

EntityBot: Supporting Everyday Digital Tasks with Entity Recommendations

Tung Vuong*
University of Helsinki
Helsinki, Finland
vuong@cs.helsinki.fi

Salvatore Andolina*
University of Palermo
Palermo, Italy
salvatore.andolina@unipa.it

Giulio Jacucci*
University of Helsinki
Helsinki, Finland
giulio.jacucci@helsinki.fi

Pedram Daei*
Aalto University
Helsinki, Finland
pedram.daei@aalto.fi

Khalil Klouche
University of Helsinki
Helsinki, Finland
khalil.klouche@helsinki.fi

Mats Sjöberg
CSC – IT Center for Science
Espoo, Finland
mats.sjoberg@csc.fi

Tuukka Ruotsalo
University of Helsinki
Helsinki, Finland
University of Copenhagen
Copenhagen, Denmark
tuukka.ruotsalo@helsinki.fi

Samuel Kaski
Aalto University
Helsinki, Finland
University of Manchester
Manchester, UK
samuel.kaski@aalto.fi

ABSTRACT

Everyday digital tasks can highly benefit from systems that recommend the right information to use at the right time. However, existing solutions typically support only specific applications and tasks. In this demo, we showcase EntityBot, a system that captures context across application boundaries and recommends information entities related to the current task. The user’s digital activity is continuously monitored by capturing all content on the computer screen using optical character recognition. This includes all applications and services being used and specific to individuals’ computer usages such as instant messaging, emailing, web browsing, and word processing. A linear model is then applied to detect the user’s task context to retrieve entities such as applications, documents, contact information, and several keywords determining the task. The system has been evaluated with real-world tasks, demonstrating that the recommendation had an impact on the tasks and led to high user satisfaction.

CCS CONCEPTS

• **Information systems**; • **Human-centered computing**;

KEYWORDS

Proactive information retrieval, real-world tasks, user intent modeling

*Tung Vuong, Salvatore Andolina, Giulio Jacucci, and Pedram Daei contributed equally.

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).
RecSys '21, September 27–October 1, 2021, Amsterdam, Netherlands
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8458-2/21/09.
<https://doi.org/10.1145/3460231.3478883>

ACM Reference Format:

Tung Vuong, Salvatore Andolina, Giulio Jacucci, Pedram Daei, Khalil Klouche, Mats Sjöberg, Tuukka Ruotsalo, and Samuel Kaski. 2021. EntityBot: Supporting Everyday Digital Tasks with Entity Recommendations. In *Fifteenth ACM Conference on Recommender Systems (RecSys '21)*, September 27–October 1, 2021, Amsterdam, Netherlands. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3460231.3478883>

1 INTRODUCTION

Many of our everyday digital tasks involve integrating and reusing information that we have created and accessed before [7]. For example, writing a document may involve web searching and pulling together information from several existing sources such as, other documents, emails, spreadsheets, instant messages, data analyses, and other entities. Researchers have studied recommendations offered to users contextually while they are performing specific tasks [1–3, 5, 8, 12, 13]. In [5], contextual text and image queries are performed based on the text written by the user in a word processing application, and, in [13], reference recommendations are shown in a similar scenario. It has been a long standing aim to automatically model users based on digital traces to make inferences about their tasks to improve recommendation services. Conventionally, user models are built from explicit behavioral data [9], such as the queries users submit, links or menu options clicked, or items browsed, and they are specific to certain applications, such as news browsing [4], personal information management tools [10], or search rankings [14]. Such observations of explicit behavioral data are usually specific to an application or system and can harness the data model which is structured based on prior knowledge about the content.

In [11] we introduce EntityBot, a system that supports everyday digital tasks with cross-app rich entity recommendation. EntityBot builds an unsupervised model of a user’s topical activities from digital activity monitoring data. Subsequently, the model is used to detect the user’s context from unseen user activity and automatically

retrieve relevant entities such as people, applications, documents, or keywords based on their current task (see video illustration¹). The EntityBot was evaluated in a study with 13 participants who volunteered for digital activity monitoring for 14 days². The results from the real-time context-aware entity recommendation experiment using this data suggest that the system provides high user satisfaction and usefulness of the recommendation in realistic task scenarios.

2 DEMONSTRATION DESCRIPTION

The demonstration will be based on a predefined user model trained with interactional data regarding two fictitious tasks. During this demonstration, conference participants will receive an explanation of the two tasks and will be invited to choose one task to resume. As the participants start performing actions, such as typing a query on a web browser, the system will respond by showing relevant entity recommendations in real-time.

3 ENTITYBOT COMPONENTS

The system consists of three main components. A digital activity monitoring system 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 (source code on github³). A user interface (UI) presents the list of recommended entities.

3.1 Digital Activity Monitoring

The digital activity monitoring system is comprised of four modules. The first module is a screen monitoring (*SM*) system that captures screenshots of active windows at 2-second intervals. SM is developed into two versions: a Mac OS version and an MS Windows version. We utilized Accessibility API to implement both versions. Both perform an identical function that saves the screenshots of active windows as images. The second module is an optical character recognition (*OCR*) system that detects and extracts text from the screenshots. We utilized Tesseract 4.0⁴ for the OCR system. The third module is Entity Extraction (*EE*) system detects and extracts available entities from the OCR-processed screenshots. We utilized the IBM Bluemix Natural Language Understanding API⁵ to extract two types of entities that included people's names and keywords. The last module is *OS logger* that 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.

3.2 Modeling and Recommendation

The modeling approach receives digital activity monitoring data and prepares recommendations for display on the UI. The user may provide explicit feedback on recommended entities to help the

system update recommendations. In our solution, the relationship between entities is established mainly based on their co-occurrence and partially through temporal closeness. We define a user context at each time step as a vector. User contexts in our setting include OCR-processed screenshots, recorded OS information, and extracted entities from EE systems. Inspired by the bag-of-words model, a context is represented as a bag of individual entities (applications, keywords, named entities, and non-keyword terms). The logged contexts are stored 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. Due to large number of context vectors and thousands of entities, we compress X into lower-dimensional spaces, such that co-occurring entities should get similar representations. For that, we perform truncated singular value decomposition (truncated SVD) on X to get the projection matrix $W_K \in \mathbb{R}^{|C| \times K}$ that enables us to project entities into a latent K -dimensional space. We defined the user interest as a linear model in this latent space,

$$r^E = XW_K\theta, \quad (1)$$

where $r^E \in \mathbb{R}^{|E|}$ is the vector containing relevance of all entities (we use r_i^E to refer to the i^{th} element) and θ is the K -dimensional latent user interest.

We follow the keywords-documents connection idea in [6] 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 as:

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

where r_j^C refers to the relevance of the j^{th} context (with some abuse of notation), and $p_{(i|j)}$ is the likelihood of the i^{th} entity being present in the j^{th} 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 $r^C = \hat{X}^T r^E$. Finally, by using Equation 1 we can directly connect the user interest to contexts

$$r^C = \hat{X}^T XW_K\theta. \quad (3)$$

In the online phase of the study, the recent "contexts" are fed to the model to predict the user interest. In addition, the user can provide explicit feedback on the recommended entities through the user interface. The feedback is 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.

¹https://youtu.be/pCMVpl_T4P0

²Participants could stop/resume the monitoring at any time. The study received ethical approval from the University of Helsinki in Finland

³<https://github.com/HIIT/Entity-Recommendation-for-Everyday-Digital-Tasks>

⁴<https://github.com/tesseract-ocr/tesseract/wiki>

⁵<https://www.ibm.com/watson/services/natural-language-understanding/>

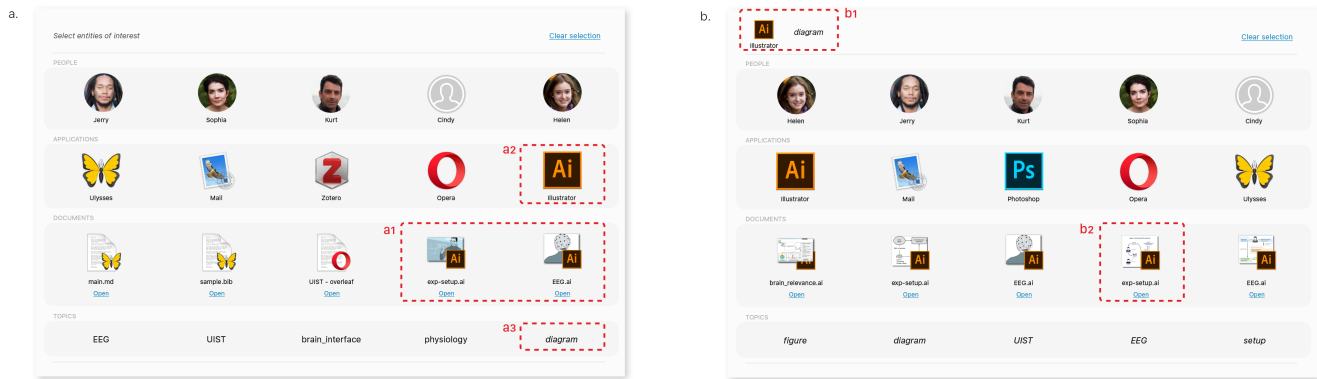


Figure 1: Two states of EntityBot’s user interface [11]. Recommended entities are displayed within four rows, here with five items each: people, applications, documents, and topics. The user can select entities of interest by clicking on them, which updates the recommendations. Example: In (a), the user 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.

3.3 User Interface

Figure 1 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 feature was implemented in our experimental setup.

3.3.1 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). 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, the URL of the webpage, and the unique ID of the email. 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 and the changes are reflected on the UI.

3.3.2 Allowing the selection of entities of interest by the user (explicit feedback). 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.

3.3.3 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.

4 CONCLUSIONS

In this demo, we introduce EntityBot [11] to the RecSys community. EntityBot supports everyday digital tasks by predicting the information needs of the users and providing them with the right information at the right time through interactive entity recommendations.

ACKNOWLEDGMENTS

Partially funded by the EU H2020 project CO-ADAPT, the MIUR (PON AIM), and the Academy of Finland (322653, 328875, 336085, 319264, 292334).

REFERENCES

- [1] Salvatore Andolina, Khalil Klouche, Tuukka Ruotsalo, Patrik Floréen, and Giulio Jacucci. 2018. Querytogether: Enabling entity-centric exploration in multi-device collaborative search. *Information Processing & Management* 54, 6 (2018), 1182–1202.
- [2] 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 (Hong Kong, China) (DIS ’18)*. Association for Computing Machinery, New York, NY, USA, 1295–1307. <https://doi.org/10.1145/3196709.3196734>
- [3] Salvatore Andolina, Valeria Orso, Hendrik Schneider, Khalil Klouche, Tuukka Ruotsalo, Luciano Gamberini, and Giulio Jacucci. 2018. SearchBot: Supporting voice conversations with proactive search. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 9–12.

- [4] Daniel Billsus and Michael J. Pazzani. 2000. User Modeling for Adaptive News Access. *User Modeling and User-Adapted Interaction* 10, 2 (2000), 147–180. <https://doi.org/10.1023/A:1026501525781>
- [5] Jay Budzik, Kristian J. Hammond, and Larry Birnbaum. 2001. Information access in context. *Knowledge-Based Systems* 14, 1–2 (2001), 37–53. [https://doi.org/10.1016/S0950-7051\(00\)00105-2](https://doi.org/10.1016/S0950-7051(00)00105-2)
- [6] Pedram Dae, 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* (Sonoma, California, USA) (IUI '16). ACM, New York, NY, USA, 71–75. <https://doi.org/10.1145/2856767.2856803>
- [7] Susan Dumais, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin, and Daniel C. Robbins. 2003. Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Toronto, Canada) (SIGIR '03). Association for Computing Machinery, New York, NY, USA, 72–79. <https://doi.org/10.1145/860435.860451>
- [8] 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* (Sheffield, United Kingdom) (SIGIR '04). ACM, New York, NY, USA, 594–594. <https://doi.org/10.1145/1008992.1009137>
- [9] Gerhard Fischer. 2001. User Modeling in Human–Computer Interaction. *UMUAI* 11, 1-2 (March 2001), 65–86.
- [10] 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* (Madison, Wisconsin) (UAI'98). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 256–265. <http://dl.acm.org/citation.cfm?id=2074094.2074124>
- [11] Giulio Jacucci, Pedram Dae, Tung Vuong, Salvatore Andolina, Khalil Klouche, Mats Sjöberg, Tuukka Ruotsalo, and Samuel Kaski. 2021. Entity Recommendation for Everyday Digital Tasks. *ACM Trans. Comput.-Hum. Interact.* 28, 5, Article 29 (2021). <https://doi.org/10.1145/3458919>
- [12] 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* (Paris, France) (CHI '13). ACM, New York, NY, USA, 681–684. <https://doi.org/10.1145/2470654.2470750>
- [13] 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* (Gold Coast, Queensland, Australia) (SIGIR '14). ACM, New York, NY, USA, 807–816. <https://doi.org/10.1145/2600428.2609585>
- [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* (Salvador, Brazil) (SIGIR '05). ACM, New York, NY, USA, 449–456. <https://doi.org/10.1145/1076034.1076111>