

Report on the Second International Workshop on the Evaluation on Collaborative Information Seeking and Retrieval (ECol'2017 @ CHIIR)

Leif Azzopardi¹, Jeremy Pickens², Chirag Shah³, Laure Soulier⁴, Lynda Tamine⁵

¹ School of Computing Science, University of Glasgow - UK *Leif.Azzopardi@glasgow.ac.uk*

² Catalyst Repository Systems - USA *jpickens@catalystsecure.com*

³ Rutgers University - USA *chirags@rutgers.edu*

⁴ Sorbonnes Universités - UPMC - LIP6 - France *laure.soulier@lip6.fr*

⁵ Université de Toulouse UPS IRIT - France *tamine@irit.fr*

November 2, 2017

Abstract

The 2nd workshop on the evaluation of collaborative information retrieval and seeking (ECol) was held in conjunction with the ACM SIGIR Conference on Human Information Interaction & Retrieval (CHIIR) in Oslo, Norway. The workshop focused on discussing the challenges and difficulties of researching and studying collaborative information retrieval and seeking (CIS/CIR). After an introductory and scene setting overview of developments in CIR/CIS, participants were challenged with devising a range of possible CIR/CIS tasks that could be used for evaluation purposes. Through the brainstorming and discussions, valuable insights regarding the evaluation of CIR/CIS tasks become apparent – for particular tasks efficiency and/or effectiveness is most important, however for the majority of tasks the success and quality of outcomes along with knowledge sharing and sense-making were most important – of which these latter attributes are much more difficult to measure and evaluate. Thus the major challenge for CIR/CIS research is to develop methods, measures and methodologies to evaluate these high order attributes.

1 Introduction

The paradigm of Collaborative Information Seeking (CIS) and Retrieval (CIR) refers to methodologies and technologies that support collective-knowledge sharing within a work team in order to solve a shared complex problem [5]. Accordingly, the main underlying challenge is to satisfy the mutual beneficial goals of both individual users and the collaborative group while minimizing the cognitive effort associated with the collaborative interactions [12]. Indeed, collaborative search is also known as a social process [4] in which users leverage from other users' interactions and social signals (e.g., bookmarks and annotations). In the recent years, several studies [8, 14, 9] have been carried out to understand the behavioral process of

users in terms of question asking on search engines vs. social networks. In this context, social platforms (e.g., social networks [10, 2] and community question-answering [7, 3]) have been acknowledged as a place where users collaborate to solve an information need. This provides numerous opportunities for new and novel research within the field of CIR/CIS where the gap between social and collaborative search can be brought together.

However, the evaluation of CIS/CIR and social & collaborative IR models means dealing with several layers of complexities and a variety of confounding factors such as the multi-user and multi-level contexts, the exploratory aspect of the search through multi-session search activities, the multiplicity of relevance factors, the individual vs. collective value of relevance, the search interfaces supporting the collaborative interactions as well as the social interactions themselves. While, substantial research advances in the evaluation of non-collaborative information retrieval and seeking tasks have been achieved through international evaluation campaigns such as TREC, CLEF and NTIR, to date and to our knowledge, no standardization effort has been achieved for the evaluation of CIS/CIR and social-CIR.

During the first ECol Workshop there was strong consensus that there is an important need to investigate the evaluation challenge in CIS/CIR/social-CIR with the aim of creating common evaluation frameworks that would foster the research area. A first edition of this workshop was organized at CIKM 2015 [1] in which we discussed the differences in evaluation between standard Information Retrieval tasks and CIS/CIR tasks [13]. The second edition of the ECol Workshop¹ at CHIIR 2017 [13] focused discussion on tasks for evaluation with the aim to develop an evaluation framework and set of tasks for CIS/CIR and social-collaborative IR.

2 General Introduction

Chirag Shah kicked off the discussions with a general overview of CIR/CIS entitled, “Collaborative Information Seeking. A brief history” [11]. During this talk, Chirag presented an overview of social and collaborative information seeking frameworks, particularly from the search iteration side in which user-to-user (social) interaction dimensions are prevalent. The state-of-the-art of collaborative information seeking approaches have been reviewed and differentiated: ethnographic, passive support, and active support. The evaluation issue has been addressed from both the methodologies and measurements perspectives. The talk also highlighted the ideal domain-applications of collaborative information seeking (legal, education, health, intelligence analysis, etc.) and pointed out the remaining challenges and relevant opportunities in the research area. Chirag then provided an overview of evaluation metrics adapted for collaborative information seeking. Finally, he summed up the main scientific events dealing with the topic and a roadmap of:

1. what we currently know: usefulness of collaboration, collaboration cost, issues related to awareness, privacy, control, transparency, communication as well as coordination, and,
2. what should be mainly investigated in the future: fundamentals of social and collaborative information seeking, including social aspects in IR, satisfying the collective relevance.

¹<https://www.irit.fr/ECol2017/>

The introduction session was highly interactive and turned into a fruitful discussion on the challenges and limitations of evaluating CIR/CIS.

3 Collaborative Work/Search Tasks

After the break, the participants reconvened to focus on discussing various contexts and settings for collaboration. For example, the people or agents involved in the collaboration, i.e. friends, colleagues, group members, Siri, Google, etc. And their relationship between each person and their goals in the collaboration. For example, in [6], the authors presented an example where analysts, while co-workers, have different levels of access to the collection, and are trying to collect intelligence, but are often not aware of the end-goal, just what is considered relevant or not. For example, when friends are searching for a holiday destination, while certain criteria are desired, often compromises need to be made during the process. And so, the time taken is not necessarily the driving factor, but the overall satisfaction of the decision made. Here, information sharing becomes quite important in order to form a consensual decision, and to understand the compromises that are being made. However, in tasks where the search task can be broken down into independent, discrete units of work, for example, assessing legal documents for disclosure purposes, then the task could be made more efficient by a divide and conquer approach. The different scenarios lead to different evaluation goals:

- *further*: find more,
- *awarer*: know more,
- *faster*: perform the task faster, and,
- *better*: make a better decision.

It was felt that the latter tasks, which were divisible, were seen as less challenging because it boiled down to efficiency. While the former ones would be more difficult – and required people to negotiate and interact - and to do so in situations where there maybe information asymmetry (expert vs novice, lead lawyer vs junior lawyers, etc). This lead us to consider the following questions:

- RQ1: How to build effective collaboration while searching? What are the impacting factors?
- RQ2: How to evaluate collaboration in the case of complex search tasks, e.g. learning tasks, decisional tasks?

Dealing with the first research question, participants argued that the quality of the collaboration depends on:

- the motivation of collaborators. Depending on motivations, depends on how much effort they put in, and also how much they get out of the collaborative task, and so both individual outcomes and group outcomes need to be considered.
- the asymmetry between participants in terms of role, resource, skill and knowledge. More specifically, depending on what people know, do not know, and believe to know (or what they believe others to know) can have an impact on success. This gives rise to difficulty in leading an effective search and coordinating people with respect to what they are able to accomplish.

Tasks	Evaluation				
	<i>Usability</i>	<i>Understandability</i>	<i>Correctness</i>	<i>Efficiency</i>	<i>Outcome Quality</i>
Report writing	✓	✓	✓		
Programming challenges			✓	✓	
Journalist (fact-finding, comparative analysis)	✓	✓	✓		
Transcription, annotation		✓	✓		
Interview, pitches	✓	✓		✓	
Market research (competitiveness, state of the art)	✓	✓			✓
Building briefs	✓	✓			
Make a lecture		✓	✓		✓
Make a wiki page	✓	✓			✓
Advanced obituary (biography)		✓	✓		✓

Table 1: Evaluation criteria of educational tasks willing to be performed in a collaborative environment.

- the trust between collaborators was also seen to be an important factor. Specifically the above factors and the prior working relationship are likely further impact the collaboration and its success.

Since the different search/work tasks can be so varied and complex, the second stage of the discussion focused on “collaborative learning” tasks specifically. This was seen as a relevant task as most participants teach courses and have students undertake collaborative assignments. To this end, we enumerated a number of different learning activities assigned as tasks, (see Table 1 for a non-exhaustive list of tasks and their related issues) and then discussed the different factors that were important in terms of evaluation. More particularly, we identified the following categories of criteria:

- Correctness aims at measuring the veracity and accuracy of the information found and produced, w.r.t the specified need. In addition an output, such as a biography, or news report, may also need to be created in a specific structure or format, and address particular questions.
- Efficiency aims at measuring the time taken and the speed at which the task can be performed, and how the collaboration leads to higher rates of gain (or not), and/or minimizing the load between collaborators. For example, programming challenges often require collaborators to take on different roles, depending on their expertise, but if the collaboration is to be efficient, all collaborators need to contribute (otherwise they would be idle).

-
- Understandability aims at measuring how well the team can make sense of the information obtained, found, and produced during the collaboration. This, of course, assumes that all the goal is that all participants learn from the process. In the context of a learning task, though, it is typically desirable for all participants to have gained new knowledge in the process and to understand the sum (and synergy) of their inputs.
 - Usability aims at measuring to what extent the output is easy to use given the specified objectives. For instance, a wiki page should enable the user to easily read and learn information, while the code from a programming assignment should be easy to read, and execute.
 - Outcome Quality aims at measuring how good the decision, the product, and/or output of the learning task was, given the collaboration. For instance, a market research report would lead to qualitative outputs that would inform what kind of strategy to pursue, or the code from a programming challenge would solve the problem correctly, efficiently and cleanly.

The motivation of this discussion relies on the fact that we are good at measuring IR systems that rank documents, however most collaborative tasks need specific evaluation frameworks that consider the higher level goals and the outputs associated with them. A global challenge to all of these tasks involved how we evaluate the efficacy, efficiency and effectiveness of the communications between participants and how that affects the task outcomes and success.

4 Summary

The second edition of the ECol workshop provided a comprehensive overview of current research work on collaborative information retrieval and seeking, perceived as an emerging and important topic in the short-term. The discussion focused on characteristics leading to an effective collaborative search, where we focused in on the application domain of learning, to deepen our understanding within this particular use case. This led to valuable insights regarding the evaluation of CIR/CIS tasks in the learning context. For particular tasks, efficiency and/or effectiveness is most important, however for the majority of tasks the understandability, usability and quality of outcomes along with knowledge sharing and sense-making were seen as more important. However, these latter attributes were considered much more difficult to measure and evaluate. Thus the major challenge for CIR/CIS research is to develop methods, measures and methodologies to evaluate these high order attributes.

References

- [1] *ECol '15: Proceedings of the 2015 Workshop on Evaluation on Collaborative Information Retrieval and Seeking*, New York, NY, USA, 2015. ACM.
- [2] D. Castilho, P. Melo, D. Quercia, and F. Benevenuto. Working with friends: Unveiling working affinity features from facebook data. In *ICWSM*, 2014.
- [3] S. Chang and A. Pal. Routing questions for collaborative answering in community question answering. In *ASONAM*, pages 494–501. ACM, 2013.

-
- [4] B. M. Evans and E. H. Chi. An elaborated model of social search. *IP&M*, 46(6):656–678, 2010.
 - [5] J. Foster. Collaborative information seeking and retrieval. *Annual Review of Information Science and Technology*, 40(1), 2006.
 - [6] N. N. Htun, M. Halvey, and L. Baillie. An interface for supporting asynchronous multi-level collaborative information retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, CHIIR '17*, pages 225–234. ACM, 2017.
 - [7] B. Li and I. King. Routing questions to appropriate answerers in community question answering services. In *CIKM*. ACM, 2010.
 - [8] M. R. Morris, J. Teevan, and K. Panovich. What do people ask their social networks, and why?: A survey study of status messages. In *CHI*, pages 1739–1748, 2010.
 - [9] A. Oeldorf-Hirsh, M. R. Morris, J. Teevan, and G. Darren. To search or to ask: The routing of information needs between traditional search engines and social networks. In *ICWSM*, 2014.
 - [10] S. A. Paul, L. Hong, and H. Chi. Is twitter a good place for asking questions? a characterization study. In *ICWSM*, 2011.
 - [11] C. Shah. Social and collaborative information seeking: State of the union. In *Proceedings of the 2015 Workshop on Evaluation on Collaborative Information Retrieval and Seeking, ECol '15*, pages 1–1, New York, NY, USA, 2015. ACM.
 - [12] C. Shah and R. González-Ibáñez. Evaluating the synergic effect of collaboration in information seeking. In *ACM SIGIR*, pages 913–922, 2011.
 - [13] L. Soulier, L. Tamine, T. Sakai, L. Azzopardi, and J. Pickens. Report on the first international workshop on the evaluation on collaborative information seeking and retrieval (ecol'2015). *SIGIR Forum*, 50(1):42–48, 2016.
 - [14] J. Teevan, D. Ramage, and M.-R. Morris. Twitter search: A comparison of microblog search and web search. In *WSDM*, pages 35–44, 2011.