Report on the SIGIR Workshop on "entertain me" : Supporting Complex Search Tasks

Nicholas J. Belkin[1] **Charles L. A. Clarke**[2] **Ning Gao**[3]
Jaap Kamps[4] **Jussi Karlgren**[5]

[1] Rutgers, USA
[2] Waterloo, Canada
[3] Peking University, China
[4] University of Amsterdam, The Netherlands
[5] SICS Stockholm, Sweden

Abstract

Searchers with a complex information need typically slice-and-dice their problem into several queries and subqueries, and laboriously combine the answers post hoc to solve their tasks. Consider planning a social event at the last day of SIGIR, in the unknown city of Beijing, factoring in distances, timing, and preferences on budget, cuisine, and entertainment. A system supporting the entire search episode should "know" a lot, either from profiles or implicit information, or from explicit information in the query or from feedback. This may lead to the (interactive) construction of a complexly structured query, but sometimes the most obvious query for a complex need is dead simple: `entertain me`. Rather than returning ten-blue-lines in response to a 2.4-word query, the desired system should support searchers during their whole task or search episode, by iteratively constructing a complex query or search strategy, by exploring the result-space at every stage, and by combining the partial answers into a coherent whole.

The workshop brought together a varied group of researchers covering both user and system centered approaches, who worked together on the problem and potential solutions. There was a strong feeling that we made substantial progress. First, there was general optimism on the wealth of contextual information that can be derived from context or natural interactions without the need for obstrusive explicit feedback. Second, the task of "contextual suggestions"—matching specific types of results against rich profiles—was identified as a manageable first step, and concrete plans for such as track were discussed in the aftermath of the workshop. Third, the identified dimensions of variation—such as the level of engagement, or user versus system initiative—give clear suggestions of the types of input a searcher is willing or able to give and the type of response expected from a system.

1 Introduction

There is a striking difference in how we ask another (unknown) person for information, giving a lot of context information and very precisely articulating what we want and why, and how we communicate with current search engines, typically with a series of short queries. Can't we do better? Think about a novel information access tool that actively supports a searcher to articulate a whole search task, and to interactively explore the results of every stage of the process. That is, can we support the entire search episode?

The leading example of the workshop was ourselves at the last day of SIGIR 2010 in Beijing, in a city we don't know, trying to plan our post-workshop evening (locating suitable restaurants, theaters, clubs, looking at reviews, locations and distances, individual preferences, time-tables, etc.). This could be formulated as a highly complex query or search strategy, that could be interactively constructed based on an initial plan and further feedback. Eliciting such a query from a searcher would require complex interaction, and an expressive query language combining several constraints on content as well as on structure (i.e. collection structure and annotation). However, the most natural query for this is in fact dead simple: `entertain me`. That is, the most natural way to express this highly complex information need is utterly simplistic, and all the other needed information could be filled in from implicit and explicit contextual information on the specific search request, the specific location, time, and IP, the searcher and her preferences, etc.

The overall goal of the workshop could be succinctly summarized as to make IR systems support searchers during their entire search episodes when interactively solving a complex task, such as the `entertain me` planning problem. Although a SIGIR Workshop devoted to a single query may seem extravagant, this query is just one example of the general problem of supporting simple and common requests that express complex and dynamic needs. In many modern search scenarios such as mobile apps or search verticals, the information derived from context is just as important as the query itself.

As an experiment within the workshop, the organizers were indeed planning the social event with the currently available tools: searching for on-line information (reviews, locations, cuisines, and prices), consulting travel guides and pinging friends and locals for recommendations, elaborate phone talks to restaurants about making reservations, etc. This turned out to be a more laborious task than we ever imagined—with traditional pen-and-paper notes proving essential to connect all the information. Following the proud IR tradition to subject all experiments to rigorous evaluation, the resulting plan was thoroughly "evaluated" by the workshop participants during a long evening in Beijing.

2 Workshop

We brought together a varied group of researchers covering both user and system centered approaches, who worked together on the problem and potential solutions, and on identifying the *barriers* to success and ways of addressing them. The workshop had a format that emphasized interaction—after all it was a *work*shop.

2.1 Many Open Questions

The workshop started with many open questions:

- *Complex Search Episodes* What complex information needs are currently lacking support? Which types of tasks are within our grasp? How important is transparency, trust, or authority? Will searchers accept "smart" systems? What benefits are to be had? How to generalize?
- *Eliciting Needs* What would the system need to "know"? How much information can be obtained from profiles, context, and implicit feedback? What would require explicit feedback and interaction? How much effort are searchers willing to spend? How to deal with evolving needs?
- *Complex Queries* How to articulate complex queries? With shallow queries there is little benefit in collection structure or semantic annotation; what complex queries or search strategies are needed to support complex tasks? How to construct such queries interactively?
- *Exploratory Search* How to explore (intermediate) results? How to refine the queries by drilling down, or collect partial answers by exploring and zooming out? How to combine partial answers into a comprehensive whole?
- *Understanding Context and Target Notions* How to integrate geographic, temporal, personal and social context? How can we measure and evaluate the result of working towards a complex and many-faceted information need?
- *Anatomy of a System* What components are needed? Which are currently available? What more is needed? What is the road map for realizing systems capable of supporting search episodes?

The questions can be roughly grouped into three themes: the domain angle (applications and use-cases), the searcher's angle (eliciting needs and complex queries interactively), and the system's angle (incorporating task and searcher context into a retrieval system).

2.2 Format

We started the day with a short introduction of the goals and schedule, and a "feature rally" in which each participant introduced her- or himself, and stated her or his particular interest in this area.

Next, we had a keynote that helped frame the problem, and create a common understanding of the challenges. Jussi Karlgren (SICS) discussed which information access use cases could be said to share a family likeness along the lines of a simple query, a complex information access task, a vast outcome space, and inspecific and complex success criteria more geared towards satisficing than optimising the fulfilment of information need.

We continued with short presentations of papers, roughly grouped in sessions related to common themes:

Domain Complex Search Needs and Use-Cases;

Searcher Eliciting Complex Needs and Queries; and

System Task Context and Success.

Most papers addressed multiple aspects, which complicated the grouping into sessions but greatly facilitated the discussion at the workshop.

Starting over lunch, we had three break-out sessions in parallel that focused on specific aspects or problems related to the three themes.

After the afternoon coffee, we had reports of the breakout sessions, followed by a final discussion on what we achieved during the day and how to take it forward.

3 Accepted papers

We requested the submission of short, 2 page papers to be presented as boaster and poster. We accepted a total of 11 papers and loosely grouped the papers in three themes:

3.1 Complex Search Needs and Use-Cases

Karlgren [10] was an invited keynote discussing use-cases of complex tasks bases on a single query. There may be a range of scenarios where a single, simple query is the most appropriate initiating action: due to a lack of knowledge (e.g., traditional "ASK"), to a lack of commitment or investment will (e.g., a lack of engagement), a lack of specificity or aspiration (e.g., satisficing behavior), or to a lack of bandwidth (e.g., mobile devices).

Choi et al. [4] discussed evidence-based practice, a set of procedures and guidelines on what constitutes the best available evidence for (typically partly subjective) clinical decisions. For tasks involving literature review, this makes distinct requirements on both the search results (in terms of relevance and usefulness to the task at hand) and on the search process that should be conducted in a prescribed way.

Davies and Bland [6] discussed nonstandard access methods for audio-visual material in broadcast archives, focusing on affective classification of TV programmes in the massive archive of the British Broadcasting Corporation (BBC) for entertainment reuse. Whilst traditional use of these archives is based on formal metadata descriptions of the entire program, modern use of them in the BBC's iPlayer is requires complex and flexible navigation, such as access to individual segments, skip forward to salient events, affective annotation, or recommendation, all in response to a simple query or button.

Karimi and Scholer [9] also discussed evidence based practice and policy, focusing on "systemic reviews" which are documents that synthesize available research on the topic of investigation. Writing a systemic review is a complex task that resembles topical information need, but requires a multistage search and selection process to ensure high recall. First, expert librarians select large pools of documents from medical databases. Second, all summaries are studied by the investigators to verify the review inclusion criteria. Third, the selected literature is analyzed to generate the final set of documents to be included in the systemic review.

3.2 Eliciting Complex Needs and Queries

The following paper was not presented at the workshop, since none of the authors was able to attend: C. Chen. Articulating information needs by user profile enrichment. In Belkin et al. [2], pages 9–10.

Lingnau et al. [11] discussed the "show-and-tell" system for children's interactive search. Show-and-tell is based on a book metaphor, allowing young children to explore a topic. The task is initiated by an object representation such as an image (which becomes the book cover), and textual or visual information found on various aspects of the topic can be collected into a virtual booklet, thereby creating their own narrative — with the narration providing a vehicle

whereby an inexperienced reader can be guided through a problem space and empowered to relate the topic to other readers.

Tang et al. [12] discussed interactive question answering as a natural means to clarify the searcher's intent. The paper focuses on generating the right follow-up questions that help refine or disambiguate the initially formulated queries. Experiments based on HowNet (the Chinese version of WordNet) demonstrated the potential of the approach, while also highlighting the complexities of semantics and pragmatics involved in the dialogue.

Yuan and Belkin [13] discussed spoken language interfaces, and how the change of modality affects the way in which we naturally communicate with systems. In particular, there is an unnatural aspect to textual input—certainly in languages like Chines in on resource-bounded mobile devices—leading to short statements of request. Speech and gesture interfaces can alleviate some of these problems, when suitably prompted, and allow for supporting complex tasks in a natural way.

3.3 Task Context and Success

Azzopardi [1] discussed a particular genre where complex tasks are based on simple query statements: search tasks related to sex ranging from looking for romcoms, to finding a date, to downloading adult content. These tasks constitute a significant fraction of on-line activity but receive very little to no attention in the literature, and the paper is a first step in classifying the different types of information needs in this genre. Charlie Clarke presented the paper on behalf of the author who was unable to attend due to visa problems. In the days prior to the workshop, the paper and presentation (reputedly containing vivid illustrations) raised considerable controversy. Hence it was decided to only present the initial slides and main idea, which received broad support from the audience. The discussion after the presentation noted that the type of interaction most typically associated with visual adult content is different from topical ad hoc search. It was postulated that it is a broad and general query generating a large retrieved set of visual material which is accessed linearly in a much more persistent manner than other queries. This is not specific to adult content but to other visually oriented queries related to various hobbies or pastimes: wildlife, ornithology, vehicles, historical images of various types.

Clarke and Song [5] discussed contextual suggestion systems—a novel mixture of search engines and recommender systems—catering for suggestions that take the searcher's context into account, that are sufficiently novel related to the other suggestions, and highlight the contrastive elements of each suggestion. Promising initial experiments on contrastive captioning for a "dining in Beijing" scenario are reported. This paper engendered a debate as to what the distinction between a recommendation system and the systems under consideration in the workshop would be. The participants agreed that the distinction would be difficult to make but that it hinged to some extent on what the end result of the interaction would be. A recommender system provides the user a small and limited list of items whereas the desired system needs to trace a path through a complex outcome space, with the need to explain "you should do this because..."

Efron and Organisciak [7] discussed an approach based on a palette mixing metaphor for highly contextual queries. Many such needs are solved by relying on multiple information sources (such as mobile apps or search verticals). These lower level interactions are combined to solve the overall task with a high level information seeking strategy that resembles palette mixing. The identification of this distinct information seeking models, immediately leads to

a number of useful observations on ways of supporting it.

Gossen et al. [8] discussed the evaluation of exploratory search systems for complex needs. While evaluation methods for system effectiveness (TREC/Cranfield experiments) as well as for usability aspects (controlled lab or longitudinal studies) are available, they are based on radically different principles and are difficult to combine. The creation of a benchmark is challenging, in particular when creative aspects of exploration need to be simulated in a realistic way in order to guarantee a reusable benchmark.

4 Breakout Sessions

The lively discussion of the paper sessions continued in three breakout groups each discussing a particular aspect of the problem in a forward looking way.

4.1 Eliciting Complex Needs and Queries

Ian Ruthven (Strathclyde) chaired a breakout group on "Eliciting Complex Information Needs." Focusing on complex needs that can be expressed by simple queries, the key question is what information we can infer and do not have to ask the searcher to enter explicitly. There are at least three distinct levels of context: location (where am I?), personal (who am I?), and task state (what am I doing or about to do?). The context information can be used both for understanding the precise task at hand as well as for filling in specific details. Profiles from earlier interactions (both personalisation or customization) can be used to derive priors on the likelihood of particular scenarios, estimating unknown values, or recommendations. Supporting such a task will necessarily be interactive and resembling a dialogue: requiring user engagement and selection, employ rich interfaces (not question based), and need a clear benefit to the user (transparency). Serendipitous search requires elements of surprise or unpredictability. At starting point might be the mood of the searcher. This may be inferred from sensors in mobile devices, or the query—is "entertain me" effectively saying "I am bored"? Especially personal mobile devices such as smartphones or tablets PCs seem the obvious area of application, being characterized by having extensive contextual information about the person, her location, app preferences, etc., and by a limited capacity to enter explicit information due to a lack of full keyboard and usage under rough conditions.

There is no generic model of this type of interaction yet, but it is clear that there are lots of possibilities. A fruitful direction would be to classify the types of complex needs and how these should be supported (relative to the traditional concept of relevance). The general conclusion was that the breakout group identified many elements of contextual information that could, in principle, be derived without the need for extensive explicit feedback. Hence, the elicitation of complex needs seems to be within our grasp.

4.2 Task Context and Success

Charlie Clarke (Waterloo) chaired a breakout group on "Evaluating Complex Search Tasks." The breakout discussed general approaches to evaluation and their pros and cons: based on explicit user feedback (e.g., user studies and surveys), based on implicit feedback (e.g., click logs), or based on editorial judgments (e.g., traditional test collections). Different evaluation approaches make a different trade-off in terms of realism versus generality and reusability.

Most time was devoted to trying to define a task suitable for a traditional test collection based on editorial judgments. This led to a proposal for a track based on contextual suggestions, where the need is expressed by a simple, non-informative query like `entertain me` (or a specific app on a mobile device) in combination with a range of contextual information. E.g. a topic statement like:

- 26 year old woman in toronto; 11:00 in the evening; summer; likes dancing; skydiving; indian food and romantic movies

This proposal, informally referred to as a "Dating Track," was regarded as an interesting combination of search and recommendation. In its simplest form it can be evaluated as a ranked list of suggestions, each of them judged individually against the topic statement (a highly personal profile). In future extensions, a comprehensive plan could be derived also taking into account dependencies between the results, such as common-sense restrictions (e.g., one diner is quite enough, but multiple bars might be appropriate) and logistics (e.g., the movie should start after diner, this great restaurant isn't worth a 6 hour drive, skydiving is beyond the budget). Discussion on a contextual suggestion track continued in the aftermath of the workshop, with the Lonely Planet Guides or a dedicated crawl of travel sites as a suggested corpus.

4.3 Complex Search Needs and Use-Cases

Jussi Karlgren (SICS) chaired a breakout group on "Variational Dimensions for Single Query Information Access Applications and Use Cases." The objective was to identify commonalities and systematic differences between the example interaction scenarios presented at the workshop. Some of the variational dimensions discussed were potentially definitional, categorising use cases—other more continuous describing a systematic variation across many different types of usage. The objective was to bring together the identified dimensions of variation and to relate them to success criteria and evaluation schemes, both with respect to benchmarking and validation of usage hypotheses.

Some of the dimensions of variation discussed in the group included *user engagement* with some envisioned systems being geared towards lean-forward and others more towards lean-back interaction; *user vs system initiative*, where some systems might push information to users, others wait for the users to request it; systems might vary according to *user expertise*, *topicality vs appeal*, *persistence* or *complexity* of results, and numerous session-oriented factors such as learning curves, evolution of information need and infrastructural characteristics. The general question of demarcation of the systems under consideration against recommendation systems and general case exploratory interfaces was brought up, with respect to the dimensions of variation.

As an example excercise some of the systems discussed in session were traced through some of the variation dimensions on a flipchart (Couch Potato, BBC Video Archive [6], Medical Review [4], and Show & Tell [11]), to tease out differences between various usage scenarios answering to the workshop topic.

5 Conclusions

The results of the breakout groups, as discussed in Section 4 above, were presented to the workshop in the final plenary session chaired by Mounia Lalmas (Yahoo! Research). There

was a strong feeling that we made substantial progress. Specifically, each of the breakout groups contributed to our understanding of the way forward. First, there was general optimism on the wealth of contextual information that can be derived from context or natural interactions without the need for obtrusive explicit feedback. Second, the task of "contextual suggestions"—matching specific types of results against rich profiles—was identified as a manageable first step, and concrete plans for such as track were discussed in the aftermath of the workshop. Third, the identified dimensions of variation—such as the level of engagement, or user versus system initiative—give clear suggestions of the types of input a searcher is willing or able to give and the type of response expected from a system. More generally, the workshop discussed what sort of family likeness could be found between the cases discussed—ranging from the most prototypical `entertain me` over variants related to recommender systems and `heal me` to systems for establishing a scientific basis for research work in a more precision oriented task setting. The discussion centered on features such as (1) complexity of the outcome space, (2) the challenge for users to establish concrete success criteria for the interaction (as opposed to e.g. known item search situations), and (3) the possibility of the interaction evolving during a session or over a sequence of sessions, if a changed knowledge state of the user would be taken into account. In many of the cases discussed, serendipity and a certain level of unpredictability of results was viewed as being desirable, whereas in others—such as in a scientific setting—this would be viewed as detrimental. In many of the cases, again, not unrelatedly, relevance was viewed as less useful as a target notion for evaluating system performance.

As to the character of the workshop, the participants voiced general broad support for the workshop's interactive character and the group discussions, and how this perfectly complemented the more formal presentations during the SIGIR conference. In particular, in the discussion insights from user-oriented researchers in information science naturally combined with those of system-oriented researchers in computer science.

Last, but certainly not least, the workshop continued far into the night with drinks in the Sanlitun (三里屯) district and a splendid dinner at the YuanYuan (圆苑) restaurant—with 25 researchers from 20 different countries—with further, even more intense, discussion about complex tasks with simple queries and (scientific) life in general. This then seamlessly continued into further drinks and talks with a smaller group during the night in downtown Beijing...

Acknowledgments We would like to thank the ACM and the SIGIR for hosting this workshop, in particular Noriko Kando, Luo Si, and JianYun Nie for their outstanding support in the organization. Special thanks to Jimmy Lin for his outstanding negotiating skills during the social event—giving us an amazing mileage out of the available budget. We would also like to thank the program committee: Gene Golovchinsky, Donna Harman, Noriko Kando, Evangelos Kanoulas, Gabriella Kazai, Diane Kelly, Mounia Lalmas, Birger Larsen, Ian Ruthven, Falk Scholer, Mark Smucker, Anastasios Tombros, Daniel Tunkelang, Ryen White, and Max Wilson and the five program chairs. Final thanks are due to the paper authors, and the participants for a great and lively workshop. Details about the workshop including the presentations and slides are online at `http://staff.science.uva.nl/~kamps/entertainme/`.

References

[1] L. Azzopardi. Searching for unlawful carnal knowledge. In Belkin et al. [2], pages 17–18.

[2] N. J. Belkin, C. L. A. Clarke, N. Gao, J. Kamps, and J. Karlgren, editors. *Proceedings of the SIGIR'11 Workshop on "entertain me" : Supporting Complex Search Tasks*, 2011. ACM Press.

[3] C. Chen. Articulating information needs by user profile enrichment. In Belkin et al. [2], pages 9–10.

[4] S. Choi, B. Ryu, S. Yoo, and J. Choi. Applying metasearch to medical literature retrieval for evidence-based medicine. In Belkin et al. [2], pages 3–4.

[5] C. Clarke and W. Song. Why is this restaurant different from all other restaurants? (captioning for contextual suggestion). In Belkin et al. [2], pages 19–20.

[6] S. Davies and D. Bland. Affective classification of large scale broadcast archives. In Belkin et al. [2], pages 5–6.

[7] M. Efron and P. Organisciak. A palette mixing model of information seeking for complex queries. In Belkin et al. [2], pages 21–22.

[8] T. Gossen, S. Haun, and A. Nuernberger. How to evaluate exploratory user interfaces? In Belkin et al. [2], pages 23–24.

[9] S. Karimi and F. Scholer. Systematic reviews: A complex search episode for evidence based policy and practice. In Belkin et al. [2], pages 7–8.

[10] J. Karlgren. The use case perspective for single query information access (invited). In Belkin et al. [2], pages 1–2.

[11] A. Lingnau, I. Ruthven, and M. Landoni. Show and tell: supporting childrens search by interactively creating stories. In Belkin et al. [2], pages 11–12.

[12] Y. Tang, F. Bu, Z. Zheng, and X. Zhu. Towards interactive qa: suggesting refinement for questions. In Belkin et al. [2], pages 13–14.

[13] X. Yuan and N. Belkin. Supporting complex tasks in a spoken language interface. In Belkin et al. [2], pages 15–16.