Supporting Exploratory Search Through User Modeling

Kumaripaba Athukorala [‡], Antti Oulasvirta [†], Dorota Glowacka [‡], Jilles Vreeken [†], Giulio Jacucci [‡]

[‡] Helsinki Institute for Information Technology HIIT
[‡] Department of Computer Science, University of Helsinki
[†] Max Planck Institute for Informatics and Saarland University

Abstract. Exploratory search is becoming more common as the web is used more increasingly as a medium for learning and discovery. Compared to traditional known-item search, exploratory search is more challenging and difficult to support because it initiates with poorly defined search goals, while the user knowledge and information-needs constantly change throughout the search process. Although information-retrieval algorithms have improved greatly in providing users with relevant information to specific queries, there is still room to improve our understanding of users' exploratory information-seeking strategies as well as design of systems supporting exploratory information-seeking. Modeling the user behavior and predicting dynamically changing information-needs in exploratory search is hard. Over the past decade there has been increasing attention on rich user interfaces, retrieval techniques, and studies of exploratory search. However, existing work does not yet support the dynamic aspects of exploratory search. The aim of this research is to explore different aspects of how to understand and support exploratory search, including user studies, intent visualization and user modeling.

Keywords: User Modeling, Exploratory Search, Scientific Information-Seeking

1 Introduction

Search can be broadly divided into two categories: known-item search and exploratory search. In known-item search the user has a specific search result in mind. On the other hand, in exploratory search the goal is ill-defined and changes as the search progresses [1]. Traditional information retrieval techniques concentrate mostly on known-item search. However, exploratory search is becoming more important as the web is becoming a major source for learning and discovery [2].

Exploratory information-seeking is known to be complex and hard to support due to its inherently open-ended and dynamic nature [3]. It arises in situations where there is a need to find information from a domain in which the user has a general interest but not specific knowledge [1]. Exploratory search has also

been defined based on the distinct characteristics of the search process such as submitting tentative queries, selectively seeking and passively obtaining cues about the next steps, and iteratively searching with evolving information-needs. In this paper, we concentrate on exploratory search in the scientific information-seeking context. To be more precise, we use a scientific essay writing scenario, where a student has to write an essay on a research topic in which she has a general interest but lacks knowledge to formulate queries to gather the necessary literature. This type of search involving exploration of unfamiliar research areas for the purpose of learning is found to be one of the most challenging literature search purposes [4].

Over the last decade many techniques have been proposed to provide better support for exploratory information-seeking, such as results clustering [5], relevance feedback [21], faceted search [7], and novel visualizations to support the creation of unfamiliar information spaces [8]. Even though these solutions help in improving exploration, exploratory search involves many different phases. For example, it begins with an imprecise query and then through several successive iterations of exploring the retrieved information and reformulating queries, the scope of the information need narrows down. This iterative and evolving nature of exploratory search makes it very difficult to identify the constantly changing information needs of the user and different phases of exploration.

Systems that suggest queries, provide interactive keyword visualizations, cluster results, and provide similar help to better support exploration need to "know" whether the suggested queries/keywords and selected clusters are too narrow or too broad for the current information need of the user. Hence, in exploratory search it is important to predict which stage of exploration the user is in with respect to the evolving state of his or her knowledge. One way to address this problem is by understanding user behaviors with queries with varying specificity in exploratory searching, which, in turn, will allow us to build a user model to predict whether a given query is too broad or too specific for the current information need of the user. Another method is by providing visualizations of systems interpretation of user needs and allow the user to provide feedback. The main goals of this research is to investigate exploratory search behaviors of academics, and build interaction models and visualizations that allow information retrieval systems to infer the state of exploration from the observable aspects of user interactions.

2 Related Work

Over the past decade researchers from, among others, information retrieval (IR), human-computer interaction (HCI), and cognitive science communities have made many attempts to better support the user in tasks involving exploratory search by developing retrieval techniques, user interfaces and conducting studies aim at understanding user behaviors in exploratory search.

In the context of information retrieval, existing contributions include relevance feedback based retrieval [21], faceted search [7], and result clustering [5].

However, evidence from user studies shows that results clustering and relevance feedback based methods are rarely used due to high cognitive overload of selecting relevant results and providing feedback, and the problem of the context trap [21]. Faceted search is found to be overly demanding as users have to go through a large number of options [7]. Furthermore, studies have shown that exploratory search requires more active user engagement with the search results [9]. The lack of success of systems such as relevance feedback is often attributed to user interface designs failing to conveniently provide feedback at suitable levels of granularity [7].

In response, a number of new techniques were designed to visualize search results and capture user feedback. Some of them include rich user interfaces combined with learning algorithms to support users to comprehend the search results [8], and visualization and summarization of results [10]. All these solutions are giving users more control, however, they fail to take the moment-by-moment information-needs of the user into consideration [11]. This is where user modeling can greatly improve existing approaches to exploratory search.

User behavior in exploratory information-seeking is studied with intents: predicting cognitive styles [12], identifying search and query formulation strategies [13], and constructing user models to predict the domain knowledge [14]. Early studies showed emergence of different search strategies depending on the users familiarity with the topic. Crucially, user studies show that users spend more time evaluating unfamiliar topics than familiar ones [14], domain knowledge and experience with a search tool impact search behavior [15], and that search strategies change over time when domain knowledge increases [16]. Existing models are useful in customizing results according to user preference [17] and knowledge, however, they do not capture situations where domain experts search information in narrower sub-fields of a familiar domain. Information Foraging Theory (IFT) provides several quantitative models of user search [18], yet existing work on IFT does not consider the effect of evolving user knowledge and queries. Overall, behavioral studies clearly point to the dynamic nature of the exploratory information-seeking process and the effect of prior knowledge on users' search strategies, which lends support to the assumptions behind the models we develop in this research. Our aim is to design user models that predict moment-by-moment information-needs of the user through observable user behaviors to improve the performance of retrieval algorithms.

3 Information-Seeking Behaviors and Intent Visualization

Exploratory search is very common among academics. Therefore, we conducted a study to investigate how academics search for scientific information and what challenges they face. This was a mixed method study involving interviews, diary logs, user observations, and a survey. The findings suggested that exploring unfamiliar research areas was one of the most common purposes of scientific information-seeking and it is the most difficult task to perform [4]. Results of this study provided useful insights into the problem of exploratory search.

In the initial stages of exploratory search, users have poor knowledge about the information space. Therefore, visualizations of the underlying information space can help the users to make better sense of the search topic. As a part of this research a prototype search tool called SciNet was developed [19–21]. With SciNet, the user can perceive the state of user model through the interactive visualization and provide feedback by moving keywords. User studies that compared SciNet indicated that it helps users to more effectively find relevant, novel and diverse results.

4 Distinguish Exploration from Navigational Search

We also conducted a study to compare exploratory search and navigational/known-item search. The results indicated that unlike in known item search in exploratory search there is a higher percentage of fixations even on results at the bottom of the ranked list of Search Engine Results Pages (SERPs) (See Figure 1). These results are useful in building a model to distinguish exploratory search from navigational search.

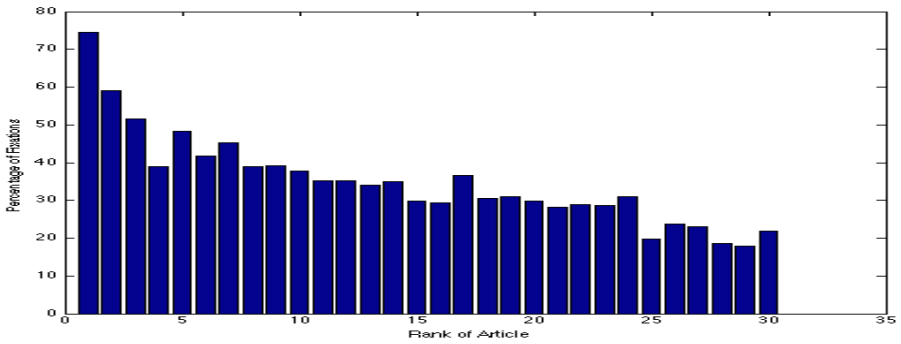

Fig. 1. Percentage of fixations at each article in the SERPs displaying 30 results in an exploratory search task. This figure indicates that in exploratory search users fixate more on results than in known item search.

5 Interaction Model to Predict Stages of Exploration

We also designed a model by combining insights from research into exploratory search and Information Foraging Theory (IFT) [18]. According to IFT, information gain can be modeled as a linear function of time when the results are ordered by relevance to the query. Further, IFT states that this information gain function will qualitatively shift towards a diminishing returns curve if new interface elements, such as result clustering, are introduced. Hence, IFT shows how information gain is affected by the user interface changes.

Our research is motivated by this model. If we keep the user interface constant, the information gain function should change according to the stage of exploration. We define the stages of exploratory search as broad, intermediate, and specific. In the broad stage the user has very little knowledge about the search topic and will issue a vague search query addressing a very broad information space. In the intermediate stage, the user would have some idea about the topic and would reformulate queries referring to sub-areas in the search topic. In the specific stage, the user would have gained a good enough knowledge and would use queries referring to very specific search topics in the area. We refer to this as subjective specificity. Sometimes a user might start the search with a very specific search query without having any knowledge about the area. Such a situation might arise when a search engine suggests keywords to the user, or when the user picks up some new terms randomly from the results without actually learning about them. If a search engine can predict the subjective specificity of the search results then it would be very useful to personalize the search results.

Our model captures how information gain in exploratory search is affected by this subjective specificity. The key idea is that the same search result can have very different information content for a user depending on how well it matches their current information needs. Consider two users who differ in the specificity of their goals and the extent of previous knowledge about a given topic, for example an undergraduate student writing a short overview essay on a well-known topic versus an experienced researcher gathering information about the latest developments in a specialized field. Their responses would differ, the former user probably spending more time on every item and the latter quickly scanning for the most informative items.

Empirical evaluation shows that our model captures the effects of query-specificity as well as the known effect of both prior knowledge and experience. Through a preliminary study we show the feasibility of using our model in a running IR system for predicting query-specificity.

6 Future Contributions

An important future challenge is to investigate in a real exploratory information-seeking scenario the performance of the formal model that we developed to predict the specificity of search results. We have already conducted a preliminary classification study which found that a system using only a simple classifier can obtain informed estimates on the specificity of a query while the user is interacting with its results. In the future, we will incorporate our model in a running IR system and further validate its usefulness in enhancing performance of exploratory search tasks.

References

1. White, R. W., Kules, B., Drucker, S. M.: Supporting exploratory search. J. Comm. ACM, 49(4), 36–39 (2006)

2. Teevan, J., Alvarado, C., Ackerman, M. S., Karger, D. R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. In: CHI, 415–422 (2004)
3. Marchionini, G.: Exploratory search: from finding to understanding. J. Comm. ACM, 49(4), 41–46 (2006)
4. Athukorala, K., Hoggan, E., Lehtio, A., Ruotsalo, T., Jacucci, G.: Informationseeking behaviors of computer scientists: Challenges for electronic literature search tools. In: AGM of Assoc. Inf. Sci. Tech. (2013)
5. Ferragina, P., Gulli, A.:A personalized search engine based on Web-snippet hierarchical clustering. J. Software: Practice and Experience, 38(2), 189–225 (2008)
6. Kelly, D., Fu, X.: Elicitation of term relevance feedback: an investigation of term source and context. In SIGIR, 453–460 (2006).
7. Yee, K. P., Swearingen, K., Li, K., Hearst, M.: Faceted metadata for image search and browsing. In: CHI, (2003)
8. Chau, D. H., Kittur, A., Hong, J. I., Faloutsos, C.: Apolo: making sense of large network data by combining rich user interaction and machine learning. In CHI, pp 167–176 (2011)
9. Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., Yee, K.: Finding the flow in web site search. J. Commun. ACM, pp. 42-49 (2002)
10. Kki, M.:Findex: search result categories help users when document ranking fails. In CHI, pp 131–140 (2005)
11. Teevan, J., Alvarado, C., Ackerman, M. S., Karger. D. R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. In CHI, pp 415–422 (2004)
12. Saito, H., Miwa, K.: A cognitive study of information seeking processes in the www: the effects of searchers knowledge and experience. In: WISE, pp 321327 (2001)
13. Payne, S. J., Howes, A.: Adaptive interaction: A utility maximization approach to understanding human interaction with technology. In Synt. Lec. Hum.-Cent. Inf., pp 1–111 (2013)
14. Holscher, C., Strube, G.: Web search behavior of internet experts and newbies. J. Computer Networks, 33(1), pp 337-346 (2000)
15. Vakkari, P., Pennanen, M., Serola, S.: Changes of search terms and tactics while writing a research proposal: A longitudinal case study. In: Inf. Proc. Manag., 39(3), pp 445–463 (2003)
16. Wildemuth, B. M.: The effects of domain knowledge on search tactic formulation. J. Assoc. Inf. Sci. Tech., 55(3), pp 246–258 (2004)
17. Kelly, D., Teevan, J.: Implicit feedback for inferring user preference: a bibliography. In: SIGIR, pp. 18–28. (2003)
18. Pirolli, P., Card, S.: Information foraging in information access environments. In: CHI, pp 51–58, (1995)
19. Głowacka, D., Ruotsalo, T., Konyushkova, K., Athukorala, K., Kaski, S., Jacucci, G.: Directing exploratory search: Reinforcement learning from user interactions with keywords. In: IUI (2013)
20. Ruotsalo, T., Athukorala, K., Glowacka, D., Konyushkova, K., Jacucci, G., Kaski, S.: Supporting Exploratory Search Tasks with Interactive User Modeling. In:ASIST (2013)
21. Ruotsalo, T., Peltonen, J., Eugster, M., Głowacka, D., Konyushkova, K., Athukorala, K., ...and Kaski, S. Directing exploratory search with interactive intent modeling. In CIKM, pp. 1759–1764, ACM. (2013)