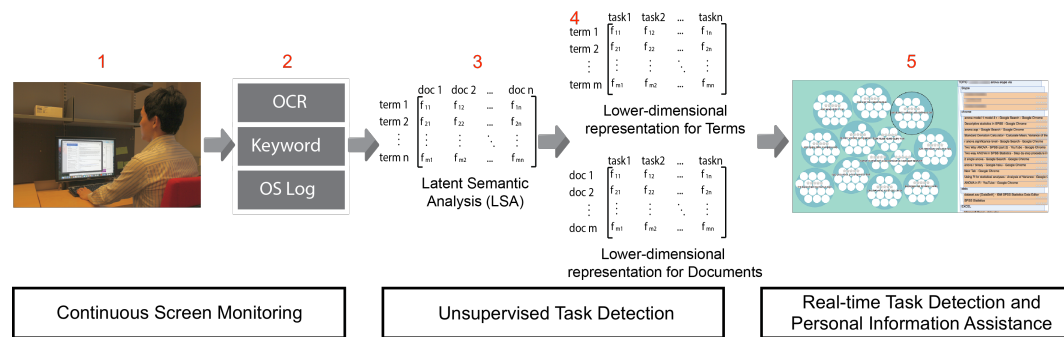
## Naturalistic Digital Task Modeling

To capture naturalistic behavior as accurately and thoroughly as possible, we employed a methodology wherein the participants' computer screens were continuously recorded and all information appearing on their screens during the data recording period was captured. The screen recordings consisted of a collection of screen frames which were then fed into an unsupervised machine learning method designed to detect user tasks.

Figure 1 presents the flow of the data processing and task detection pipeline. First, screen frames of active windows are saved as images at 5-second intervals representing naturalistic user behavioral data. Second, Optical Character Recognition (OCR) software converts those images into texts for task modeling, keywords are extracted for task labeling, and Operating System (OS) data are collected for document retrieval. Third, the text, the keywords and OS data are stored in a vector that represent the document. Fourth, all vectors representing all documents for a particular user are fed into the LSA, which is then run, to find a lower-dimensional representation of this data. LSA produces vectors for tasks, where each vector is linked to documents, and the corresponding input data, such as terms and keywords. Fifth, the resulting task models are used to detect naturalistic tasks in real-time, and proactively provide personal information assistance.

### *Continuous Screen Monitoring*

Unlike other logging methods, screen recording has no limitations, in term of application range or user input, and apart



**Figure 1:** The work-flow of the data processing and task modeling. The digital activity monitoring data were modeled using Latent Semantic Analysis (LSA). LSA learns a latent lower-dimensional representation of the input data. Each dimension in the lower dimensional space can be interpreted as a representation of a task and used to retrieve task-relevant documents and labels.

from audio, it is capable of capturing every input and presentation of content that involves human-computer interaction. We used a logger to record images of active windows at five-second intervals. Further, the logger recorded OS information, such as files and folders associated with the screen frames, and the names of the active windows and timestamps. The logger was developed via two OS-dependent versions: We used Core Graphics API to implement the Mac OS version and Desktop App UI to implement the Windows OS version. Both were native OS libraries that performed identical functions of capturing screen frames and saving them as images. To access to the clean text version of the screen frames, we utilized very accurate OCR engine, Google Vision API <sup>1</sup>.

#### *Unsupervised Task Detection and Labeling*

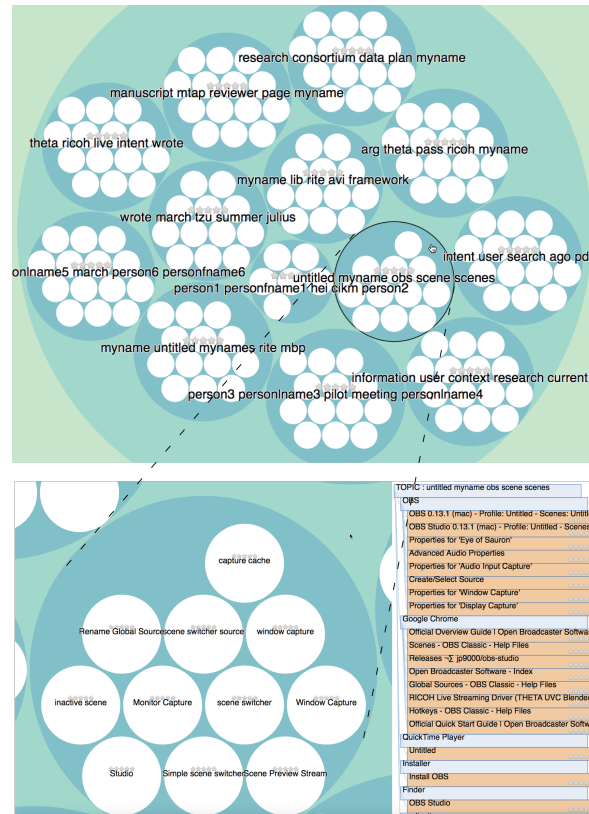
To uncover the task structure in the collected and OCR processed documents of the screen captures, we utilized LSA. LSA learns a latent lower-dimensional representation of the input data. Each dimension in the lower-dimensional space can be interpreted and labeled as a representation

of a task, and used to retrieve task-relevant documents and labels.

The LSA yielded a lower-dimensional representation, but as an unsupervised method, it is not directly interpretable. To facilitate interpretability of the lower-dimensional representation for the participants, we developed an approach to labeling the tasks (i.e. find keyword descriptions for the dimensions in the lower-dimensional output space). A vector-space model of information retrieval with cosine similarity was applied, to rank the keywords extracted in the task detection pipeline. Keywords occurring in the 100 top-ranked documents (OCR processed screen captures) and extracted via the IBM Cloud API <sup>2</sup> were ranked simply on the basis of their term frequency. The term frequency was determined to be a good measure as the keywords were already being extracted via an entity detection process and did not contain general terms. The upper part and left bottom part of the Figure 2 depicts how the tasks are labeled and visualized.

<sup>1</sup><https://cloud.google.com/vision/>

<sup>2</sup><https://www.ibm.com/watson/services/natural-language-understanding/>



**Figure 2:** Overview of how tasks are labeled and task-relevant document retrieval. *Top:* the task view with all detected tasks and the associated labels. Each task is visualized as a circle. *Bottom left:* the real-life user task is "Video capture with Theta". A set of keywords from OCR processed screen captures are used to label the task e.g., video recording features, software menus, and tools used in the task. *Bottom right:* retrieved documents in response to the task were from various applications depending on a user's personal tasks, such as OBS software wherein new scenes or video projects were created, Web browsers by means of which the user looked up Theta-related tutorials and OBS's help information, Quick Time Player to re-play recorded videos, Finder through which the user explore location of OBS application, and so forth.

### Real-time Task Detection and Information Assistance

To apply the resulting task model in real-time task detection scenarios, the output of the screen monitoring system was fed into the LSA model, which produced a prediction of the task that the user was performing. The model was able to correctly classify unseen input resulting from user interactions as a given task in the task model, and proactively retrieve the relevant related labels.

The task modeling and labeling indicate the extent to which the system can accurately detect user tasks and render them interpretable for the user. This does not, however, reflect the model's usefulness for the user. To provide useful assistance to users, we constructed a document-retrieval method that retrieved a ranked list of documents matching each of the tasks. The rationale behind this effort was to make it possible to study whether the task model could be used on a personal computer, to re-find documents that could be useful resources for the user in the task context. The vector-space model of information retrieval with cosine-similarity ranking was used to retrieve and rank the documents represented in the lower-dimensional space. The input was the lower-dimensional task vector, and the model ranked documents, using the data from the original higher-dimensional space. The right bottom part of the Figure 2 depicts the mode of retrieving task-relevant documents.

## Evaluation

### Data Collection

Ten people, from both university and industrial settings, were recruited to participate in the experiment. The participants signed a consent form regarding data usage, privacy, and the experimental procedure. We followed the ethical guidelines laid out by our University<sup>3</sup>. Naturalistic behav-

<sup>3</sup><https://www.helsinki.fi/en/research/research-environment/research-ethics>

ioral data was collected by installing the logger on personal laptops and setting it to run continuously for a duration of 14 days. During the data collection period, participants were advised to avoid stopping the logger unless it was necessary. All screen frames recorded by the logger, including OCR processed texts, extracted keywords, and associated OS logs were stored on the local hard drive of each participant's laptop. Further, participants were asked to keep a diary of their daily tasks. We defined a digital task as any sequencing activity related to a certain topic, spanning time, such as hours, days, or weeks. We intentionally focused on macro tasks that could consist of several microtasks. However, we encouraged participants to use their own conceptual understanding of which activities could be regarded as meaningful tasks.

#### Setup & Results

The participants assessed the correctness of the task models, using a user interface (UI) shown in Figure 2. Task-detection accuracy was computed as the binarized output. Participants were asked to compare task descriptions in diaries to the task labels on UI, and explicitly indicate either (0), no tasks matched for a task or (1), for a correctly formulated task. Overall, the participants reported a total of 119 tasks and an average of 72.27% task-detection accuracy.

For real-time task detection and document retrieval, we asked every participant to select and resume 6 tasks from the diaries. The participants' task was interrupted at 30-second intervals (up to 120 seconds), and we asked them to look at the task detection system, to see if the task was detected correctly; participants assessed the relevance of the retrieved documents using the following scale: (0) not relevant; (1) slightly relevant; (2) moderately relevant; (3) highly relevant; and (4) absolutely relevant. We used relevant score, Normalized Discounted Cumulative Gain

Attempt	1	2	3	4
Score	3.31(0.61)	3.21 (0.67)	3.18 (0.79)	2.67 (0.98)
NDCG	0.94	0.79	0.89	0.98
P@1	1	0.78	0.9	1
P@10	0.95	0.82	0.87	0.92
P@20	0.93	0.89	0.87	0.83

**Table 1:** Document precision at 1, 10, 20 in the real-time task detection and document retrieval experiment. Results are reported with respect to attempts (task interruption on 30 seconds interval).

(NDCG), and precision at N to measure the performance of document retrieval. All measures were computed at every point of interruption (at 30, 60, 90, and 120 seconds). The overall result, after all attempts (mean over the tasks and participants at 120 seconds), shows a rate of task-detection accuracy of 95% (57 out of 60 tasks). Table 1 shows results of all measures for the document retrieval.

#### Implications

Detecting naturalistic tasks sheds lights on information retrieval studies, as it provides opportunities to learn a user task model from a single source of screen monitoring. The model is also useful for detecting tasks in real-time, and enables proactive information retrieval. User tasks, such as work tasks or leisure tasks, are often analyzed and used as an important context, influencing information search behavior [11]. Different search tasks could require more search effort than others.

Given the knowledge of the importance of task factors on search behavioral factors, we can potentially inform the work of designers and developers focused on supporting particular types of tasks in information search. The current "one size fits all" user interaction with search engines may not be optimal across different tasks, and the design of the next generation of information access systems could

benefit from considering whether it makes sense to employ user task contexts to identify, and even predict the need for, specific kinds of search support that might benefit the user. Contextual factors and search effort could be linked to task categories, thereby constituting to a useful step in adapting information retrieval environments.

### Conclusion

In this workshop paper, we proposed a novel approach to detecting user tasks, using real world data to support personal information access. Unlike any prior research of which we are aware of, user behavioral data is collected in a most naturalistic way, by simply recording a computer screen. The resulting task models are promising, not only for detecting naturalistic tasks, but they could also be useful for studying interdependencies between search behavior and the associated real-life tasks. More specifically, in some tasks, the need for information could be triggered by information observed on the screen, whereas other tasks could rely on intrinsically-driven information. Thus, we believe that the unobtrusiveness of personal information assistance system can only be achieved through intelligent actions that coordinate timely recommendations and filter information, pursuant to a user's demand and the information's relevance, with respect to user tasks and goals. Our work could be extended, in the sense that it also accounts for interruption costs based on task characteristics, to maintain the benefits of proactive systems [15].

### Acknowledgements

This research was partially funded by TEKES (Re:Know), the Academy of Finland (278090, 305739), and Nokia Foundation Scholarship.

### REFERENCES

1. 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 (DIS '18)*. ACM, New York, NY, USA, 1295–1307. DOI : <http://dx.doi.org/10.1145/3196709.3196734>
2. Daniel Billsus and Michael J. Pazzani. 2000. User Modeling for Adaptive News Access. *User Modeling and User-Adapted Interaction* 10, 2 (01 Jun 2000), 147–180. DOI : <http://dx.doi.org/10.1023/A:1026501525781>
3. 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 : [http://dx.doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6<391::AID-ASI1>3.0.CO;2-9](http://dx.doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9)
4. Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. 2014. Lessons from the Journey: A Query Log Analysis of Within-session Learning. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)*. ACM, New York, NY, USA, 223–232. DOI : <http://dx.doi.org/10.1145/2556195.2556217>
5. Eric Horvitz, Jack Breese, David Heckerman, David Hovel, and Koos Rommelse. 1998. The Lumière Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI'98)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 256–265. <http://dl.acm.org/citation.cfm?id=2074094>. 2074124
6. Wen Hua, Yangqiu Song, Haixun Wang, and Xiaofang Zhou. 2013. Identifying Users' Topical Tasks in Web

- Search. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM '13)*. ACM, New York, NY, USA, 93–102. DOI : <http://dx.doi.org/10.1145/2433396.2433410>
7. Markus Koskela, Petri Luukkonen, Tuukka Ruotsalo, Mats Sjöberg, and Patrik Floréen. 2018. Proactive Information Retrieval by Capturing Search Intent from Primary Task Context. *ACM Trans. Interact. Intell. Syst.* 8, 3, Article 20 (July 2018), 25 pages. DOI : <http://dx.doi.org/10.1145/3150975>
  8. Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Fabrizio Silvestri, and Gabriele Tolomei. 2011. Identifying Task-based Sessions in Search Engine Query Logs. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM '11)*. ACM, New York, NY, USA, 277–286. DOI : <http://dx.doi.org/10.1145/1935826.1935875>
  9. Tuukka Ruotsalo, Giulio Jacucci, Petri Myllymäki, and Samuel Kaski. 2014. Interactive Intent Modeling: Information Discovery Beyond Search. *Commun. ACM* 58, 1 (Dec. 2014), 86–92. DOI : <http://dx.doi.org/10.1145/2656334>
  10. 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* (2018). Accepted. To appear.
  11. Miimaria Saastamoinen and Kalervo Järvelin. 2016. Queries in authentic work tasks: the effects of task type and complexity. *Journal of Documentation* 72, 6 (2016), 1114–1133. DOI : <http://dx.doi.org/10.1108/JD-09-2015-0119>
  12. Fabrizio Silvestri. 2010. Mining Query Logs: Turning Search Usage Data into Knowledge. *Found. Trends Inf. Retr.* 4, 1&#8212;2 (Jan. 2010), 1–174. DOI : <http://dx.doi.org/10.1561/15000000013>
  13. Bin Tan, Xuehua Shen, and ChengXiang Zhai. 2006. Mining Long-term Search History to Improve Search Accuracy. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '06)*. ACM, New York, NY, USA, 718–723. DOI : <http://dx.doi.org/10.1145/1150402.1150493>
  14. Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2005. Personalizing Search via Automated Analysis of Interests and Activities. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05)*. ACM, New York, NY, USA, 449–456. DOI : <http://dx.doi.org/10.1145/1076034.1076111>
  15. Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Interruptibility Prediction for Ubiquitous Systems: Conventions and New Directions from a Growing Field. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 801–812. DOI : <http://dx.doi.org/10.1145/2750858.2807514>
  16. Jinfeng Zhuang, Tao Mei, Steven C.H. Hoi, Ying-Qing Xu, and Shipeng Li. 2011. When Recommendation Meets Mobile: Contextual and Personalized Recommendation on the Go. In *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11)*. ACM, New York, NY, USA, 153–162. DOI : <http://dx.doi.org/10.1145/2030112.2030134>