Workshop on Social and Collaborative Information Seeking (SCIS)

Chirag Shah
Rutgers University
chirags@rutgers.edu

Rob Capra
UNC Chapel Hill
rcapra@unc.edu

Preben Hansen
Stockholm University
preben@dsv.su.se

Abstract

This is a report from the Workshop on Social and Collaborative Information Seeking (SCIS), held on May 14-15, 2015 at Rutgers University. Twenty-eight participants contributed to the workshop in the form of short talks and work sessions. Outcomes from the workshop include a "research roadmap" of important future research needed in the area of social and collaborative search. The report presented here summarizes the short talks and the discussions that took place in six breakout groups. The report ends with pointers for future directions (research roadmap) and a few take-away points.

1 Introduction

Increasingly, people are utilizing collaboration and sharing technologies to address needs in their work and personal lives. Information plays a key role in today's world, and many problems require us to use social and collaborative ties to search for and locate information. Examples range from corporate teams doing business intelligence gathering to a couple planning their vacation to a diabetes patient looking for information and support regarding treatment options.

Recently, researchers in the fields of information and computer science have been studying how people work in social and collaborative situations to search for information, and how information systems can support these needs. Innovative research has resulted in new tools and services for social/collaborative information seeking (SCIS), and the development of systems for studying social/collaborative search behaviors. However, research to support collaborative search is still young, and there are many challenges to be addressed. These include creating suitable data collection and analyses methods, constructing new evaluation frameworks, and developing integrated systems that incorporate people's social and collaborative behaviors.

This workshop was sponsored by the Center for Discrete Mathematics & Theoretical Computer Science (DIMACS) at Rutgers University, as part of their focus on Information Sharing and Dynamic Data Analysis, funded by the National Science Foundation. The workshop brought together scholars from a variety of disciplines and backgrounds who are experts and innovators in studying collaborative search systems, with a goal of outlining a "research roadmap" of challenges and

opportunities as an outcome of the workshop. For one and a half days, these scholars presented, discussed, debated, and synthesized ideas related to social and collaborative information seeking theories and applications.

This report is a brief summary of the workshop. A longer, more detailed workshop report (over 50 pages) is available at: http://dimacs.rutgers.edu/Workshops/SCIS/SCIS2015Report.pdf.

2 Summary of the Talks

After an introduction to the workshop, 18 participants gave short presentations on their current areas of research. These presentations are summarized below.

Chirag Shah (Rutgers University) opened the session by talking about the importance of space and time as dimensions in social and collaborative information seeking, and highlighting the additional roles of communication, awareness, affects, and group sizes. Next, Michael Twidale (University of Illinois at Urbana-Champaign) talked about how technology learning can often be both a search activity as well as a social activity, and that knowledge about strategies and tactics can have large impacts. Roberto González-Ibáñez (Universidad de Santiago de Chile) then talked about the importance of affective processes such as emotions in collaborative information seeking. Doug Oard (University of Maryland, College Park) discussed how challenges in cross-language information retrieval (CLIR) could help inform the research agenda for CIR, and presented as an example a collaborative translation task with people having different skills (e.g., some collaborators knowing the source language, others knowing the target language). Following this, Soo Young Rieh (University of Michigan) proposed developing a set of evaluation measures for social search divided into four categories: performance measures, informational outcomes, social outcomes, and user experience.

Christopher Leeder (Rutgers University) presented results from a user study that showed that groups found more useful sources and greater information coverage while working together. In contrast, individuals did better regarding query effectiveness and the amount of relevant sources. Yinglong Zhang (University of Texas) argued that the success of collaborative work is largely based on whether members can trust each other in a collaborative group. He also discussed how culture is one of many important factors that may influence the development of trust in collaboration. Aiko Takazawa from University of Illinois at Urbana-Champaign discussed a case study involving seven Japanese women living in Finland who created an ad-hoc, self-organized humanitarian aid group in response to the 2011 Great Tohoku Earthquake and Tsunami disaster in Japan. She reported how they collaboratively searched and used information with available technologies, and highlighted the role of social media in the group's work.

Robert Capra (University of North Carolina at Chapel Hill) discussed how systems could support users' activities in collaborative search, including planning, communicating results, monitoring progress, creating shared representations of structure, and performing synthesis of findings. He also described his recent research to investigate benefits of showing the "search trails" of previous users to future searchers who are doing similar tasks. Following this, Daqing He (University of Pittsburgh) talked about context-sensitive supports for collaborative information retrieval (CIR). He argued that it is important to take into account collaborators' search histories in a team. Sandra Toze (Dalhousie University) explored needs at the group-level within social and collaborative search activities. She presented a conceptual model of Group Information Process that provides a base that can be used for

further research. Next, Jeremy Pickens (Catalysts Inc.), talked about explicit task-constrained collaborative information seeking and discussed the importance of role asymmetries in certain domains, describing how the information needs are not always jointly negotiable between different collaborative partners when solving a task. Following this, Simon Knight (Open University, UK) focused on collaborative information seeking tasks as complex performance assessments. In order to study this, he developed two types of tasks: collaborative information seeking, and collaborative multiple document processing.

Mark Ackerman (University of Michigan) talked about collaborative information access in the health domain. He described the long-term information needs and information behaviors of people with chronic medical conditions, and talked about several projects to help people sense-make their conditions and the information they are gathering. Kaitlin Costello (University of North Carolina at Chapel Hill) also discussed health information seeking, focusing on how patients may teach other patients how to evaluate health information in online support groups (for example by teaching them crosschecking techniques). Javed Mostafa (University of North Carolina at Chapel Hill) discussed how health information seekers often involve and depend on other "co-consumers" of health information such as caregivers and physicians. To help support this, Mostafa proposed an approach to develop different profiles (e.g., one for the elderly end-user, one for the caregiver, and one for the physician) that can be used with machine learning approaches in collaborative information retrieval environments. Finally, Jyothi Vinjumur (University of Maryland, College Park) talked about her research on e-discovery, focusing on how legal professionals and technology (such as retrieval techniques) could collaborate to ensure proper production in a cost effective way, including factors. In her talk, Vinjumur discussed how factors such as context, cognition, and annotator expertise affect the process and quality of the review process.

3   Breakout Groups

Across the two days of the workshop, six breakout groups formed and discussed research challenges in the following areas: cross-language retrieval, e-discovery, health information, learning, search trails, and evaluation.

Group #1: Cross-Language Retrieval – This breakout group focused on issues related to collaboration in cross-language retrieval. The group discussed challenges in specific domains including e-discovery, cross-language access to cultural heritage collections, and military decision making. The group identified a number of research issues including: how best to use limited human translation resources, how to integrate human and machine translation, and how to communicate with a curator who does not know the searcher's language, and how to integrate collaborators with different levels of domain and search expertise.

Group #2: E-Discovery – This group focused on the legal process of e-discovery and the use of technology this process. The group set as their objective to reduce the cost of manual document review in the e-discovery process by using collaboration. They outlined a Semi-Automated Collaborative Technology Assisted Review (CTAR) system that would use algorithmic mediation to gather high quality relevance judgements. In their proposed system, human reviewers would annotate documents and the system would provide support (1) by training a document ranker to help order the documents to be manually reviewed, and (2) to visually highlight features of the document that indicate relevance.

Group #3: Health Information – This group discussed collaborative information seeking issues related to health. The group identified and discussed questions including: what relevance cues to patients use to evaluate collaborative health information, what roles to caregivers and health care providers play in collaborative information seeking, how does information seeking change over time for patients with chronic conditions. The group discussed how collaborative information behaviors are a type of "wicked problem", with requirements that are constantly evolving. To address this, the group proposed to develop exemplary cases in medical collaborative information behavior with a goal to generalize and characterize tasks. The group identified several areas that might lead to important cases: new diagnosis of a life-changing chronic illness, end-of-life and hospice care decisions, and diagnosis of illness with multiple care or treatment options requiring patient decision-making.

Group #4: Learning – This group identified one focal interest of the SCIS attendees – the relationship between learning and SCIS. SCIS can be thought of from two perspectives in a learning context: (1) searching to learn, (2) learning to search. The two might be seen combining, e.g. in information literacy contexts where students learn how to find and evaluate high quality content. In many cases, students need to learn how to use the (SCIS) tool, but we also might care about students finding the answer, understanding the bigger picture, learning to do SCIS better, or just engaging in information seeking faster/more efficiently (as an outcome of the SCIS). The group compiled numerous references on these topics, along with a list of resources relating to education/learning and search (these resources are detailed in the full workshop report).

Group #5: Search Trails – This group discussed the potential of studying search trails as a method of student learning of search skills and domain knowledge. A search trail is a record of an individual's actions and interactions during a search session, which can include the search terms entered, pages visited, and paths traversed. Two main questions emerged: what are the potential uses of search trails, and what are benefits to others? The group discussion identified possible benefits of exposing students to search trails and suggested possible study designs to incorporate individual searching and learning with social/collaborative activities. For instance, study participants could search individually, and then come together to view the group's results, comparing and reflecting on differences and similarities. Studies could also follow the classroom instruction model of "think/pair/share" by searching first as individuals, then comparing results with a partner, and reporting on their discussion to the larger group.

Group #6: Evaluation – This group started with a general discussion about different research problems involving social and collaborative information seeking, followed by brainstorming on how to evaluate knowledge change as a result of SCIS activities/tasks. The group proposed a number of measures/instruments that can be used for evaluation in SCIS, and provided a sample study design to demonstrate and discuss such usage. The research method proposed by this group involved a mixed-method approach in which observations, interviews, surveys, and log data analysis are used. The group also recommended using pre-test and post-test to evaluate learning during an SCIS activity. It was suggested that the following variables need to be controlled or at least kept in mind while conducting a study: space (co-located, remotely located), time (synchronous, asynchronous), common ground, prior knowledge, education level (institutional), group size, resources/tools used or available, and task difficulty/complexity.

4   Research Roadmap

The last two sessions of the workshop on the second day focused on discussing future directions for SCIS. Given what we know, what we have learned and discussed at this workshop, where do we go from here? What are some of the next steps for us and others interested in SCIS research to follow? Here are some of the points, in no particular order, that came up during the final session discussions.

- What are the fundamentals of SCIS research? Why should people care about SCIS? What is the big impact on the society? One answer – SCIS allows us to address problems that are too difficult or even impossible for individuals to do.
- What other important problems might be addressed with the use of socio-collaborative connections? One approach is to look at existing problems and consider what happens when you put 'collaborative' put in front of it. Example: Collaborative analytics.
- Algorithmic approaches are important and SCIS would benefit from having people with algorithms backgrounds involved (e.g., traditional CS and engineering system-oriented). Bringing system-focused and user-focused researchers and developers together to design SCIS systems and services could have benefits such as: (1) showing users what we are doing, (2) providing algorithmic mediation with classifiers, features, (3) encouraging people to collaborate.
- In addition to research and development, we should also focus on education. We could work on developing curriculum/courses for SCIS. The content developed through collaboration of scholars in SCIS area could be available online for educators to easily incorporate into their existing curriculum. For instance, we could suggest list of foundational readings and a set of activities to explain/reinforce SCIS behaviors. A teacher or course developer could use them as needed in their courses. Potential courses/subjects that could incorporate SCIS topics include: information retrieval (IR), human-computer interaction (HCI), and human information behavior (HIB).
- Those of us who teach classes in SCIS, IR, HCI, and HIB could give collaborative search activities to students and reflect on them. We could also create a pool of interesting topics for students to select for both smaller and larger assignments such as bachelor or master thesis.
- Another idea was to organize a summer program or seminar to educate/train students and new scholars in the area.
- We also discussed possible funding opportunities for work on SCIS and ways to communicate to funding agencies the great potential for impactful research in this area. Possible sources include NSF, IARPA (incisive analysis), DARPA, NIST, NLM, NIH, NSF-SCH. Outside of the US, there are other funding agencies such as the European Commission, Canadian funding agencies, etc. It will be important to address what the computational aspects are that will be advanced by SCIS research. What are computational models that can be developed, used, and evaluated?
- Another possible and highly valuable activity is to create test collections, tasks, evaluations that we can all share. This could be explored as part of a NIST TREC track, and/or in coordination with other test collection activities such as the CLEF in Europe, NTCIR in Japan, and FIRE in India.

5   Conclusion

These are the take-away points and themes from the workshop.

- The areas of social and collaborative information seeking have broad interest, with many application domains where impacts could be made.
- Mediated collaboration was a topic of increased interest – how can systems help mediate information seeking processes that involve humans with different skill sets, languages, roles, etc.?
- Collaboration around health information seeking was an area of strong interest. This area has a number of dimensions that make it interesting – often there are multiple people involved with different roles and skills, are all working together.
- Learning in collaborative search is another area with strong interest. How can we encourage, foster, and measure learning through collaborative information seeking? These are big challenges, but they also have great potential.
- A big selling point of introducing or considering SCIS is to show where information-intensive tasks that are normally done by individuals could benefit by applying social and/or collaborative considerations. For instance, the intelligence analysis area could benefit by subject experts, search experts, and decision makers being connected in a way that does more than just chaining them in the process of discovering patterns of interest from a massive amounts of information coming in. A student may benefit by having a study buddy or peer mentor mediated by a tutor in a learning environment.
- There have been several interesting and impactful works done in SCIS, stemming from diverse disciplines such as health and education, but now is a time to bring scholars from these disciplines to work on the next generation of problems together. This will require us to create ways in which we can start sharing resources (tools, systems, study design templates), data sets, methods, and findings easily and effectively. It will also require educating not just our students and colleagues, but also funding agencies. Some of these efforts may need assistance from agencies to begin and support their work until they receive enough momentum to be sustainable.

To help support SCIS research, after the workshop an email list (scis@infoseeking.org) and a community website for sharing resources (http://scis.infoseeking.org) were created.

6   Acknowledgements

We would like to thank DIMACS for their funding and support, making this workshop possible. This workshop would also not have been successful without the diverse and distinguished set of participants we had.