

SciNet: Interactive Intent Modeling for Information Discovery

Tuukka Ruotsalo¹, Jaakko Peltonen^{2,4}, Manuel J.A. Eugster², Dorota Glowacka³,
Aki Reijonen³, Giulio Jacucci³, Petri Myllymäki³, Samuel Kaski^{2,3}
Helsinki Institute for Information Technology HIIT

¹Helsinki Institute for Information Technology HIIT, Aalto University, Finland

²Aalto University, Department of Information and Computer Science, Finland

³University of Helsinki, Department of Computer Science, Finland

⁴University of Tampere, School of Information Sciences, Finland
first.last@hiit.fi

ABSTRACT

Current search engines offer limited assistance for exploration and information discovery in complex search tasks. Instead, users are distracted by the need to focus their cognitive efforts on finding navigation cues, rather than selecting relevant information. Interactive intent modeling enhances the human information exploration capacity through computational modeling, visualized for interaction. Interactive intent modeling has been shown to increase task-level information seeking performance by up to 100%. In this demonstration, we showcase *SciNet*, a system implementing interactive intent modeling on top of a scientific article database of over 60 million documents.

Keywords

Intent modeling; Interactive information retrieval; Personalization; Visual information seeking

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]

1. INTRODUCTION

Recent behavioral studies show that a large portion of users' information-seeking activities are exploratory and characterized by the complex and evolving nature of users' information needs [4]. As a result, the users face the problem of entering correct terms that describe their evolving search intents, so that the desired information can be retrieved on subsequent iterations [2]. We propose *Interactive intent modeling* promoting resourceful interaction between humans and information retrieval systems to enable information discovery that goes beyond search [6]. It tackles the vocabulary mismatch problem [2] by providing the user with knowledge about potential intents to discover, visualizes the intents as directions in the information space around the user's present position, and transfers the exploratory search to a recognition task in which

users can direct their search via an interactive visualization to interact with the system's intent model [7].

Interactive intent modeling is based on two principles [6]: 1) Visualizing the current search intents and directions in the information space, and 2) Interactive adaptation of the intent model balancing exploring the information space and exploiting the user feedback.

2. THE SCINET SYSTEM

The *SciNet* system, shown in Figure 1, implements interactive intent modeling on top of over 60 million scientific documents. The search starts with the user typing in a query, which results in a set of keywords being displayed and laid out on a radial display called the Intent Radar [8], and a ranked list of articles. The estimated intents, for which the results on the right side list have been retrieved, are visualized for the user (inner circle). The angular distance corresponds to similarity of intents and the radial distance from the centre corresponds to relevance. Predicted future intents, which are estimated for users to find directions on the radar away from their present intent estimate, are visualized in the outer circle of the Intent Radar. The user can provide positive feedback by dragging keywords closer to the center of the radar, and negative feedback by dragging them further away. Multiple keywords can be dragged at each iteration.

Figure 1 shows two iterations of a search session: the Intent Radar visualization of the first iteration (left) and the second iteration with both the visualization and the final result list. The user first issues a query "information visualization" and the system visualizes the intent model (inner circle) and suggested potentially interesting intents (outer circle) as keywords, and a ranked document list. The user increases the relevance of "visual information retrieval" (red arrow) by dragging it to the center of the radar. The system then computes and visualizes new intent model represented as a set of keywords (inner circle), including "image representation and indexing," "perception," and "visual data exploration." The system also computes projections in the intent space via pseudo-feedback and suggests relevant search directions for the user (outer circle), including "www," "information fidelity," "information visualization techniques," "interaction." The user can continue the search by selecting to exploit to more specific topics or to explore to more general but yet relevant topics.

3. ALGORITHMIC BUILDING BLOCKS

Learning the intent model. Given the evolutionary nature of search, as also demonstrated in our example search scenario, it is important to not only exploit the feedback elicited from the user

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). Copyright is held by the owner/author(s).

SIGIR'15, August 09-13, 2015, Santiago, Chile.

ACM 978-1-4503-3621-5/15/08.

DOI: <http://dx.doi.org/10.1145/2766462.2767863>.

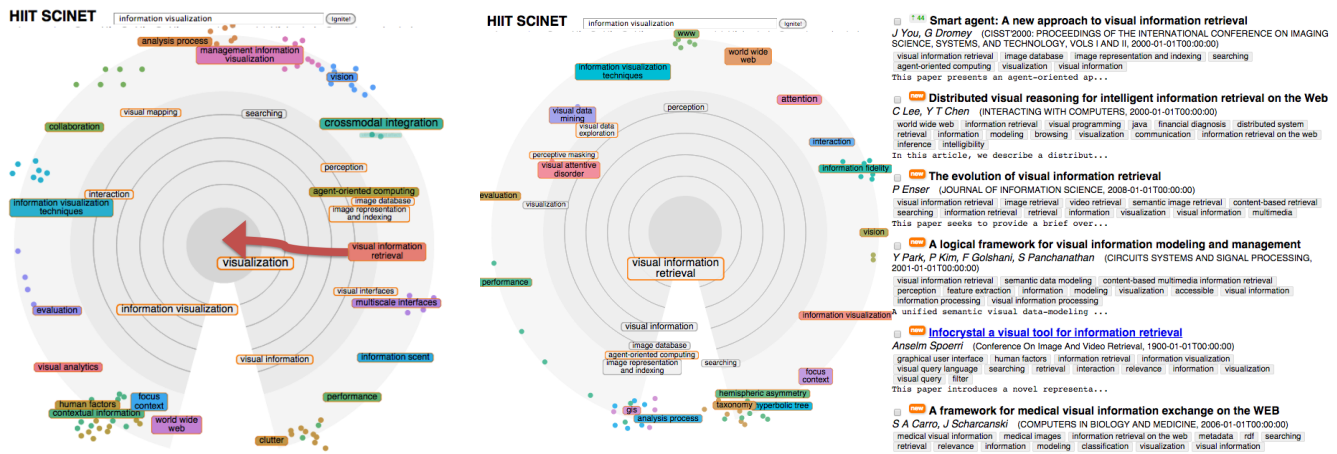


Figure 1: Two iterations of a search session with an initial query “information visualization”. Left: The Intent Radar visualization of the first iteration. The user drags the keyword “visual information retrieval” to the center of the Intent Radar (red arrow). Right: The second iteration. Both the Intent Radar visualization and the final result list are shown.

but also balance it with exploration. To put it simply, users need to be able to focus on a specific location of the information space (exploit) and also be able to broaden their search through more general areas (explore). To learn such models on-line, we use the *exploration-exploitation paradigm* of reinforcement learning [1, 3]. In this paradigm, the model (the search engine) and its environment (the user) form an online loop, and learning involves finding an optimal balance between exploration (selecting items from uncharted parts of the information space) and exploitation (selecting items most likely to be relevant given the current intent model) to acquire feedback on subsequent iterations. The learning mechanism predicts the model consisting of keyword relevances based on user interactions during the search session. The model is then used in combination with a statistical language model to rank the documents.

Interactive visualization. We optimize a data-driven layout for the search intents by probabilistic modeling-based nonlinear dimensionality reduction [9]. The task of the layout algorithm is to place keywords so that neighboring keywords on the display have neighboring characterizations. To highlight the structure in the outer circle layout, we apply a simple agglomerative clustering to angles of keywords in the outer circle and show for each cluster the label of the predicted most relevant keyword [7].

4. SUMMARY OF RESULTS

The *SciNet* system was studied in task-based experiments in which users were given 30 minutes to solve research tasks using information retrieval systems operating on a database of over 60 million scholarly articles. The system setup with interactive intent modeling was compared against conventional information retrieval system with a list-based visualization and interaction with typed queries. Interactive intent modeling was found to significantly improve users’ task performance on complex tasks, helping users to move away from their initial query context, thus allowing them to increase recall by over 50% while preserving precision. Moreover, session-level recall for novel information that was assessed separately was found to improve by over 100% [7, 5]. The improvements can be attributed to interactive intent modeling. The interaction with intent visualization does not fully replace the query-typing interaction, but users still rely on typed keywords when ini-

tializing new search directions. However, interactive intent modeling offers an additional complementary way to express search intents to direct search towards novel, but still relevant information.

5. ACKNOWLEDGMENTS

Certain data included herein are derived from the Web of Science prepared by THOMSON REUTERS, Inc., Philadelphia, Pennsylvania, USA; the Digital Libraries of ACM, IEEE, and Springer. The search engine is commercialized by Etsimo Ltd. (www.etsimo.com).

6. REFERENCES

- P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.*, 3:397 – 422, 2002.
- G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in human-system communication. *Commun. ACM*, 30(11):964–971, Nov. 1987.
- D. Glowacka, T. Ruotsalo, K. Konyushkova, K. Athukorala, S. Kaski, and G. Jacucci. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proc. IUI’13*, pages 117–128, New York, NY, USA, 2013. ACM.
- G. Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41–46, Apr. 2006.
- T. Ruotsalo, K. Athukorala, D. Glowacka, K. Konyushkova, A. Oulasvirta, S. Kaipainen, S. Kaski, and G. Jacucci. Supporting exploratory search tasks with interactive user modeling. In *In Proc. ASIS&T*, pages 39:1–39:10, Silver Springs, MD, USA, 2013. American Society for Information Science.
- T. Ruotsalo, G. Jacucci, P. Myllymäki, and S. Kaski. Interactive intent modeling: Information discovery beyond search. *Commun. ACM*, 58(1):86–92, Dec. 2014.
- T. Ruotsalo, J. Peltonen, M. Eugster, D. Glowacka, K. Konyushkova, K. Athukorala, I. Kosunen, A. Reijonen, P. Myllymäki, G. Jacucci, and S. Kaski. Directing exploratory search with interactive intent modeling. In *Proc. CIKM ’13*, pages 1759–1764. ACM, 2013.
- T. Ruotsalo, J. Peltonen, M. J. Eugster, D. Glowacka, A. Reijonen, G. Jacucci, P. Myllymäki, and S. Kaski. Intenttradar: Search user interface that anticipates user’s search intents. In *Proc. CHI EA ’14*, pages 455–458. ACM, 2014.
- J. Venna, J. Peltonen, K. Nybo, H. Aidos, and S. Kaski. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *J. Mach. Learn. Res.*, 11:451–490, 2010.