

Research Challenges in Search as Learning

Claudia Hauff
Delft University of Technology
Delft, The Netherlands

Kevyn Collins-Thompson
University of Michigan
Ann Arbor, USA

Preben Hansen
Stockholm University
Stockholm, Sweden

ABSTRACT

Today's Web search engines are engineered and optimized to fulfill individual users' lookup tasks. This efficiency, however, also means that we largely view search systems as tools to satisfy immediate information needs, instead of rich environments in which humans heavily interact with information content, and search engines act as intelligent dialogue systems. In recent years, there has been a growing recognition of the importance of studying and designing search systems to foster discovery and enhance the learning experience during the search process outside of formal educational settings. To foster this line of research, we have recently hosted a seminar at Schloss Dagstuhl, bringing together HCI, IR, psychology and pedagogy researchers. In this talk, we will outline the current research paradigms in search as learning as well as the numerous research challenges this group of researchers identified.

1 MOTIVATION

Today's Web search engines—often also powering domain-specific and site-specific search—are optimized to fulfil individual users' lookup tasks. This efficiency, however, also means that we largely view search systems as tools to satisfy immediate information needs, instead of rich environments in which humans heavily interact with information content, and search engines act as intelligent dialogue systems, facilitating the communication between users and content. Web search engines are not designed for complex search tasks that require *exploration and learning*, user collaborations and involve different information seeking stages and search strategies, despite the fact that more than a quarter of Web searches are complex.

In recent years, there has been a growing recognition of the importance of studying and designing search systems to foster discovery and enhance the learning experience during the search process outside of formal educational settings. Searches that lead to learning, are naturally complex. Research progress in this area, however, is slow, with many more open questions than answers. Several critical bottlenecks and major impediments to advancements in the so-called *search as learning* area exist, including:

- (1) the reliance on small-scale lab studies to evaluate novel approaches which severely limit the diversity of investigable factors as well as the validity and generalizability of the findings;
- (2) the lack of awareness among researchers' initiatives in this very multidisciplinary area of work; and,
- (3) the lack of a shared research infrastructure.

Keeping these bottlenecks in mind, the seminar focused on four interconnected research themes, for each of which a considerable number of challenges as well as potential solutions and approaches were discussed (which are now published at [1]):

- (1) understanding search as a human learning process;
- (2) the measurement of learning performance and learning outcomes during search;
- (3) the relationship between the learning process and the search context;
- (4) the design of functionalities and search system interventions to promote learning.

2 SELECTED CHALLENGES

Below are a number of challenges that are particularly relevant to IR researchers:

- (1) Search query logs offer a very limited view into users' minds; we have to make educated guesses on their learning intent, their prior expertise and their context based on noisy signals. In order to make strides into understanding learning we require large-scale data with more semantic meaning behind it. How would such data look like and how can we collect it at scale?
- (2) How can we measure to what extent robust learning (learners achieving a deep conceptual understanding) or transfer learning (learners employing learnt concepts in novel situations) is taking place during the search process?
- (3) How can we help users that want to learn something but already struggle early on in the search process when formulating an initial query based on their information need? A common use case here are laymen medical inquiries (e.g. "pain in my side").
- (4) How can we support users that are searching for information with search requests that are already providing a certain answer frame (e.g. "vaccinations are bad", "climate change is not real")?
- (5) How can we track what users do with the information after the search to make sense of the information, and recognize the value of the search system?
- (6) What scalable measures, based on search behaviours and document characteristics are good approximators of learning gains? Across which periods of time (if we think for instance about sequences of queries across sessions) can we reliably measure learning gains?
- (7) Learning through failure: users may also learn when their information needs are not satisfied and their goals are not achieved. How can we deal with this scenario?
- (8) How should a conversational agent (presumably trained automatically on vast quantities of text) deal with questions to which no clear consensus answer exist ("Does God exist")?

REFERENCES

- [1] Kevyn Collins-Thompson, Preben Hansen, and Claudia Hauff. 2017. Search as Learning (Dagstuhl Seminar 17092). *Dagstuhl Reports* 7, 2 (2017), 135–162. DOI: <http://dx.doi.org/10.4230/DagRep.7.2.135>