

Report on the Second Workshop on Supporting Complex Search Tasks

Marijn Koolen¹ **Jaap Kamps²** **Toine Bogers³**
Nicholas Belkin⁴ **Diane Kelly⁵** **Emine Yilmaz⁶**

¹ Huygens ING/KNAW, Netherlands

² University of Amsterdam, Netherlands

³ Aalborg University Copenhagen, Denmark

⁴ Rutgers University, NJ, USA

⁵ University of Tennessee Knoxville, TN, USA

⁶ University College London, UK

Abstract

There is broad consensus in the field of IR that search is complex in many use cases and applications, both on the Web and in domain-specific collections, and both in our professional and in our daily life. Yet our understanding of complex search tasks, in comparison to simple look up tasks, is fragmented at best. The workshop addressed many open research questions: What are the obvious use cases and applications of complex search? What are essential features of work tasks and search tasks to take into account? And how do these evolve over time? With a multitude of information, varying from introductory to specialized, and from authoritative to speculative or opinionated, when should which sources of information be shown? How does the information seeking process evolve and what are relevant differences between different stages? With complex task and search process management, blending searching, browsing, and recommendations, and supporting exploratory search to sensemaking and analytics, UI and UX design pose an overconstrained challenge. How do we know that our approach is any good? Supporting complex search tasks requires new collaborations across the whole field of IR, and the proposed workshop brought together a diverse group of researchers to work together on one of the greatest challenges of our field.

The workshop featured three main elements. First, two keynotes, one on the complexity of meaningful interactive IR evaluation by Mark Hall and one on the types of search complexity encountered in real-world applications by Jussi Karlgren. Second, a lively boaster and poster session in which seven contributed papers were presented. Third, three breakout groups discussed concrete ideas on: (1) search context and tasks, (2) search process, and (3) evaluation of complex search tasks. There was an general feeling that the discussion made progress, and built new connections between related strands of research in IR.

1 Introduction

One of the current challenges in information access is supporting complex search tasks. A user's understanding of their information need and their overall task develops as they interact with the system. Supporting the various stages of the task involves many aspects of the system, e.g. interface features, presentation of information, retrieving and ranking. Many search systems treat the search process as a series of identical steps of submitting a query and consulting documents. Yet information seeking research has shown that users go through different phases in their search sessions, from exploring and identifying vague information needs, to focusing and refining their needs and search strategies, to finalizing their search. To be able to support exploring and discovering strategies we need to understand the characteristics of different tasks including open-ended, leisure-focused sessions. This is a highly complex problem that touches upon and bridges areas of information seeking, interactive information retrieval, system-centered (ranking, evaluation), and user interface design.

The background for this workshop is derived from the Interactive Track of the CLEF/INEX Social Book Search Lab [5–7], which investigates scenarios with complex book search tasks and develops systems and interfaces that support the user through the different stages of their search process.

The overall goal of the workshop was to create and foster an interdisciplinary forum where researchers can exchange and contribute to the development of alternative experiments and prototypes. The main aim was to better understand how to support complex search tasks, addressing many open research questions to be explored, including:

Context What are the obvious use cases and applications of complex search? In what sense are these “complex”? What generic characteristic do they share? How can search become an integral part of its context, and the context integral part of search?

Tasks What are essential features of work tasks and search tasks to take into account? And how do these evolve over time? How do complex tasks be decomposed into manageable sub-tasks, and partial results composed into comprehensive answers? How can we monitor and support task progress?

Heterogeneous sources With a multitude of information, varying from introductory to specialized, and from authoritative to speculative or opinionated, when to show what sources of information? When to show more or other types of information than directly requested by the searcher? Do we know when the user has gotten enough?

Search process How does the information seeking process evolve and what are relevant differences between different stages? What search tactics and search strategies are effective? How can we promote the use of effective search strategies? How does the information need evolve and what are relevant success criteria for the end result and intermediate steps? How can we cast these as effective complex queries, and how to (interactively) construct such queries?

UI and UX Does the need of complex task and search process management, blending searching, browsing, and recommendations, and supporting exploratory search to sense-making and analytics, make UI and UX design an overconstrained challenge? What affordances are

required and in what stage of the search process? How can we make the search process transparent to the user? How and when does the initiative shift between system and user?

Evaluation How do we know that our approach is any good? Can we carve out one or a range of generic aspects testable on a suitable benchmarks? Is there enough empirical evidence to ground simulated interactive search? What kind of novel retrieval models are needed to combine topical, contextual and preferential aspects?

SCST 2017 was a half day workshop on supporting complex search tasks—a *workshop* proper where discussion was central, and all attendees were active participants. The workshop brought together a varied group of researchers with experience covering both user and system centered approaches, to work together on the problem and potential solutions, and identify the *barriers* to success and work on ways of addressing them.

The rest of this report will follow the program of the workshop. The workshop started with a round of introductions where each attendee introduced him- or herself, and explained their own interest in the topic. Next, it featured two short keynotes (discussed in §2) which helped frame the problems and reach a shared understanding among all workshop attendees of the issues involved. Mark Hall (Edge Hill University) talked about interactive IR evaluation and Jussi Karlgren (Gavagai, Sweden) talked about the hidden complexity of seemingly simple business information needs. This was followed by a boaster and poster session in which nine papers (discussed in §3) were presented. In the next session, participants divided over three breakout groups (discussed in §4), with in-depth discussion on three topics of importance in the area: (1) search context and tasks, (2) search process, and (3) evaluation of complex search tasks. In the final session the results and progress of the workshop was discussed and preliminary conclusions were drawn (discussed in §5).

2 Keynote

The workshop started with two short keynotes, one from an academic perspective and one from an industry perspective, to set the stage and ensure all attendees were on the same page.

2.1 Where does it end? Complex Search Tasks and Evaluation

Mark Hall (Edge Hill University, UK) presented the academic keynote in which he explored the blurring boundary between complex search tasks and the larger work tasks that motivate the search. This has important implications for what aspects of the process we should evaluate and how we do the evaluation in a meaningful and measurable way.

The task is one of the core concepts around which Information Retrieval (IR) and particularly its evaluation are structured. The traditional search task is a set of instructions that motivate a query and which enable the assessment of the query results' relevance. With the introduction of the user into the evaluation process the task evolved into a set of instructions that the user is provided with when they start the evaluation process. As these tasks become ever more complex the boundary between the IR task and the larger work task that motivates the IR task is starting to blur. This transition poses a major question for evaluation, as it also blurs the question of whether the evaluation is judging the quality of the IR system, the quality of the data for completing the

work task, or the user’s ability to identify with and complete the larger work task. In his keynote, Mark explored some of these issues.

2.2 Complex Aspects of Seemingly Simple Information Needs

Jussi Karlsgren (Gavagai, Sweden) presented the industrial perspective and discussed how the typical information needs of corporate customers are often posed in short and basic questions but are surprisingly hard to formulate in meaningful queries, and answering them requires complex processes of curating and aggregating diverse and disparate data.

In commercial practice customers have seemingly simple information needs. What are customers and potential customers saying about my brand online? Are our customers happy? What do our customers want from us? How could we make them happier? These questions are easy to pose and intuitively will provide a basis for commercial action if information to meet them is given by a system. But at least three non-trivial challenges lie in the way of fulfilling information needs of this kind: (1) formulating an information need in terms a system can accommodate; (2) curating, aggregating, and refining the relevant information sources of reasonable quality; (3) understanding client strategies and sometimes internal unstated organisational mechanisms on the client side. These are all in some sense gaps between technology offerings and business needs which need to be addressed for information access technology to progress from ad-hoc search.

3 Accepted Papers

The workshop invited short paper contributions to be presented as posters. We received 11 submissions and accepted 9 (for an 81% acceptance rate). Each paper was reviewed by at least three reviewers. Paper contributions are presented as a 1-minute boaster talk and as a poster during the interactive poster session.

Rutter et al. [12] discuss a case study of a type of complex task that at face value is simple and straightforward, but turns out to be complex to resolve: how do you make a phone safe for a child? There are a lot of opinions online, many possibilities for actions, many variations in hardware and software, but ultimately no single clear and correct answer for everyday phone users.

Bogers et al. [3] report on the experiences and challenges in organizing the CHIC and SBS Interactive Tracks from 2013 to 2016 in the form of a list of important properties. These properties could inform the design of new IIR evaluation campaigns and related researcher communities in ways that expand our understanding of information (seeking) behavior.

Koesten and Singh [10] focus on how a large governmental data portal in the UK supports users in conducting complex search tasks involving data, identify problems with the used interface, and discuss potential research directions to improve interfaces for complex, data-related search tasks.

Hoeber et al. [8] examine the use of exploratory search strategies for purposive sampling from large text collections. The use of exploratory search strategies that leverage visual analytics enables them to consider the relevance of the data in addition to more traditional sampling methods.

Egusa et al. [4] investigate the use of concept maps—graphical representations that allow people to represent their knowledge explicitly—to evaluate the effects of interactive complex search. Their

study showed a significant change in the concepts maps produced before and after executing a complex search.

Huurdeeman [9] proposes a framework for the design of search user interfaces for complex search tasks. His framework covers three different types of features—personalizable features, informational features, and input & control features—and discusses the different stages of complex information seeking where these features are relevant.

Ventocilla et al. [13] suggest a bottom-up approach to displaying and exploring relations and correlations in datasets. Using billiards as a metaphor, a graph-representation of (cor)relations in a dataset are unfolded in directions based on the user’s choices. This provides an intuitive exploratory faceted search interface with quantitative analyses calculated at run-time.

Novin [11] argues that studies on complex search tasks should make their designs more context-based, which will make them more applicable to real-world scenarios, as well as more reproducible and falsifiable. The paper reviews literature on cognitive experiments that stress the importance of situation on actions and proposes a outside-in approach where the context is defined first, then the work task, after which different experimental variables can be considered.

Arora and Jones [2] conducted a user study to investigate how users perceive relevance and importance of highlighted document fragments related to specific search topics, to better understand how to generate effective summaries of documents. The results provide insights on what types of information are effective for satisfying information needs and why users find some parts more relevant than others.

4 Breakout Sessions

The second half of the workshop consisted of 3 breakout groups, seeded from the open research questions (see §1) and the contributed papers (see §3). The goal of the breakout groups was to discuss concrete activities to make progress on this topic, for instance in the form of a workshop or a new Interactive IR task at e.g. TREC or CLEF. The three themes were (1) search context and tasks, (2) search process, and (3) evaluation of complex search tasks.

4.1 Search Context and Tasks

The first break-out session focused on tasks and context and had 17 participants. Many participants agreed that while several taxonomies of task facets have been proposed, there is no true methodical approach for how to construct authentic tasks yet. One proposal was to use Bloom’s taxonomy of learning outcomes could be used to prompt different kind of behaviors, although not every task has learning focus. Another actively debated question was the role that context plays in tasks: is it even possible to talk about complex tasks without involving context? The context that people are in influences the complexity of that task (mobile vs. desktop, time pressure vs. relaxed, etc.) One potential barrier to progress on research on tasks is the lack of a common set of instruments that as a minimum could be used in interactive IR experiments with a possibility for extension. Unfortunately, this has not been done yet and it makes experiments difficult to compare. A proposed solution to this was a repository of interactive IR data to allow researchers to share and compare their data.

At the end of the session, several actionable outcomes were identified to further the state of research on task-based search:

- A repository for interactive track data. A sensible first step for this could be a workshop where people discuss how such datasets would need to be described and shared.
- More resource papers that describe existing datasets for IIR. These could be submitted to the SIGIR resource track, SIGIR Forum or possibly to CHIIR.
- A participatory workshop for organizing and clustering tasks, starting with the database of tasks initiated by Luanne Freund and Barbara Wildemuth. This workshop could also try to prioritize the types of tasks the community should focus on.
- Domain-specific workshops for generating sets of authentic tasks for a specific domain by inviting real-world domain experts and IIR experts

4.2 Search Process

The third break-out session focused on the search process of complex search tasks and had 24 participants. The majority of the discussion in the breakout group focused on how to create a shared evaluation task that would be of interest to a wide range of IIR researchers.

The basic idea that the group came up with was to organize a shared task or workshop to determine what methods, tools, techniques are commonly used to investigate a given problem. That would give an idea of what any future, shared IIR evaluation task would have to take into account to generate interest from a broad spectrum of IIR researchers (both user and system-focused researchers).

The shared task or workshop would set a specific, complex task. It was left open how this task should be specified or obtained. Each participant would then be free to use whatever methods they wanted to use to collect data and address the task. At the workshop each participant reports and shares what methods they used, what data they collected, and what results they got. In addition to the shared dataset that the task and workshop would produce, the increased understanding of what methods are in use would also allow for the development of new tools and systems to support IIR evaluations and IIR shared evaluation tasks.

4.3 Evaluation of Complex Search Tasks

The third break-out session focused on evaluation and had 15 participants. It focused on the challenge of reducing dependence on standard datasets for shared tasks. Reliance on shared datasets risks over-fitting and takes away research efforts for contexts where data cannot (easily) be shared but deserve attention. Datasets or streams are to an increasing degree heterogenous, comprising mixed media, very various text genres, levels of abstraction and editorial oversight and structured and unstructured data. Reproducibility of research cannot rely only on identical data, but on using comparable data. As a concrete activity the group suggested a so-called *extract-a-thon*:

- Before the workshop, participants review and describe the methodology section of published papers, including how the data used are described and characterized, and assess if the information is sufficient for reproducing research.
- At the workshop, participants go through these identified characteristics, to assess if a reported experiment is based on data which is described carefully enough to be useful for the respective research and application interests. An example of the challenge would be a statement in a paper claiming that topic modeling works badly for social media, where in fact the actual problem is that the dataset is heterogeneous.
- At end of workshop, participants suggest concise ways of characterizing data that can be used across publications to fit the respective (and diverse) research interests.
- The outcome of the workshop is a check list and quality certification to share with reviewers in the field and make assessment easier and more uniform. This is somewhat similar to what is currently considered a hygiene factor for quality assurance of research in the medical field (e.g. the Cochrane Training¹).

The group noted that it is important to include ways of evaluating retrieval with mixed media collections (video, text). Text collections are easier to describe as there are more standard ways of doing this so, but the categories used may not be sufficiently fine-grained to reproduce an experiment.

5 Conclusions

Finally, the breakout groups reported to the audience with short pitches for their concrete plans for addressing these themes. The workshop provided a comprehensive overview of current work on supporting complex tasks in a variety of settings, and fostered new collaboration within our field on one of the most important topics in the coming years. Although there is a clear sense of direction emerging, it is less easy to pinpoint concrete insights or lessons learned. There was a general feeling that we made progress, and that the open discussion with participants across the fields of user-centered and system-centered information retrieval and human computer interaction was useful. There was great support for holding another edition of the workshop at a future CHIIR.

Last, but certainly not least, the workshop lived up to its proud reputation of social events, leading to new papers, spin-off workshops, and new friendships. This tradition was continued with an informal program at *Restaurant Schrøder* in Oslo, attended by workshop participants and other CHIIR attendees interested in the workshops topic, combining great discussion with food and drinks.

6 Acknowledgments

We thank the CHIIR workshop chair Preben Hansen and the local organizers for their support. We also thank the program committee members for their timely and constructive reviews. Last,

¹See <http://training.cochrane.org/>

but not least, we thank all presenters and attendees for their lively and engaged discussions. The workshop proceedings are available at <http://ceur-ws.org/Vol-1798> .

References

- [1] *Proceedings of the Second Workshop on Supporting Complex Search Tasks*, volume 1798, 2017. CEUR-WS. URL <http://ceur-ws.org/Vol-1798>.
- [2] P. Arora and G. Jones. How do Users Perceive Information: Analyzing user feedback while annotating textual units within Web Documents. In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 7–10, 2017. URL <http://ceur-ws.org/Vol-1798/>.
- [3] T. Bogers, M. Gäde, M. Hall, V. Petras, and M. Skov. Lessons Learned from the CHiC and SBS Interactive Tracks: A Wishlist for Interactive IR Evaluation. In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 11–14, 2017. URL <http://ceur-ws.org/Vol-1798/>.
- [4] Y. Egusa, M. Takaku, and H. Saito. Evaluating Complex Interactive Searches Using Concept Maps. In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 15–17, 2017. URL <http://ceur-ws.org/Vol-1798/>.
- [5] M. Gäde, M. M. Hall, H. C. Huirdeaman, J. Kamps, M. Koolen, M. Skov, E. Toms, and D. Walsh. Overview of the SBS 2015 Interactive Track. In L. Cappellato, N. Ferro, G. J. F. Jones, and E. SanJuan, editors, *Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum, Toulouse, France, September 8-11, 2015.*, volume 1391 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2015. URL <http://ceur-ws.org/Vol-1391/78-CR.pdf>.
- [6] M. Gäde, M. M. Hall, H. C. Huirdeaman, J. Kamps, M. Koolen, M. Skov, T. Bogers, and D. Walsh. Overview of the SBS 2016 interactive track. In *Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016.*, volume 1609 of *CEUR Workshop Proceedings*, pages 1024–1038. CEUR-WS.org, 2016. URL <http://ceur-ws.org/Vol-1609/16091024.pdf>.
- [7] M. M. Hall, H. C. Huirdeaman, M. Koolen, M. Skov, and D. Walsh. Overview of the INEX 2014 Interactive Social Book Search Track. In L. Cappellato, N. Ferro, M. Halvey, and W. Kraaij, editors, *Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014.*, volume 1180 of *CEUR Workshop Proceedings*, pages 480–493. CEUR-WS.org, 2014. URL <http://ceur-ws.org/Vol-1180/CLEF2014wn-Inex-HallEt2014.pdf>.
- [8] O. Hoeber, L. Hoeber, R. Snelgrove, and L. Wood. Interactively Producing Purposive Samples for Qualitative Research using Exploratory Search. In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 18–20, 2017. URL <http://ceur-ws.org/Vol-1798/>.

- [9] H. C. Huurdeman. Dynamic Compositions: Recombining Search User Interface Features for Supporting Complex Work Tasks. In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 21–24, 2017. URL <http://ceur-ws.org/Vol-1798/>.
- [10] L. Koesten and J. Singh. Searching Data Portals - More Complex Than We Thought? In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 25–28, 2017. URL <http://ceur-ws.org/Vol-1798/>.
- [11] A. Novin. Contextualizing the Cognition Crisis. In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 29–32, 2017. URL <http://ceur-ws.org/Vol-1798/>.
- [12] S. Rutter, V. Blinzler, C. Ye, M. B. Twidale, and M. L. Wilson. Complex Search Task: How to Make a Phone Safe for a Child. In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 33–35, 2017. URL <http://ceur-ws.org/Vol-1798/>.
- [13] E. Ventocilla, J. Bae, M. Riveiro, and A. Said. A Billiard Metaphor for Exploring Complex Graphs. In *CHIIR 2017 Second Workshop on Supporting Complex Search Tasks (SCST 2017)*, pages 36–40, 2017. URL <http://ceur-ws.org/Vol-1798/>.
