Supporting Complex Information Needs via Large-Scale Collaborative Search

Felipe Moraes
Delft University of Technology
Delft, The Netherlands
f.moraes@tudelf.nl

The need for exploration comes from users' information needs that are highly complex: open-ended and multi-faceted information needs that are derived from complex search tasks such as learning tasks. Researchers have shown that exploratory search can be better supported through explicit search collaborations. More importantly, previous research has suggested that leveraging collaborations of a large group of users may lead to an improved user experience. However, there is still a lack of evidence in support of this hypothesis as past studies were usually conducted in restricted environments (laboratory studies with small groups of participants and simulated work tasks). In this paper, we briefly discuss how we will shed light on this research field, and we present current work and future directions.

Complex information needs; Exploratory search; Collaborative search;

1. INTRODUCTION

Exploratory search encompasses complex information needs that are often open-ended and multi-faceted, in which users usually pose multiple queries and iteratively interact with the search results and the search engine results pages (SERPs) (Marchionini 2006). This need for exploration arises from various context-aware search tasks such as personal search, professional search, and learning as search tasks (White and Roth 2009; White 2016; Goker et al. 2009). The latter has received increasing research attention with recent advances in large-scale online learning offered through portals such as Coursera[1] and edX[2].

Research on Human-Computer Interaction (HCI) and Information Retrieval (IR) have shown that exploratory search can be better supported in terms of efficiency, material coverage and knowledge gains when conducted in an explicit collaboration (Shah 2010). Importantly, previous work has suggested that search collaborations conducted in larger groups may lead to an improved user experience; this is in contrast to the rather small groups of users (up to 5) that have been the focus of collaborative search research so far (Morris et al. 2010; Shah et al. 2016). As many existing collaborative search studies are limited by the nature of the study (lab-based, short-term, with a participant pool of less than 50 study participants, usually a homogeneous group working with simulated information needs (Hearst 2014)), empirical evidence in support of this hypothesis is still missing.

Our goal in this work is to **scale collaborative search to a large number of searchers**. Starting off with the findings in small-group collaborative search, we aim to develop an in-depth understanding of collaborative search "at scale". In particular, we will design and build search tools that can be deployed in Massive Open Online Courses (MOOCs), which offer us (i) a large number of potential study participants (i.e. "learners")[3], (ii) real complex information needs derived from learners' study needs, (iii) long-term search needs, and, (iv) a demographically heterogeneous group of searchers. In the process, we will also be able to verify and validate previous research findings, e.g. (Golovchinsky et al. 2012; Morris and Horvitz 2007; Paul and Morris 2009) in a domain that is several magnitudes larger and more realistic than the lab setup of prior works.

In this paper, we provide an overview of our research project in which the expected main contributions are:

- Novel algorithmic mediation approaches to automatically divide the collaborative search space in the best possible manner by taking the search collaborators' preferences, cultural backgrounds, and expertise into account.

---

[1] http://www.coursera.org
[2] http://www.edx.org/

[3] Although privacy of the data shared among learners is a concern in this scenario, we consider it (for now) as out-of-scope.

- Techniques to foster collaborators' awareness of each others' activities, that should incur a low cognitive cost.

- Provide insights into the costs and benefits of collaborations at such large scale and determine when the tipping point (the costs of collaboration outweigh the benefits) is reached.

- An open-source collaborative search framework that allows us to perform a wide range of live studies of the algorithms to be developed, and the setup of the collaborating groups.

In the remaining sections of this paper we provide a brief discussion on our research design, current work, and future directions.

2. RESEARCH DESIGN

In this section, we briefly discuss our research design for the doctoral research project. We have the following research statement:

Research Statement An effective exploratory search experience can be obtained via a large group of collaborators that share complex information needs. In particular, explicit, asynchronous, and distributed large-scale collaborative search with algorithmic mediation can lead to a *more effective* search experience and *greater coverage* of the search space than individual search or small-group collaborative search.

To provide empirical support for this statement, we designed this doctoral research project to be conducted in four research stages as shown in Figure 1:

1. Exploring Existing Technologies

We will empirically investigate the boundaries of existing collaborative search technologies. For that, we will reproduce previous small-scale lab-studies in a large-scale environment (by employing them in the MOOC setting), e.g. (Golovchinsky et al. 2012; Morris and Horvitz 2007; Paul and Morris 2009). As a result, we aim to show: (i) what group sizes can existing approaches to explicit collaboration search support effectively; (ii) what impact the group coherence and diversity have on search effectiveness and coverage; (iii) at what point the cognitive effort of collaboration outweigh the benefits; (iv) whether guiding principles of small-group collaborative search (Brennan et al. 2008) are valid in large-group settings; and (v) what additional principles are required to enable large-group collaborations.

2. Iterative Collaborative Search

We will begin the design and development of a dedicated open-source large-group collaborative search framework. In this stage, we will investigate how

iterative support Little et al. (2010); Fisher et al. (2012) can benefit collaborative search. In particular, we will explore: (i) how can iterative search be supported algorithmically and based on which criteria should the search space be partitioned across collaborators; (ii) how can users make sense of the iterative search process; and (iii) for how many iterations can we observe positive effects on knowledge gain, depth of understanding, and speed of content assimilation.

3. Role-based Collaborative Search

After investigating the trade-offs of iterative collaborative search, we will turn to an alternative, orthogonal strategy: distributing different roles to the collaborators based on their preferences, expertise, and availability (Pickens et al. 2008; Shah et al. 2010). In particular, we aim to show: (i) how many different type of roles (and which ones) are beneficial to assign in large-group collaborative search, and how can collaborators effectively share a single role; (ii) what effects different information retrieval strategies have on the effectiveness of algorithmic mediation; and (iii) what is the interplay between iterative collaborative search and role-based collaborative search, and does it lead to additional gains in our assessment metrics when combined in a single iterative and role-based collaborative search setup.

4. Learning via Search Metrics

The realization of empirical studies (that is, the deployment of large-group collaborative search tools in actual MOOCs) plays a vital role in the three stages outlined thus far. A valuable by-product of these to be deployed collaborative search prototypes will be query logs as well as the learner logs (learners' behaviour on the MOOC platform, including their actions on videos and their performance on quizzes) the learners generate. In this stage, we will use the indicators of learning as a foundation to investigate: (i) how can the amount of learning that is taking place in the search process be quantified on a group and individual user level; (ii) which models of user or group behavior are sufficiently accurate and elementary to enable the development of feasible model-based learning metrics; (iii) when and by whom (all users or only a subset of users involved in the search) does learning occur in the collaborative process.
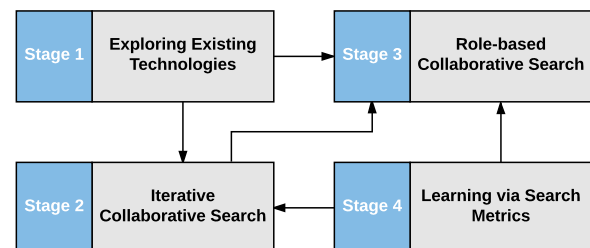

Figure 1: *A diagram about relationship between the four different stages.*

3. CURRENT WORK

As a zero step towards this doctoral research project is the understanding of how learners *individually* search on the Web to *complement* their learning process. We are currently working on a Web search framework (powered by Bing) that will be deployed inside MOOCs, enabling learners to search on the Web from within the MOOC platform. This setup enables us to observe how learners translate their learning needs into searches and increase our understanding of individual learners search needs during the learning process. Once we obtain those insights, we will move towards our first research stage and implement collaborative search tooling inside the framework. Our current work is guided by the following research questions:

$RQ1$ At what stages of the learning process (Wilson and Wilson 2013) do learners search on the Web to increase their knowledge?

$RQ2$ How does the complexity of the information needs relate to the learning process?

$RQ3$ How do learners materialize their information needs? To what extent does prior knowledge influences the expressiveness of learners' information needs?

4. NEXT STEPS

An immediately future direction of the current work is to investigate how to better support complex information needs for individual users. Specifically, we are interested in validating past research that relied on laboratory studies. Promising candidates here include comparing term suggestions and query suggestions (Kelly et al. 2009), distinguishing low-quality from high-quality query suggestions (Kelly et al. 2010), and comparing different interfaces (structured, standard and query suggestion) (Azzopardi et al. 2013).

ACKNOWLEDGEMENTS

REFERENCES

Azzopardi, L., D. Kelly, and K. Brennan (2013). How query cost affects search behavior. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 23–32. ACM.

Brennan, S. E., X. Chen, C. A. Dickinson, M. B. Neider, and G. J. Zelinsky (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition 106*(3), 1465–1477.

Fisher, K., S. Counts, and A. Kittur (2012). Distributed sensemaking: Improving sensemaking by leveraging the efforts of previous users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, New York, NY, USA, pp. 247–256. ACM.

Goker, A., H. Myrhaug, and R. Bierig (2009). Context and information retrieval. *Information retrieval: Searching in the 21st century*, 131–57.

Golovchinsky, G., A. Dunnigan, and A. Diriye (2012). Designing a tool for exploratory information seeking. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, New York, NY, USA, pp. 1799–1804. ACM.

Hearst, M. A. (2014, March). What's missing from collaborative search? *Computer 47*(3), 58–61.

Kelly, D., A. Cushing, M. Dostert, X. Niu, and K. Gyllstrom (2010). Effects of popularity and quality on the usage of query suggestions during information search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 45–54. ACM.

Kelly, D., K. Gyllstrom, and E. W. Bailey (2009). A comparison of query and term suggestion features for interactive searching. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pp. 371–378. ACM.

Little, G., L. B. Chilton, M. Goldman, and R. C. Miller (2010). Exploring iterative and parallel human computation processes. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, New York, NY, USA, pp. 68–76. ACM.

Marchionini, G. (2006, April). Exploratory search: From finding to understanding. *Commun. ACM 49*(4), 41–46.

Morris, M. R., D. Fisher, and D. Wigdor (2010, November). Search on surfaces: Exploring the potential of interactive tabletops for collaborative search tasks. *Inf. Process. Manage. 46*(6), 703–717.

Morris, M. R. and E. Horvitz (2007). Searchtogether: An interface for collaborative web search. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, New York, NY, USA, pp. 3–12. ACM.

Paul, S. A. and M. R. Morris (2009). Cosense: Enhancing sensemaking for collaborative web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, New York, NY, USA, pp. 1771–1780. ACM.

Pickens, J., G. Golovchinsky, C. Shah, P. Qvarfordt, and M. Back (2008). Algorithmic mediation for collaborative exploratory search. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, New York, NY, USA, pp. 315–322. ACM.

Shah, C. (2010). Collaborative information seeking: A literature review. *Advances in Librarianship (Advances in Librarianship, Volume 32), Emerald Group Publishing Limited 32*, 3–33.

Shah, C., R. Capra, and P. Hansen (2016, January). Workshop on social and collaborative information seeking (scis). *SIGIR Forum 49*(2), 117–122.

Shah, C., J. Pickens, and G. Golovchinsky (2010, November). Role-based results redistribution for collaborative information retrieval. *Inf. Process. Manage. 46*(6), 773–781.

White, R. W. (2016). *Interactions with search systems*. Cambridge University Press.

White, R. W. and R. A. Roth (2009). Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services 1*(1), 1–98.

Wilson, M. J. and M. L. Wilson (2013). A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. *Journal of the Association for Information Science and Technology 64*(2), 291–306.