

# Report on the Workshop on Social and Collaborative Information Seeking (SCIS)

May 14-15, 2015  
DIMACS Center, Rutgers University

## Organizers

**Chirag Shah<sup>1</sup>, Rob Capra<sup>2</sup>, Preben Hansen<sup>3</sup>**

1. Rutgers University, [chirags@rutgers.edu](mailto:chirags@rutgers.edu)
2. UNC Chapel Hill, [rcapra@unc.edu](mailto:rcapra@unc.edu)
3. Stockholm University, [preben@dsv.su.se](mailto:preben@dsv.su.se)

This Workshop was presented under the auspices of the DIMACS Special Focus on Information Sharing and Dynamic Data Analysis with support from DIMACS and its partners, and from the National Science Foundation under grant #1144502.

---

**DIMACS**

---



## **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.

A workshop, sponsored by the Center for Discrete Mathematics & Theoretical Computer Science (DIMACS) at Rutgers University, 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 is the report from the workshop. While great care has been taken to synthesize and present the ideas and outcomes from the workshop, there are certainly many "raw" elements here that one might have to take with a pinch of salt. Having said that, we do hope this report provides several pointers and a fair bit of inspiration for great many challenges and opportunities that lie ahead in the area of SCIS.

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. These participants are also instrumental in producing this report. This was truly a socio-collaborative endeavor!

*Chirag Shah, Rob Capra, Preben Hansen*

## **Executive Summary**

The workshop was held May 14-15, 2015 at the Rutgers University at the Center for Discrete Mathematics & Theoretical Computer Science (DIMACS), and was sponsored by Rutgers DIMACS as part of their focus on Information Sharing and Dynamic Data Analysis, funded by the National Science Foundation.

Participants were solicited through invitations and via a call for participation that was widely circulated to information retrieval and information science listservs. In total, 28 participants attended the workshop (see list below). Participants had a range of backgrounds in information and computer science fields and the invited participants intentionally included a balance of senior to junior researchers, including PhD students. Prior to the workshop, invited participants submitted short abstracts describing their interests and backgrounds related to social and collaborative information seeking. These abstracts are posted on the workshop website (<http://dimacs.rutgers.edu/Workshops/SCIS/abstracts.html>) and are included in Appendix A of this workshop report.

The workshop was designed as a “working” workshop with a goal of outlining a “research roadmap” of important future research needed in the area of social and collaborative search. The program was structured into four main phases:

- Thursday morning -- Introductions and Background. During the first morning, the sessions focused on introducing the goals of the workshop and introducing the participants to each other’s work, backgrounds, and interests. Time was allotted for the invited participants to give brief 7-minute presentations about their recent work and interests (see the “Short Talks” section of this report), and other participants introduced themselves informally.
- Thursday early afternoon -- Identify research areas. On Thursday afternoon, we formed breakout groups, each charged with identifying a list of open research questions and topics important to SCIS. The responses generated by each group were then presented and discussed in a plenary session. A summary of the research questions, organized by theme, is presented in the section, “Research Questions”.
- Thursday late afternoon -- Address specific research questions in detail. After a break on Thursday afternoon, we re-formed new breakout groups and asked each group to select one or two of the research questions discussed during the previous plenary session. For the selected question(s), the groups were asked to come up with at least two different ways they might study the question, and to discuss the pros and cons of different approaches. Again, after these breakout

### SCIS 2015 Workshop Report

groups, we held a plenary session in which each group presented their ideas, and the whole group discussed the approaches.

- On Friday morning, most of the breakout groups from the end of Thursday reformed and continued to work in more depth on the research designs they started on the previous day. Two groups adjusted their focus on Friday. Thus, over the course of last two breakout sessions, six groups discussed detailed approaches to study specific research questions. During a final plenary session, on Friday, the groups presented the results and discussed them with the whole group. After the workshop, each group wrote a summary of their discussion and the approaches they developed. These summaries are presented later in this report as Appendix B.

### Workshop Participants

Mark Ackerman	University of Michigan
Talal Ahmed	Rutgers University
Yigal Bejerano	Alcatel-Lucent, Bell-Labs
Nick Belkin	Rutgers University
Rob Capra	University of North Carolina at Chapel Hill
Dongho Choi	Rutgers University
Kaitlin Costello	University of North Carolina at Chapel Hill
Roberto González-Ibáñez	Universidad de Santiago de Chile
Daqing He	University of Pittsburgh
Simon Knight	The Open University, UK
Chris Leeder	Rutgers University
Mohammad Ali Maddah-Ali	Alcatel-Lucent, Bell-Labs
Ziad Matni	Rutgers University
Matthew Mitsui	Rutgers University
Javed Mostafa	University of North Carolina at Chapel Hill
Douglas Oard	University of Maryland
Connie Pascal	Rutgers University
Jeremy Pickens	Catalyst Repository Systems
Soo Young Rieh	University of Michigan
Babak Saleh	Rutgers University
Chirag Shah	Rutgers University
Aiko Takazawa	University of Illinois at Urbana-Champaign
Sandra Toze	Dalhousie University, Canada
Michael Twidale	University of Illinois
Jyothi Vinjumur	University of Maryland
Ellen Voorhees	NIST

Hassan Zamir  
Yinglong Zhang

University of South Carolina  
University of Texas at Austin

## **Workshop Program**

### **Thursday, May 14**

8:00-9:00	Registration and breakfast
9:00-9:20	Introduction to the workshop by the organizers
9:20-10:30	Short talks by participants (8 x 7 minutes)
10:30-11:00	Coffee break
11:00-12:30	Short talks by participants (12 x 7 minutes)
12:30-1:30	Lunch
1:30-1:40	Welcome by Rebecca Wright, Director of DIMACS
1:40-1:55	Discussions to plan the breakout sessions
1:55-3:00	Breakout session-1 (includes time to report back to full group)
3:00-3:30	Coffee break
3:30-4:45	Breakout session-2
4:45-5:30	Breakout progress reports to whole group, full group discussion
5:30-6:30	Reception
6:30-7:00	Travel to restaurant
7:00-9:00	Dinner at Panico's restaurant (downtown New Brunswick)

### **Friday, May 15**

8:30-9:00	Breakfast
9:00-10:30	Breakout session-3
10:30-11:00	Coffee break
11:00-12:00	Outlining a research roadmap
12:00-1:00	Lunch

## Short Talks by Invited Participants

### Thursday, May 14

#### 9:20-10:30 *Short talks by participants (7 minutes each)*

1. Social and Collaborative Information Seeking (SCIS): Space, Time, and Beyond: Chirag Shah, Rutgers University
2. Searching for Help: How Learning Technologies Involve Collaborative Search: Michael Twidale, The University of Illinois at Urbana-Champaign
3. Affective Dimension in Collaborative Information Seeking: Roberto González-Ibáñez, Universidad de Santiago de Chile
4. Collaborative Cross-Language Search: Doug Oard, University of Maryland, College Park
5. Evaluation Measures in Social Search: Soo Young Rieh, University of Michigan
6. Library Research as Collaborative Information Seeking: Chris Leeder, Rutgers University
7. Culture and Trust in Collaborative Information Seeking: Yinglong Zhang, University of Texas, Austin
8. Searching to Help: Collaborative Information Seeking in a Disaster Relief Context: Aiko Takazawa, The University of Illinois at Urbana-Champaign

#### 11:00-12:30 *Short talks by participants (7 minutes each)*

1. System Support for Collaborative Information Seeking: Rob Capra, University of North Carolina at Chapel Hill
2. Context-Sensitive Supports for Collaborative Information Retrieval: Daqing He, University of Pittsburgh
3. Exploring the Group within Social and Collaborative Search: Sandra Toze, Dalhousie University
4. Reducing E-Discovery Cost with Collaborative Review Process: Jyothi Vinjumur, University of Maryland, College Park
5. Collaborative Information Seeking Tasks as Complex Performance Assessments: Simon Knight, Open University, UK
6. Collaborative Information Access in Health: Mark Ackerman, University of Michigan
7. Collaborative Cross-checking: Patients Teaching Patients How to Evaluate Health Information in Online Support Groups for Chronic Kidney Disease: Kaitlin Costello, University of North Carolina at Chapel Hill
8. Task-Constrained Collaborative Information Seeking: Jeremy Pickens, Catalysts Inc.
9. Social Searching and Information Recommendation Systems: Hassan Zamir, University of South Carolina
10. Collaborative Search Challenges for Adaptive and Personalized Search for the Elderly: Javed Mostafa, University of North Carolina at Chapel Hill

## **Summary of the Talks**

The short talks that followed the introduction of the workshop, was divided into two sessions during the first part of day 1. All in all, there were 18 short presentations, each 7 minutes long, followed by questions from the workshop participants.

First presenter out to talk was Chirag Shah from Rutgers University. He talked about the two important dimensions in social and collaborative information seeking (SCIS): space and time. Specifically, these two dimensions were discussed based on different lab and field studies conducted at Rutgers. Finally, Chirag Shah also pointed to the roles of other important dimensions like communication, awareness, affects, and group sizes.

Next, Professor Michael Twidale, The University of Illinois at Urbana-Champaign, talked about how learning technologies could involve collaborative search, within situations where people learn how to use for example informational resources in new ways such as adapt, appropriate, combine and modify them. Twidale pointed out that technology learning is often both a search and a social activity and different strategies could improve efficiency, thus, this may have implications for design, policy and education.

Then, Roberto González-Ibáñez, Universidad de Santiago de Chile, focused in his talk on the importance of affective processes such as emotions in collaborative information seeking (CIS). This was first done by presenting an overview of the relevance of the affective dimension, followed by some findings from research on the affective dimension. Finally, Roberto González-Ibáñez, invited the research community to elaborate on new research questions and hypotheses.

Douglas Oard, University of Maryland, College Park, focused on a rather unknown research area within CIS: collaborative cross-language information retrieval (CLIR). Douglas Oard discussion about a research agenda that include collaborative CLIR and adds language expertise. This was exemplified through a presentation of a project in which collaborative translation task with people having different skills (one knowing the source language, the other the target language). Drawing on that experience, Douglas Oard summarized in what way the CLIR setting is different, and how those differences might help to inform the research agenda for collaborative information retrieval.

Next, Soo Young Rieh, University of Michigan discussed evaluation measures in social search. Soo Young Rieh proposed to develop set of evaluation measures that could be used in social search. The reason for this is that the context of collaboration could be more diverse than what's defined in CIS frameworks. These measures were derived from two empirical studies by Ji Yeon Yang (2013) and Grace Jeon (2014). Basically,

*SCIS 2015 Workshop Report*

four categories of measures were discussed. (a) IIR performance measures, would involve metrics for information diversity, besides the traditional measures such as precision and recall. In the category of (b) informational outcomes, measuring information quality based on comprehensiveness, novelty, trustworthiness of information, could be done. Next (c) social outcomes involve the users' appreciations of other people's attempt, effort, responsiveness, and understanding of information needs. In the final category (d) of user search experience, Soo Young Rieh suggest that subjective assessment of overall search experience could be measured.

Then, Christopher Leeder, Rutgers University, presented results of a study with students conducting collaborative research using library resources. The Coagmento (<http://coagmento.org/>) collaborative search system was used while participants were working on a class assignment. Christopher Leeder shared the results from the study that involved both drawbacks as well as benefits to collaborative information seeking. Findings show that groups found more useful sources and greater information coverage while working together. In contrast, individuals did better regarding query effectiveness and working with amount of relevant sources.

Then, Yinglong Zhang, University of Texas, Austin, discussed the important topic of trust as a common motivation for individuals to involve in collaborative search tasks. It is argued that the success of collaborative work is largely based on whether members can trust each other in a collaborative group. Furthermore, culture is an important factor that to a large extent influences the development of trust in collaboration. One such cultural aspect is that people from different cultures can interpret the same problem based on their own cultural knowledge and beliefs. The gaps in problem representation can increase misunderstanding and conflicts in collaboration, therefore weakening trust among the individuals.

The last presenter before the break, Aiko Takazawa, The University of Illinois at Urbana- Champaign, discussed a case study involving seven Japanese women living in Finland creating an ad-hoc self-organized humanitarian aid group in response to the 2011 Great Tohoku Earthquake and Tsunami disaster in Japan. This study unfolded the work of a group of people and of how they collaboratively searched and used information with available technologies. Since this study was done in a real setting, the presenter also highlighted how the 'messiness' of a natural setting with intertwined and different kinds of both individual and collaborative information activities could provide deep insights in CIS. One of the goals was to discuss how to make sense of social data drawn from a microcosm case study.

*SCIS 2015 Workshop Report*

After the break, another 8 presentations were performed. First out was Rob Capra, University of North Carolina at Chapel Hill, discussing that people do not only search information together in collaborative activities, but also coordinate their activities, involving planning, communicating results, monitoring progress, creating shared representations of structure, and performing synthesis of findings. Rob Capra presented a project called ResultSpace, a collaborative search environment, and focused on strategies and triggers to shift strategies and the influence of awareness. He also described his recent research to investigate the benefits of making “search trails” of previous users search paths available to future searchers who are doing similar tasks.

The next topic presented by Daqing He, University of Pittsburgh, was about context-sensitive supports for collaborative information retrieval (CIR). The presentation described studies on exploring the effectiveness of various search contexts in supporting collaborative information retrieval. Especially, the effectiveness of the search context based not only from the user’s own search history, but also taking into account all the different partner’s search histories in a team. Further to that, the team’s explicit collaboration behaviors were also studied. This was done studying the participants’ chats. One major finding was that context-sensitive CIR has its own uniqueness.

Next, Sandra Toze, Dalhousie University, explored the group within social and collaborative search activities with the important question: How can we study groups? It was argued that the “group” represents a unique level of specific attributes such as interaction, interdependence, awareness and shared understanding, and that these need to be better understood and supported. Thus, the focus has been to understand and model information needs, seeking and use at the group level. A longitudinal study was presented and the findings point to that there is a need to develop series of tools and series of information tasks and to facilitate more effective group work. She presents a conceptual model of Group Information Process (GIP) that provides a base that can be used for further research.

Jeremy Pickens from Catalysts Inc., talked about task-constrained collaborative information seeking. In contrast to implicit collaboration, explicit collaboration involves a wider range of possibilities. For example, when two or more searchers collaborate on a task, the individual roles do not need to be symmetric. Jeremy Pickens discuss the importance of role asymmetries in certain professional domains and the needs themselves are not always jointly negotiable between different collaborative partners when solving a task. For example, one collaborator may have a defined need, and the other collaborators may need to support that need in the overall task. One question discussed related to this problem is how collaborative systems can support this kind of team activity.

Next, Simon Knight, Open University, UK, focused on collaborative information seeking tasks as complex performance assessments. This was done through the development of two tasks: collaborative information seeking, and collaborative multiple document processing. The performance assessment was developed for higher-level literacy using the Coagmento (<http://coagmento.org/>) browser add-on and analysis focuses on the relationship between CIS processes and learning outcomes.

The presentation by Mark Ackerman, University of Michigan, had its focus on collaborative information access in the health domain. More specifically, the research presented was from people with chronic medical conditions, like diabetes or depression, which involve long-term information needs. Medical situations, in general, are complex and information seeking for these conditions occurs as a combination of both individual and social information gathering, using many information sources. The questions discussed during the presentation dealt with how to help people sense-make their conditions and the information they are gathering from their various social worlds. One example was a project about information in bone marrow transplantation. Findings shows that there was a shift to teams involving not only clinicians, but also relatives and different caregivers.

Kaitlin Costello, University of North Carolina at Chapel Hill, also presented research from the medical domain. The topic was social information behaviors and more specifically on patients teaching patients how to evaluate health information. The presenter called this process for collaborative cross-checking and may occur when one person consults multiple sources of information in order to verify certain information. Therefore collaborative cross-checking is collaborative information literacy practices in which users attempt to teach other users how to verify information. Finally, Kaitlin Costello, pointed out that collaborative cross-checking may serve multiple functions, such as, refute the misinformation from the original post; offer evidence supporting the correct information, and fosters an understanding of how we may evaluate information by offering clear instructions.

A third presentation within the medical domain, by Javed Mostafa, University of North Carolina at Chapel Hill, highlighted that within the health domain, elderly users often involve and depend on other “co-consumers” of health information. This may result in that the caregivers or physicians sometimes find themselves in the role of search result “interpreters” and suggesting improved search strategies. This specific issue is investigated through machine learning approaches in which combined cumulative knowledge from user profiles in collaborative information retrieval environments. One of the major techniques used was multi-agent learning to support personalization. The

*SCIS 2015 Workshop Report*

idea is to use three different profiles: one for the elderly end-user, one for the caregiver, and one for the physician. Machine learning techniques are then used to update and combine these profiles. Thus, one of the goals is to develop theoretical frameworks and models that can support our understanding of collaborative consumer-health information seeking.

The last speaker, Jyothi Vinjumur, University of Maryland, College Park, presented research that focused on e-discovery and how the legal professionals and the technology could collaborate to ensure proper production in a cost effective way. Even though technology have been developed to handle information better, the amount of information have increased much more which make the reviewing process cumbersome. Jyothi Vinjumur mentions that context, cognition, annotator expertise, all affect the process as well as the quality of the review process. The conclusion of this complex reviewing process is that the process could be done more effective in both time and performance as a collaborative task, rather than as a solitary activity.

**Research Questions (Breakout Session 1)**

Small groups brainstormed research questions in CIS, which were then presented back to the main group at the end of the session. The questions/points from this session fell in the following categories.

**THEORY:**

- Dynamics of collaboration: roles of the collaborators
- Dynamics of collaboration: within-group processes, roles
- Dynamics of collaboration: knowledge flows
- Dynamics of collaboration: multiple knowledge sources
- Dynamics of collaboration: so what is being transferred in the group?
- Dynamics of collaboration: products, assessing stopping
- Dimensions of CIS: purpose - why are we doing this, motivation
- Dimensions of CIS: power differentials, roles
- Dimensions of CIS: temporality, same time, asynch, within 1 hour, 1 month, 1 decade
- Dimensions of CIS: number of participants (scale, size): 2, 3, 4-8, 9-24, 25-100, 100-1000, 1000-10,000,000
- Dimensions of CIS: explicitness, ad-hoc-ness
- Dimensions of CIS: affect, trust, cognition
- Dimensions of CIS: technologies used
- Dimensions of CIS: location: where are the participants
- CIS pathologies: too much talking
- CIS pathologies: tech overhead
- CIS pathologies: just the intersection of knowledge
- CIS pathologies: not being very collaborative
- CIS pathologies: no complementarity - just a set of individuals

**EPISTEMOLOGICAL:**

- What/how to evaluate: when is CIS worthwhile?
- What/how to evaluate: is there a group level cognition/learning?
- What role does distance (geographical, cultural, affective, experiential, etc.) play in collaborative information behavior?
- What are the knowledge boundaries, individuals' experience?

*SCIS 2015 Workshop Report*

- What skills/interest/qualities help people in doing collaborative search?
- What cues do people use in assessing credibility or relevance (of the system? of other people?) in CIS?
- How do levels of collaboration (social context, culture, cross expertise, and potential reuse) play in CIS?
- What is difference between CIR in general knowledge and specific knowledge, health, politics, education, legal, scholarly
- Where the collaborative is in the search/seeking process?
- Where does collaborative search happen that we're not aware of? (What can we learn from how they are doing collaborative search (1.0))
- Where is collaborative search not happening but it could? And why not? What barriers?
- How do you tease out the influence in collaborative encounters? How much influence do individual people, system components, have on encounters? Can we tease out the effect of that influence?
- Does being able to tease out influence even matter in collaborative information seeking?
- How do you quickly get a grasp of who brings what to the team?
- When do you get synergistic effects? How do you measure them?
- As researchers, what practical issues in CIB can we tackle in the next 1-3 years?
- Motivation/benefit: What are the benefits of collaboration? What motivates it? - Impact/consequences of the CIS, Application (areas/goals like health + education; improve recall and precision), Gathering/seeking vs. ranking/evaluation of documents, Temporal aspect: Value of collaboration at different stages, Satisficing vs. deeper understanding (Peer pressure vs. group dynamics vs. task motivation?)

**FACTORS TO CONSIDER:**

- Information flow among different components, how to facilitate the flow of information
- Use of structure (hierarchy, flat structure, mediation and other roles) in CIS
- Temporal difference, sometime very intensive among different partners, sometimes is relatively slow in pace; temporal aspects in a CIS episode
- Collective vs. individual vs. subgroup outcomes

### *SCIS 2015 Workshop Report*

- When obscuring info is handy, being usefully vague
- Fluid shift CIS IIS, within groups, doing. how the transitions occurs
- Different expertise, how this can be integrated into the collaborative, not just different roles. how to know the expertise. how to evaluate the expertise. could be domain expertise and search expertise. can we trust the person's expertise
- Organizational context: in a hospital, with family, in a class
- How to facilitate sharing esp in strict role boundaries
- Learning and retrieving from people as well as the info resources / systems
- How one search event determines what you do next
- Domain, topic area: Medical, legal, engineering, shopping
- Knowledge / experiential context of participants
- Scale and granularity - size of groups, level of collaboration (and information)
- What counts as the 'value' of information in a CIS context
- How to get data set. What could the right dataset, evaluation framework, what could be the right dataset
- Failures of different subparts of the CIS process

### EVALUATION:

- Metrics: which ones to use?
- Does recall / precision work well in CIS?
- Metrics: usefulness
- Metrics: what is collaborative relevance?
- What/how to evaluate: whether the information transferred
- What/how to evaluate: is the knowledge flow hierarchical (from knows most down?) does the expert learn too?
- What kind of system supports can be developed, evaluated
- How to evaluate individuals in a team
- Learning from collaborative search, how, what measure?

### APPLICATIONS:

- What are the contexts or scenarios where we can create tools for support of CIS?
- With systems what are contexts where we CANNOT support CIS? (Are explicit tools always possible or even desirable?)

*SCIS 2015 Workshop Report*

- How can lightweight, flexible tools for collaborative info behavior behaviors?  
Does everything have to be a heavyweight “system”?
- How do mobile access and multiple types of devices play in CIS?
- What affordances should be built into CIB systems?
- What kind of basic system components could be? it is related to what we see collaborative seeking is. such as social sharing, visualization, sense making, recommendation
- Can we design for spontaneous collaboration systems for serendipitous encounters?
- What elements of collaboration are ephemeral and can't be captured by systems? Ad-hoc-ery?

**Research Studies (Breakout Sessions 2 and 3)**

After a break on Thursday afternoon, we re-formed new breakout groups and asked each group to select one or two of the research questions discussed during the previous plenary session. For the selected question(s), the groups were asked to come up with at least two different ways they might study the question, and to discuss the pros and cons of different approaches. After the breakout sessions, we held a plenary session in which each group presented their ideas, and the whole group discussed the approaches.

On Friday morning, the breakout groups from the end of Thursday reformed and continued to work in more depth on the research designs they started on the previous day. Two groups adjusted their focus on Friday. Thus, over the course of last two breakout sessions, six groups discussed detailed approaches to study specific research questions. During a final plenary session, on Friday, the groups presented the results and discussed them with the whole group. After the workshop, each group wrote a summary of their discussion and the approaches they developed. These summaries are presented later in this report in Appendix B. The groups were the following.

- Group #1 – Contexts for SCIS, Cross-Language Retrieval
- Group #2 – Contexts for SCIS, E-Discovery
- Group #3 – Health Information in SCIS
- Group #4 – Learning and SCIS
- Group #5 – Search Trails
- Group #6 – How to evaluate SCIS systems

## Research Roadmap (Final Session)

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 that came up during the final session discussions.

- Developing curriculum/courses for SCIS: we could collaboratively design content and courses for online teaching.
- Give collaborative search activities to students and reflect on them - this could even be in regular classes that cover any aspect of information seeking/retrieval or HCI.
- Organize a summer school - need funding
- Develop a special issue - but it's been done before and recently (2010 IP&M special issue on CIR, 2014 IEEE Computer special issue on CIS)
- Let's think big - talk to people with funding opportunities - possible places NSF, IARPA (incisive analysis), DARPA, NIST, NLM, NIH, NSF-SCH
- What about European Commission? Canadian funding agencies? Grand? Nectar?
- To NSF - what part of computation aspect this will advance? What are computational models?
- Creating test collections, tasks, evaluations that we can all share - perhaps using NIST?
- Another way to think about is is looking at existing problems and asking what happens when you put 'collaborative' put in front of it? Example: Collaborative analytics.
- Think about the fundamentals of SCIS research. Why would anyone care? 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.
- How do we get more algorithms (traditional CS and engineering system-oriented) people involved?
- Things we could do as researchers and developers providing SCIS systems and services: (1) show users what we're doing, (2) provide algorithmic mediation with classifiers, features, (3) encourage people to collaborate
- Centralized resources? - *Action*: create a repository for everyone to deposit and share resources (Chirag Shah)
- Talking to potential funders, collaborators, and students? Ask yourself - what's getting in the way of doing the next step?

## Conclusion

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

- The areas of social and collaborative information seeking have received a much more diverse treatment and interest in the recent years than previously thought.
- 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 another 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. In other settings, information may be gathered from diverse sources in loose collaboration.
- Learning in collaborative search seems to be of great interest to many. 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, 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 could 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.

Note that after the workshop, an email list ([scis@infoseeking.org](mailto:scis@infoseeking.org)) and a community website for sharing resources (<http://scis.infoseeking.org>) were created. As of the release of this report in late June 2015, the website is still being constructed.

## Appendix A: Invited Talks Abstracts

**Mark Ackerman**, University of Michigan, Ann Arbor

*Title: Collaborative Information Access in Health*

People with chronic medical conditions, such as diabetes or depression, have long-term information needs. These conditions persist. Diabetes cannot be cured, only controlled over one's lifetime, and depression can wax and wane over one's lifetime. Many other chronic conditions, with similar information needs, exist. Information seeking for these conditions occurs as a combination of individual and social information gathering, often in a rich ecology of information sources. Information gathering may be proactive, but it can also be quite passive. Often, information is gathered casually for later use, but it can also be garnered when a crisis occurs or when the condition changes for the individual. Often the need is very contextualized, as must be the information, since it is peculiar one's lived experience - the specifics of one's body and one's socio-economic conditions.

Medical situations, in general, are exceeding complex and often involve large numbers of clinicians and auxiliary personnel, as well as the family, extended family, and friends. As one gets older, the number of conditions and co-occurring conditions increases, increasing the complexity for the user. The average Medicare patient has 23 clinicians he/she is trying to juggle.

We are currently exploring a number of projects to help people sense-make their conditions and the information they are gathering from their various social worlds. Our research group has explored diabetes and hypertension in an underserved community, diabetes support, depression monitoring, information scaffolding in adult bone marrow transplant (BMT), information artifacts in pediatric BMT, and the use of socially-derived "translations" in order to understand medical advice more clearly. We have also examined in-hospital information exchanges among nurses and doctors, and we will shortly be starting studies of spinal cord injury patients and community-based depression management. We are also constructing a video-based prototype so as to help people navigate the differing and sometimes conflicting advice they get from their various communities and social worlds as well as a tablet-based system to help in-patient BMT caregivers understand their child's condition.

---

**Rob Capra**, University of North Carolina, Chapel Hill

*Title: System Support for Collaborative Information Seeking*

In collaborative information seeking, collaborators not only conduct searches for information, but also must coordinate their activities, including planning, communicating

results, monitoring progress, creating shared representations of structure, and performing synthesis of findings (Morris & Teevan, 2010; Poltrock et al., 2003). Prior research on collaborative search has investigated systems that help groups communicate and coordinate activities during collaborative search. In this short talk, I will discuss several ways that systems could provide more "structural" support for collaborative search. Explicit system support for shared, modifiable representations of task and results structure could help teams in planning, searching, and sensemaking during a collaborative search process. This structure could also be leveraged by search algorithms to present results that match particular sub-goals. My goal is to present ideas that will generate discussion at the workshop.

---

**Kaitlin Light Costello**, University of North Carolina, Chapel Hill

*Title: Collaborative crosschecking: Patients teaching patients how to evaluate health information in online support groups for chronic kidney disease*

Patients living with lifelong health conditions often search for information about their health throughout their illness trajectory (Johnson & Case, 2012). Many patients diagnosed with chronic conditions are increasingly turning to the Internet for health information (Fox & Duggan, 2013), where they may encounter online support groups [OSGs] dedicated to their chronic condition. In OSGs, patients routinely exchange information and social support with one another as they come to terms with their diagnosis, make treatment decisions, and learn what to expect as their illness progresses. Healthcare providers are often concerned that patients will find misinformation both in OSGs and on static websites, and they may even deter patients from using the Internet as an information source because of concerns about credibility (Chung, 2013). In OSGs, misinformation is often corrected relatively quickly by other users (Esquivel, Meric-Bernstam, & Bernstam, 2006).

However, my research suggests that users do not merely correct misinformation when they encounter it in an OSG. Participants in a recently completed two-year grounded theory study examining the information behaviors of patients diagnosed with chronic kidney disease [CKD] in OSGs attempt to teach other users how to evaluate the credibility of health information posted in OSGs when they encounter misinformation. This is a process that I call collaborative crosschecking. Crosschecking is common among my participants, and occurs when one person consults multiple sources of information in order to verify information. Collaborative crosschecking, therefore, is a collaborative information literacy practice whereby users attempt to teach other users how to verify information by sharing their own crosschecking techniques. That is, my participants do not simply correct misinformation when they encounter it in OSGs: they walk through their own evaluation process in an attempt to teach others to evaluate the trustworthiness of information. This often results in additional users adding their own

*SCIS 2015 Workshop Report*

sources and giving their own tips for crosschecking. Interestingly, my participants tell me that they are careful when they engage in collaborative crosschecking: they use gentle language, remind users that "everyone is different" and that not all information about CKD applies to everyone, and often provide multiple references to back up their claims.

Collaborative crosschecking serves multiple functions: it refutes the misinformation from the original post, offers evidence supporting the correct information, and fosters an understanding of how to evaluate information by offering clear instructions. This collaborative information literacy practice likely extends beyond the health domain. For example, it is similar but not identical to the "call and avalanche" pattern of receiving answers to questions in online forums for massively multiplayer online games (Martin & Steinkuehler, 2010). Additional research is necessary to determine whether collaborative crosschecking occurs in other domains. Further research must also explore the effectiveness of collaborative crosschecking in disseminating information literacy skills and in correcting misinformation.

*References*

Chung, J. E. (2013). Patient-provider discussion of online health information: Results from the 2007 Health Information National Trends Survey (HINTS). *Journal of Health Communication*, 18(6), 627-648.

Esquivel, A., Meric-Bernstam, F., & Bernstam, E. V. (2006). Accuracy and self-correction of information received from an internet breast cancer list: content analysis. *BMJ*, 332(7547), 939-942.

Fox, S., & Duggan, M. (2013). *Health online 2013* (p. 55). Washington, D.C.: Pew Research Center's Internet & American Life Project. Retrieved from [http://www.pewinternet.org/files/old-media//Files/Reports/PIP\\_HealthOnline.pdf](http://www.pewinternet.org/files/old-media//Files/Reports/PIP_HealthOnline.pdf) (Archived by WebCite at <http://www.webcitation.org/6Wsi5u8S9>).

Johnson, J. D., & Case, D. O. (2012). *Health information seeking*. Peter Lang Publishing.

Martin, C., & Steinkuehler, C. (2010). Collective information literacy in massively multiplayer online games. *E-Learning and Digital Media*, 7(4), 355-365.

---

**Jeremy Pickens**, Catalysts Inc.

*Title: Task-Constrained Collaborative Information Seeking*

In information seeking, when collaboration is implicitly moderated the range of design patterns around collaborative activities is limited. However, when collaboration is explicit, a wider range of possibilities opens up. When two or more searchers collaborate on a task, roles do not necessarily have to be symmetric. Well-known role asymmetries include different levels or types of expertise and different levels or types of search activity. However, it is often assumed that the information need itself is symmetrically shared, that all collaborators have equal stake in the task, even if their roles in supporting that need differ. For some domains, such as families deciding to make large purchases or friends traveling together, this need symmetry assumption is not unreasonable. However, in certain professional domains such as law, the needs themselves are not always jointly negotiable between collaborative partners. Instead, one collaborator will define the need and the remaining collaborators are charged with the task of supporting that need. Information needs may evolve as learning takes place over the course of a collaborative session, but due to the professional nature of that task, ultimately approval of that evolving need may only be approved by certain collaborators. In such domains, it is an open question as to how collaborative systems can best support team activity, a question that we hope to discuss at the workshop.

---

**Roberto González-Ibáñez**, Universidad de Santiago de Chile

*Title: Affective Dimension in Collaborative Information Seeking*

Emotions and other affective processes have long been considered essential elements in people's lives. Despite emotion research conducted in different domains, little is known about the role of the affective dimension in the information search process of teams. Researchers have shown an active role of affective processes such as feelings and emotions in individual information seeking, however such findings do not necessarily apply to collaborative settings. This talk is aimed to reflect about the importance and challenges of research on the affective dimension in collaborative information seeking (CIS). To achieve this goal the talk is structured in three parts. First of all, an overview of the relevance of the affective dimension is provided. Second, some key findings from research on the affective dimension in individual information seeking in general, and CIS in particular are presented. Finally, research approaches, challenges, and ethical aspects in this type of studies are discussed. This talks hopes to encourage CIS researchers to explore the affective dimension in their studies, formulate new research questions and hypotheses, and share their findings with the community.

---

**Daqing He**, University of Pittsburgh

*Title: Context-Sensitive Supports for Collaborative Information Retrieval*

This brief talk presents our recent efforts on exploring the effectiveness of various search contexts in supporting collaborative information retrieval (CIR). We are particularly interesting in comprehending the unique search contexts available in CIR. Therefore, we examined the effectiveness of the search context drawing from not only the user's own search history, but also the partner's search history, as well as the team's explicit collaboration behaviors their chats. Our results demonstrate that context-sensitive CIR has its own uniqueness in considering the user's own and the partner's search history, but more importantly it can draw various contributions from the chats between team members.

---

**Simon Knight**, Open University, UK

*Title: Collaborative Information Seeking Tasks as Complex Performance Assessments*

My work has focused on the development of 2 (CIS) tasks Collaborative Information Seeking, and Collaborative Multiple Document Processing for the purposes of developing a performance assessment for higher-level literacy. The research uses the Coagmento (<http://coagmento.org/>) browser addon in a novel, and large scale, context. Students were asked to work in pairs to write summaries of 'the best supported claims' regarding a contested-scientific issue. Analysis focuses on the relationship between CIS processes and learning outcomes.

---

**Christopher Leeder**, Rutgers University

*Title: Library research as collaborative information seeking*

Today's students are accustomed to collaborative information behavior, with group work being a common requirement in educational settings. This talk presents the results of a study students conducting collaborative research using library resources. Participants used the Coagmento (<http://coagmento.org/>) collaborative search system in a library lab while working on a class assignment. The results demonstrate that there are benefits and drawbacks to collaborative information seeking. Findings showed that students working collaboratively found more useful sources and achieved greater information coverage, while individuals showed better results for query effectiveness and amount of relevant sources. Challenges that students face when conducting library research were identified. The findings of this study offer suggestions on how to support group work, and how collaborative search systems can address the challenges faced by students doing group work using library resources.

**Javed Mostafa**, University of North Carolina at Chapel Hill

*Title: Collaborative Search Challenges for Adaptive and Personalized Search for the Elderly*

Health information seeking, broadly, is one of the most popular usages of the Web (<http://www.pewinternet.org/fact-sheets/health-fact-sheet/>). In consumer-oriented health information seeking research there is a lack of attention to information seeking challenges that elderly users face. While seeking health information online, elderly users often involve and depend on other "co-consumers" of such information, particularly their caregivers and their physicians. Caregivers or physicians sometimes find themselves in the role of search result "interpreters", whereby they have to explain the search output or suggest improved search strategies. It is not uncommon for caregivers or physicians, on occasion, to engage in searching on behalf of elderly users. There is currently very little understanding of the dynamics of such collaborations in information seeking and consequently it is even rarer to find any search tools and systems designed specifically for such collaborative usage.

For the past 15 years or so, we have investigated a variety of different dimensions and challenges associated with personalized information retrieval. With the focus on delivering timely and highly accurate search results, we developed techniques and systems for content representation, profile acquisition, and interaction designs. One major area we investigated is machine learning approaches for combining cumulative knowledge from user profiles in collaborative or social information retrieval environments. Identifying interest profile similarities in social (networked) environments can be useful for a wide variety of purposes, for e.g., to address "cold start" challenges in profile acquisition or to discover cohorts of users that share interest in topics of mutual interest. Among many techniques we applied, a primary one we employed involved multi-agent learning to support personalization among groups of users.

Recently, we started exploring the problems associated with serving elderly searchers' needs by using a similar multi-agent paradigm. The vision is to acquire a set of three profiles, i.e., a profile-triplet, one for the elderly end-user, one for the caregiver, and one for the physician. The initial data for the profile-triplet is acquired from the context of an electronic health record, which represents the general health profile of an elderly patient. The subsequent step then is to apply machine learning approaches to adaptively update these profiles and combine the critical information in them to support a single user's personalization needs in an ongoing way. We recently started exploring a research collaboration with Dr. Phil Sloan, currently the Director of Program on Aging, Disability, and Long-Term Care. Dr. Sloan and his team developed a knowledge resource, called the Alzheimer's Medical Advisor (<http://alzmed.unc.edu/>), for serving information needs of caregivers who support elderly patients suffering from dementia.

*SCIS 2015 Workshop Report*

Two key goals of the collaboration are: 1) Develop intelligent, multi-agent personalization techniques that combine the information needs of patients, caregivers, and physicians to improve the quality and effectiveness of the information delivered and 2) Create theoretical frameworks and models that can improve our understanding of collaborative search, particularly as it pertains to consumer-health information seeking.

My interest in joining the workshop is to share some early conceptualizations of collaborative search challenges in the health domain, gather feedback and insights on possible directions for personalization approaches for collaborative search, and contribute toward continued development of the broader area of collaborative search through new initiatives, programs, and identification of funding sources.

---

**Douglas W. Oard**, University of Maryland

*Title: Collaborative Cross-Language Search*

In this talk, I will briefly review what little we know about collaborative Cross-Language Information Retrieval (CLIR), with an eye towards starting a discussion about a research agenda. To the usual list of talents that are distributed unevenly across a population of searchers (e.g., search strategies, domain expertise, and prior knowledge), collaborative CLIR adds language expertise. I will begin by reviewing a project on collaborative translation in which people with complementary language skills (one knowing the source language, the other the target language) worked together (with system assistance) to produce translations. I'll then briefly describe one session of a larger user study in which we sought to enhance recall, a challenge that is somewhat more difficult in CLIR than in monolingual settings, by reassigning low-yield topics to searchers who might bring different search strategies to bear. Drawing on this experience, I'll then step back to say a few words about what makes the CLIR setting different, and how those differences might help to inform the research agenda for collaborative information retrieval.

---

**Soo Young Rieh**, University of Michigan

*Title: Evaluation Measures in Social Search*

The Interactive Information Retrieval research has used a large number of evaluation measures including relevance, utility, efficiency, user satisfaction, and usability. In this talk, I propose to develop a set of evaluation measures to be used in social search that were derived from two empirical studies conducted in work places and social Q and A services. Four basic categories of measures will be discussed: performance measures, informational outcomes, social outcomes, and user experience. To measure performance, in addition to traditional criteria such as recall and precision, metrics for

*SCIS 2015 Workshop Report*

information diversity can be developed. The category of informational outcomes includes measuring information quality based on comprehensiveness, novelty, trustworthiness of information. The category of social outcomes refers to users' appreciations of other people's attempt, effort, responsiveness, and understanding of information needs. The category of user experience includes subjective assessment of overall search experience such as feeling of time was well spent, increased certainty in the problem, and perceived learning of new information. The future directions of developing and testing evaluation measures in social search will be also discussed.

---

**Chirag Shah**, Rutgers University

*Title:* Social and Collaborative Information Seeking (SCIS): Space, Time, and Beyond

Space and time are considered to be the most defining characteristics of classifying various collaborative and social activities. I will talk about new knowledge concerning these two important dimensions when it comes to SCIS tasks and how we obtained it. Specifically, I will outline various lab and field studies we have conducted to learn about different trade-offs we observe in people working in co-located vs. remote conditions, and people working synchronously vs. asynchronously. In addition, I will talk about the roles of other important dimensions in SCIS that we have identified: communication, awareness, affects, and group sizes. I will also allude to the connections between S and C elements of SCIS.

---

**Aiko Takazawa**, UIUC

*Title:* Social and Collaborative Information Seeking

Seven Japanese women living in Finland became leaders of a self-organized humanitarian aid group in response to the 2011 Great Tohoku Earthquake and Tsunami disaster in Japan. The way this group managed to send bulks of baby formula from Finland to Japan is a fascinating case to study for holistic understanding of how people collaboratively search, use, and seek information in the use of available technologies. Since this group emerged in a natural setting mediated by social media without being guided through an established affiliation among participants or managed by an outside source, its emerging process of becoming and being a group provides deep insights into the substantive context for intertwined, various kinds of both individual and collaborative information activities. I claim that such messiness in the present case represents the reality of ordinary people living in this present ICT-mediated environment, although what the group ended up doing transcended the ordinary. From a broad perspective, this case demonstrates the potential for expanding existing concepts relevant to Social and Collaborative Information Seeking research by looking at its gradually constructed

*SCIS 2015 Workshop Report*

information needs, resulting from browsing in social context, serendipitous searching, and collaborative learning.

The present case's information activities are situated in its particular circumstances as those concerned individuals shared a vague aspiration and expressed with a strong compassion for offering meaningful aid directly and immediately to the victims of the 2011 national crisis in Japan. A loosely formed assemblage of like-minded individuals started conversing on social media spaces, and their "conversations" were carried out on different platforms involving both digital and physical spaces. Moreover, a number of other processing activities (or steps) were necessary in parallel in order to implement the idea of sending the baby formula, for example: the correspondence and reporting about the local transport (e.g. talking to FinnAir Cargo), the coordination of tasks and procedures, the fundraising and procurement of the baby formula, the preparation for the exportation of emergency food rations, and the packaging and loading work. How the TTJ evolved and completed six shipments of 12,000 cartons of formula in a timely manner draws an analogy of the reality of how information activities evolve in social contexts and becomes powerful; how people engage in collaborative information seeking and search even without knowing a particular need of information.

Using the publicly available data, particularly the group's Twitter activity and weblog, as well as interview data from few participants, I am currently trying to identify basic features that explain how indirect, opportunistic collaboration among like minded individuals shaped as they worked on unstructured information tasks situated in social technologies. In this workshop, I would like to discuss how to make sense of social data drawn from a microcosm case study. I want to understand how ordinary people, information, and technologies interact and how their intertwining social and collaborative information activities enable them to find a way that brings meaningful accomplishments.

---

**Sandra Toze**, Dalhousie University

*Title:* Exploring the Group within Social and Collaborative Search

Within social and collaborative search the "Group" represents a unique level with specific attributes (interaction, interdependence, awareness and shared understanding) that need to be better understood and supported. To address this gap my research has focussed on understanding and modelling information needs, seeking and use at the group level. For my dissertation research I collected longitudinal data from seven student groups working on multiple tasks over time, and used a structured task analysis to deconstruct when and how a group 1) identifies information needs 2) satisfies these needs through seeking through various channels, and 3) how a group collectively uses this found information to solve problems, make decisions and generate something new.

*SCIS 2015 Workshop Report*

To ensure motivated participants, the student groups I recruited completed class based assignments that were independent of my research and represented a significant part of their final class grade. Data collected included 60 hours of video and log files, from 25 different group sessions. This longitudinal data collected is unique within the field of group research. The naturalistic lab study method addressed a key methodological challenge of studying groups, allowing the complex details of group work to be captured as they unfolded naturally over time.

To guide my analysis I combined a rhythm based model of group task accomplishment (Marks, Mathieu, & Zaccaro, 2001) with an information behaviour lens (Choo, 2006; Marchionini, 1995; Wilson, 1999) to create an integrative framework. I first analyzed the procedural aspects of group work and found that all the groups shifted between three phases of group activities: Planning, Doing, and Monitoring. Within each phase I then identified and described the elements of a group information process: the information tasks, information task goals, information activities, sources, tools, artefacts, roles and shifts in participation. Groups looked for information to satisfy eight different goals, requiring 19 different information activities, as well as specific sources and tools to generate new artefacts. Ten roles were observed within the groups to manage their information activities, and participation fluctuated from individual through to the group. The relationship between these elements was described. Integrative analysis revealed that the student groups did not have good mechanisms for managing information needs, and encountered the greatest difficulties trying to use information collectively. Additionally challenges when searching together were identified.

Based on the findings of my research I made recommendations for tools and processes to facilitate more effective group work. My definition and conceptual model of Group Information Process extends our understanding of information behaviour within groups, and provides a base that can be used to ground further research. In addition, and of particular interest for this workshop is the richness of my data, the process of analyzing complex group data, and the natural lab methodology. These represent new avenues for moving the research agenda forward. Currently I am investigating collaborative information use in groups by analyzing key moments from the videos using conversation analysis to examine the relationship between the information activities and the formation of shared understanding.

---

**Michael Twidale**, University of Illinois Urbana-Champaign

*Title:* Searching for help: how learning technologies involves collaborative search

As computational and informational resources become ever more abundant, we see changes in the way people learn how to use them, adopt, adapt, appropriate, tinker, tailor, combine and modify them. Examples include software developers who search as

they code, and data scientists going online to get ideas for how best to clean, combine and manipulate datasets. However such activities are not restricted to the computational elites. Across all levels, tech learning is often both a search and also a social activity, synchronous and asynchronous, co-located and remote, with colleagues and strangers.

Doing this kind of searching as part of technology learning and problem solving accentuates particular difficulties in the search process. Various strategies and tactics can dramatically improve efficiency, and equally a lack of certain skills the possession of certain misconceptions can degrade people's ability to learn and cope, and even lead them to self-define as "not- techie". This raises important implications for design, policy and education.

---

**Jyothi Vinjumur**, University of Maryland

*Title:* Reducing E-Discovery Cost with Collaborative Review Process

Seeking relevant information and yet protecting sensitive content that could be intermixed with relevant information are two different goals, but in certain situations, a balance must be struck between the two. One example is, the protection of content that is subject to a claim of attorney-client privilege when sharing responsive evidence incident to civil litigation, a process called e- discovery. In e-discovery, the use of automated retrieval techniques to retrieve responsive evidence have not brought the hoped-for cost savings, since attorneys are reluctant to trust automated methods for privilege review (example: attorney- client privilege), and therefore frequently advocate on manual review on the responsive set. Such manual assessments of privilege require expert legal knowledge causing the review procedure to be very expensive. Thus, the main question in e-discovery is not just about what the technology is able to do or how legal professionals use what we build, but also about how the legal professionals and the technology could collaborate to ensure proper production at a proportionate cost. Although legal professionals have been quick to embrace technology supported retrieval and review techniques, the increasing volume of potential electronic evidence that must be reviewed is just overwhelming. Many factors like context, cognition, annotator expertise, etc., affect the process and quality of the review process. In privilege review task, the nature of the content to be protected can be generally unknown to the reviewer in advance or different reviewers may share different opinions. In addition to the review complexity, it is not realistic for human reviewers to be infallible. The intuition in this paper is that, review process could be effective in time and performance as a collaborative task than as a solitary activity. Thus, this paper aims to build an interactive web- based system to corroborate the manual review process by facilitating explicit collaboration among annotators to review huge masses of electronic evidence with the goal to optimize for e-discovery cost (both review and training). In order to arrive at high quality and cost effective relevance/privilege assessments, this

paper proposes a first step of building a Collaborative Technology-Assisted Review (CTAR) tool that can support lawyers to make faster and more accurate judgments during review.

---

**Hassan Zamir**, University of South Carolina

*Title:* Social Searching and Information Recommendation Systems

I am strongly interested to participate at the Social and Collaborative Information Seeking workshop to learn more about how social and collaborative relationships are useful in information searching. Currently, I am at the early stage of writing my doctoral dissertation that concentrates on usefulness of social media data to recommend right information contents to the right users at right time. An active practicum on theories, models, techniques and usefulness of social and collaborative search will be very valuable to plan and select appropriate methodologies for my research. Apart from my dissertation research, I usually conduct investigations in the areas related with social media information retrieval.

Omnipresence of social media tools enable people to produce and reproduce contents instantly and share those with the world almost effortlessly. It makes the task of information seeking more complex and challenging specially to retrieve relevant items. Online social media users actively use various web 2.0 tools to report social crisis and events, protests, occurrences, natural disasters, political debates, policy dialogue and many more. To ensure quick information retrieval, social media sites adopt user-friendly mechanisms such as tagging, organizing topics by trends, searching, categorizing contents by general interest based subjects etc. However, information gathering is a combined task that requires explicit and implicit collaboration and participation of others. Social media tools harness the power of crowds purposefully, which makes the task of information recommendation easier and convenient. I have interests to develop systems that can recommend information contents to the users based on information available on social platforms. Companies like Amazon, Netflix, Pandora etc. are common examples that are widely using recommendation methods based on collaborative filtering. Applications of recommendations methods have potentials to fit in the libraries as well, although patron privacy issues need to be handled carefully. General distance computing algorithms such as Manhattan distance, Euclidean distance, Pearson correlation coefficient etc. will be helpful to identify similar books or information resources and suggest those to library patrons. Similar data mining techniques evidently work well with online social media data. Currently I am conducting my dissertation research that focuses on social movements tweets with a purpose to examine how Twitter can suggest information contents to tweeters. The role of explicit and implicit filtering in this context needs to be investigated as well. In the case of Twitter, 'favorite', 'retweet', and 'follow' information can be utilized to observe explicit behavior of tweeters.

Furthermore, implicit users' behaviors including click through, eye-tracking, information generating and sharing behavior etc. can potentially recognize personalized information preferences. Eventually, this technique has implications on grouping and referring information contents to the users.

---

**Yinglong Zhang**, The University of Texas at Austin

*Title:* Culture and Trust in Collaborative Information Seeking

Why do people need collaborative searches? One of the common motivations for individuals to use collaborative search is that people trust their friends more than strangers. In my prior research, it has been found that people are prone to judge information as irrelevant and refuse to use it when they consider it is unreliable. It seems that the success of collaborative work is largely based on whether members can trust each other in a collaborative group. Culture is one of the many important factors that heavily influence development of trust in collaboration. Individuals from different cultures can interpret and understand the same problem in a distinct way based on their own cultural knowledge and beliefs. The gaps in problem representation can increase misunderstanding and conflicts in collaboration, therefore weakening trust among the individuals. Aiming to address this issue in a context of collaborative information seeking, I am interested in investigating what factors will contribute to the development of trust in an intercultural group and how to design collaborative search systems that can make people trust each other more. Influenced by theories and methods from Human-Computer Interaction as well as cognitive science, I seek to adopt quantitative and qualitative methods to address questions of interest.

---

**Appendix B: Group Reports**

Edited by Chirag Shah

**Group #1 – Contexts for SCIS, Cross-Language Retrieval**

Prepared by Jyothi Vinjumur and Daqing He

The working group 1 discussed several emerging problems that can help to put SCIS into real and critical contexts, such as e-discovery, cross-language access to culture heritage collections, or military related decision making.

**Research Problems**

The research problems identified include

1. Where the collaborative is in the search/seeking process?
2. What is difference between CIR in general knowledge and specific knowledge, health, politics, education, legal, scholarly?
3. How do these levels of collaboration (social context, culture, cross expertise, and potential reuse) play in CIS?
4. What kind of system supports can be developed, evaluated?
5. Information flow among different components - how to facilitate the flow of information?
6. How to get data set? What could the right dataset, evaluation framework, what could be the right dataset?
7. What kind of basic system components could be? It is related to what we see collaborative seeking is, such as social sharing, visualization, sense making, and recommendation?
8. Different expertise - how this can be integrated into the collaborative, not just different roles? How to know the expertise? How to evaluate the expertise. Could be domain expertise and search expertise? Can we trust the person's expertise?
9. Temporal difference, sometime very intensive among different partners, sometimes is relatively slow in pace.
10. How mobile access and multiple types of devices play in CIS?

Then the group worked further on two specific research context. The first one concerned collaborative information access in the context of cross-language access to culture heritage collections. The other one was on e-discovery.

**Detail Design of Research Projects**

The assumed title of the research program in the context of cross-language access to culture heritage collections is "Synchronous Remote Collaborative Cross-Language Search for Access to Cultural Heritage Materials by Scholars".

### **Cross-Language Collaborative Access to Culture Heritage Collection**

In this research program, we assume that there is a centralized hybrid collection of culture heritage multiple media documents, which can be books, papers, maps, images, video, audio, artifacts, etc. All these documents can be classified in physical content (partial metadata-based catalog), digitized content (perhaps with document image retrieval) and born-digital content (with full-content search). These documents are written in multiple languages with one principal non-English language that covers 85% of the collection and 16 other languages for last 15%.

The research questions in this program include

- How to work together to do better CLIR than one person can do alone?
  - Baseline: fully automatic interactive CLIR with MT
- How best to use limited access to human “translation” service
  - Translator/Interpreter without (or with) domain expertise
    - Option 1: help craft good queries
    - Option 2: explain what has been found
    - Option 3: Do some translation (for use, for assessment, for tuning MT systems)
  - Translator/Interpreter with domain expertise
    - Option 4: assess what’s been found and iterative improve query
    - Option 5: Find documents and then translate them for you
- How best to communicate with curator who doesn’t know searcher’s language
  - Option 1: Interactive “chat translation” and ability to point
  - Option 2: Limited on-demand machine translation

The different partners can engage collaborative information seeking at various stage of the process. These partners can include:

- Collaborative Search Research Team
  - Information Retrieval Team
    - CLIR system developer
    - Format-specific indexing (e.g., XML, finding aids, OCR, speech)
  - Machine Translation Team
    - Inhat translation developer
    - Document translation grad student
  - UX Developer
    - CL-Coagmento developer grad student
    - User study grad student
- Domain “Science” Team
  - Historian, History grad students, Archivist
- Supporting staff
  - Translator, Simultaneous interpreter

## **Group #2 – Contexts for SCIS, E-Discovery (Mitigating the Cost of Review in E-Discovery using Collaborative Annotation)**

Prepared by Doug Oard, Jeremy Pickens, Daqing He, and Jyothi Vinjumur

### **1 Introduction**

In United States, civil litigation is a legal dispute between two or more parties where either hold the right to request relevant evidence from each other, a process called “discovery.” Traditionally, discovery focused on materials that are paper documents. Since most documents today are in electronic format, the meaning of “qualifying evidence” has experienced a definitional change over time. The amendments to the Federal Rules of Civil Procedure in December 2006 require that the traditional discovery process address the discovery of the Electronically Stored Information (ESI). Discovery (in the context of ESI called “e-discovery”) poses a multi-faceted challenge to information retrieval systems; the need to initially find the evidence that has been requested, and among the requested evidence the obligation to identify (to withhold) privileged evidence. While the failure to find the requested evidence jeopardizes the interests of the requesting party, failure to identify privileged evidence jeopardizes the interests of the responding party. Thus the 2006 amendments and its implications on the use of Technology-Assisted Review (TAR) in e-discovery (“Rule 26(g)”) have compelled litigators to exercise great care in the discovery of the ESI to ensure proper production at a proportionate cost.

Costs in e-discovery break down along the following lines: (1) the cost of data collection; (2) the cost of data processing, and (3) the cost of review. Since document review has been estimated to account for almost 60% of all litigation costs, the current trend is to reduce the amount of content that requires manual review. In many large cases, third-party contractors perform review. The cost of contracted reviewers depends on their hourly rate and their review speed, both of which vary with expertise and experience, and on how many documents must be reviewed. Hence it is the cost factor associated with human review that we seek to optimize using collaboration.

### **2 Research Questions**

The research objective we discuss here is to reduce the cost of manual review using collaboration. We aim to design a Semi-automated Collaborative Technology Assisted Review (CTAR) system using algorithmic mediation to gather high quality relevance judgments. The main goal of the CTAR system is to support human reviewers annotate the documents for relevance and/or privilege. While this support can take many forms, we propose to model that support in two steps; (1) By training a document ranker that ranks documents in the order in which they should be reviewed and (2) By providing system mediated visual clues (by highlighting the features that signify graded importance) to the human annotator that indicate the relevance and/or privilege

propensities generated by the model. A prototype of the current Graphical User-Interface of CTAR system is shown in the bottom-left corner of the Figure 1.

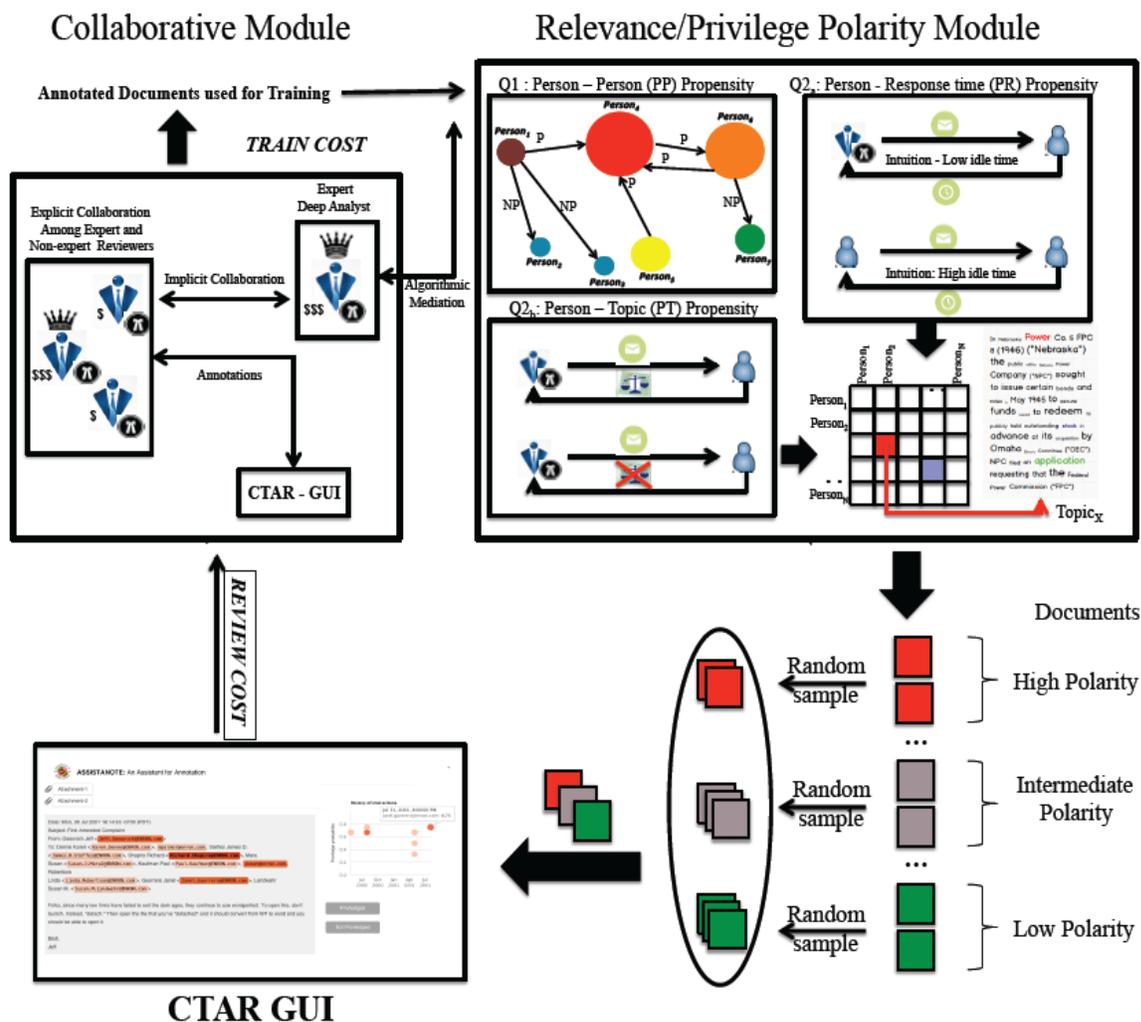


Fig. 1: Research problem overview

Thus we categorize the research problem into two modules: (1) Polarity Module and (2) Collaborative Module. These modules exhibit a cyclic dependency; that is, the knowledge captured due to human interaction in collaborative module is used by the polarity module to compute propensity scores and the propensity scores thus computed are provided as guidance for annotators during collaboration<sup>1</sup>.

The Polarity Module investigates the following research question:

<sup>1</sup> Note that propensity scores are computed for individual features in the document while polarity score is computed for the document.

*RQ1: To what extent can automation accurately determine the factors that could influence relevance and/or privilege?*

Rather than engineering an automated classification as the goal, we apply a modest step to compute a polarity score for each document. The document polarity score is used to filter and determine the order in which the document needs to be reviewed. The propensity score of each feature in the document is intended to support lawyers in making faster yet accurate judgments by providing the visual clues. The polarity scoring technique for each document depends on the individual feature propensity scores. The Polarity Module framework is graphically illustrated on the top-right corner of Figure 1.

The Collaborative Module investigates the following research questions:

*RQ2a: Does explicit collaboration between non-expert and expert reviewers reduce the e-discovery annotation cost?*

*RQ2b: Does using an expert analyst to perform deep analysis improve system performance?*

The Collaborative Module utilizes the document polarity score and the features' propensity scores. Based on the polarity scores, documents that need to be reviewed are categorized into three bins; "high-polarity", "low-polarity" and "intermediate-polarity" documents. Documents with high polarity score exhibit features with high relevance and/or privilege propensity scores, those with low polarity exhibit low propensity scores and those documents with an intermediate polarity score can be considered as the border-line document cases which are particularly hard for the classifier to predict. A random sample of documents is then drawn from each of the three bins for manual review (with high sampling rate for documents in the "intermediate-polarity" bin). The propensity scores of the features in each document provide visual scaffolds during review. The principle focus of the collaborative module is to facilitate collaboration between annotators (expert and non-experts) and validate the review process. Since one of the factors that account for the increase in review cost is dependent on reviewer expertise, our approach utilizes utility theory to compute the information gain for each reviewer to decide which set of annotations (when used for training) yield better classifier performance.

**Group #3 – Health Information in SCIS**

Prepared by Mark Ackerman, Kaitlin L. Costello, and Javed Mostafa

This group convened at the workshop to discuss collaborative information seeking issues related specifically to health. Our research questions mainly focused on the temporal, affective, and collaborative aspects of information and information platforms/software systems in the health arena. Specifically, we discussed questions such as:

- How do you deal with information needs surrounding co-morbidities?
- How can we better understand the phases of health – chronic conditions do not tend to evolve linearly, but in a more messy fashion? How do we handle this within information systems?
- What happens with respect to information management when people are involved with the healthcare system over a long period of time?
- What relevance cues do patients use to evaluate collaborative health information?
- How is information delivered? What is the role of the intermediary in collaborative information sharing (e.g. caregivers, providers).
- How do we take into account both the temporal and affective aspects of health information behavior?
- What is the interaction between information behavior and relationships with people in health?
- How does health information seeking for patients diagnosed with chronic conditions change over time? What other human sources do they engage, when, and how?

We also discussed how collaborative information behaviors in health are essentially a “wicked problem” – a phrase used to discuss issues that are very difficult to solve because the requirements of the solution are constantly evolving over time, and may even compete with themselves (Churchman, 1967). We discussed the difficulty in operationalizing the problem when it is so complex, and determined that the best course of action in the collaborative health information seeking arena is to develop exemplary cases in medical collaborative information behavior so that we can generalize and characterize tasks. This would allow researchers in the field to work on building information platforms and systems geared towards some of the common collaborative information-based tasks that occur in health. Therefore, the research question that first needs to be addressed is: What kind of cases would be important when developing exemplars and test collections for researchers in collaborative health information behavior? Based on the research already completed by the members of this group, we came up with several ideas while brainstorming:

*SCIS 2015 Workshop Report*

- New diagnosis of a life-changing chronic illness (both stigmatized and non-stigmatized illnesses)
- Diagnosis of illness with multiple care or treatment options requiring patient decision-making
- End-of-life and hospice care decisions
- Patients diagnosed with chronic illnesses making other life decisions (e.g. moving to a new city) that impact their health records-keeping and health information management

It is important to note that this list is not intended to be comprehensive. A systematic review of the literature on collaborative health information behavior is an obvious next step in order to better understand what tasks may be useful to pursue in more depth.

We concluded our working group discussions by talking about how difficult this problem is – fundamentally, this is a problem that involves a very complex interleaving interaction of people and information and the links between and among different actors and different information sources. We currently do not have flexible, usable platforms that are truly useful for managing the social aspect of health over time. This wicked problem moves past data into more conceptual issues: it is not just an issue of creating algorithms or wrangling data; it is not just about deploying technical systems: it is about how to best leverage the research community to specifically address the sociotechnical aspects of health. Health is particularly important as a research area in collaborative information behaviors specifically because it is a wicked problem and therefore involves a very complicated set of moving targets. We plan to address this more comprehensively by creating a review or survey article and perhaps by attracting researchers in the area to contribute journal articles on the topic to a special journal issue devoted to this fairly intractable social problem of collaborative health information behavior.

*References*

Churchman, C.W. (1967). Wicked problems. *Management Science* 14(4).  
Doi:10.1287/mnsc.14.4.B141.

**Group #4 – Learning and SCIS**

Prepared by Simon Knight, Chris Leeder, Sandra Toze, Michael Twidale, and Soo Young Rieh

**Key Issues**

One focal interest of the SCIS attendees was 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).

Some key questions emerged:

1. What skills help people in SCIS? Which of these are particular to SCIS versus other collaborative, or information seeking contexts?
2. What relationships can we identify between SCIS motivation, and the SCIS process/outcome? E.g. personal-health, library-search, college assignment, experimental context might lead to differing precision/recall/evaluative needs and satisficing/deeper understanding related to it, and varying processes including collaboration or information-seeking at varying stages
3. How do we measure SCIS success, and should it be at the individual or small group level (see e.g. Gerry Stahl's work)
4. Dimensions of CIS: technologies used; affect, trust, cognition; expertise level; roles; motivation; domain expertise v search expertise; output; physical location;
5. Where does collaborative search happen that we're not aware of? (What can we learn from how they are doing collaborative search?)
6. Where is collaborative search not happening but it could? And why not? What are the barriers?
7. What are the barriers to learning in SCIS?

Why does collaborative help learning? Why does collab search help learning?

- a. Externalizing learning - addresses the 'vocabulary problem', provides diversity of perspective
- b. Makes the search process (and output) an "improvable object"
- c. Because specific contexts are designed to maximize learning (i.e., these things don't just happen by themselves)
- d. Varying contexts, e.g. 'over the shoulder' suggestions, versus sustained project-based collaboration

8. How do we understand the desire to support learning processes alongside specific outcomes? E.g. issues around quality v quantity of contribution (some people say/do less, but have a larger impact). What processes does SCIS foster that are related to learning?

### Themes

- Learning to search or searching to learn
- Use of such trails (of experts or earlier searchers) to support learning - what kind of visualizations and content support learner understanding and how? Do they support learning to search, or searching to learn?
- SCIS as a context in which to explore information seeking across multiple (conflicting?) documents
- Temporality and the process of learning through searching
- Complexity of evaluation

### Research contexts

There's a need to better understand SCIS in formal and informal learning contexts. There are likely to be a diverse set of learning contexts in which SCIS occurs - co-located, remote, asynchronous, synchronous, and combining all of these at various stages. We want to understand how SCIS is linked to learning.

- Various methods (lab, field, interview, survey, observation, log data, etc.) could be used to understand the processes.
- To assess outcomes various methods could be used, including: a pre-test, post-test design; between-group design; concept maps capture structure of understanding, subject expert creates exemplar to compare to participants outcomes;
- Various contextual factors might play a role, including: space; time; Common ground amongst collaborators; Prior knowledge; Education level (institutional); Group size; Resources/Tools; Task difficulty/complexity
- Task design might be self-directed, assigned, or semi-assigned (e.g. a theme is given, but specifics are self-directed)
- Task topic/structure might be more or less exploratory. Precision oriented tasks can still be complex (e.g. Dan Russell's tasks <http://searchresearch1.blogspot.co.uk/>). Exploratory tasks might involve, e.g. researching socio-scientific topics about which there are varying (scientific or otherwise) perspectives such as fracking.

### SCIS 2015 Workshop Report

- We are often interested in processes in SCIS for learning, for example: Socialization and communication; coverage/recall/precision; search query sophistication (e.g. query depth, query-vocabulary-richness);
- In some contexts, learners might get most from reflecting on how different groups have engaged in the search process, and what they have found. Research could explore this reflective-SCIS (e.g. through collaborative sensemaking over search trails)

### Education resources

#### Resources to support educators using SCIS in their practice, or to support learners

- <http://sjknight.com/finding-knowledge/edusearch-tips/> Some tips (a bit out of date now)
- <http://agoogleaday.com> - moderately complex precision oriented tasks
- <http://searchresearch1.blogspot.co.uk/> - complex precision oriented tasks

#### Software resources:

- <http://coagmento.org>
- <http://trailblazer.io/>
- <https://addons.mozilla.org/en-US/firefox/addon/mywebsteps/>

### References/bibliography

#### Research on SCIS/relevant themes in education

1. Cunnane, S. (2011, May 12). The Danish gambit: online access, even during exams. *Times Higher Education*. UK. Retrieved from <http://www.timeshighereducation.co.uk/story.asp?sectioncode=26&storycode=416090&c=1>
2. Dimopoulos, K., & Asimakopoulos, A. (2010). Science on the Web: Secondary School Students' Navigation Patterns and Preferred Pages' Characteristics. *Journal of Science Education and Technology*, 19(3), 246–265. <http://doi.org/10.1007/s10956-009-9197-8>
3. Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research and Development*, 51(1), 21–38.
4. Ge, X., & Land, S. M. (2004). A conceptual framework for scaffolding Ill-structured problem-solving processes using question prompts and peer interactions. *Educational Technology Research and Development*, 52(2), 5–22.
5. Hong, H.-Y., & Lin, S.-P. (2010). Teacher-Education Students' Epistemological Belief Change through Collaborative Knowledge Building. *The Asia-Pacific Education Researcher*, 19(1). Retrieved from

<http://www.ejournals.ph/index.php?journal=TAPER&page=article&op=view&path%5B%5D=129>

6. Hsu, C.-Y., Tsai, M.-J., Hou, H.-T., & Tsai, C.-C. (2013). Epistemic Beliefs, Online Search Strategies, and Behavioral Patterns While Exploring Socioscientific Issues. *Journal of Science Education and Technology*, 1–10. <http://doi.org/10.1007/s10956-013-9477-1>
7. Hwang, G.-J., Tsai, P.-S., Tsai, C.-C., & Tseng, J. C. R. (2008). A novel approach for assisting teachers in analyzing student web-searching behaviors. *Computers & Education*, 51(2), 926–938. <http://doi.org/10.1016/j.compedu.2007.09.011>
8. Hyldegård, J. (2006). Collaborative information behaviour—exploring Kuhlthau’s Information Search Process model in a group-based educational setting. *Information Processing & Management*, 42(1), 276–298. <http://doi.org/10.1016/j.ipm.2004.06.013>
9. Hyldegård, J. (2009). Beyond the search process – Exploring group members’ information behavior in context. *Information Processing & Management*, 45(1), 142–158. <http://doi.org/10.1016/j.ipm.2008.05.007>
10. Jucks, R., & Paus, E. (2013). Different Words for the Same Concept: Learning Collaboratively From Multiple Documents. *Cognition and Instruction*, 31(2), 227–254. <http://doi.org/10.1080/07370008.2013.769993>
11. Knight, S. (2014). Finding Knowledge: What Is It To “Know” When We Search? In *Society of the Query Reader: Reflections on Web Search*. Amsterdam, The Netherlands: Institute of Network Cultures. Retrieved from <http://networkcultures.org/wpmu/portal/publication/society-of-the-query-reader-reflections-on-web-search/> individual chapter from <http://oro.open.ac.uk/39307/>
12. Knight, S., Arastoopour, G., Williamson Shaffer, D., Buckingham Shum, S., & Littleton, K. (2014). Epistemic Networks for Epistemic Commitments. In *International Conference of the Learning Sciences*. Boulder, CO: International Society of the Learning Sciences. Retrieved from <http://oro.open.ac.uk/39254/>
13. Knight, S., Buckingham Shum, S., & Littleton, K. (2013c). Tracking Epistemic Beliefs and Sensemaking in Collaborative Information Retrieval. In *Collaborative Information Seeking Workshop*. San Antonio, Texas, USA. Retrieved from <http://collab.infoseeking.org/events/cscw2013workshop>
14. Knight, S., & Littleton, K. ((Accepted)). Learning through Collaborative Information Seeking. In P. Hansen, C. Shah, & C.-P. Klas (Eds.), *Collaborative Information Seeking: Best practices, New Domains, New Thoughts*. Springer. Retrieved from <http://collab.infoseeking.org/cisbook2015/>
15. Knight, S., & Littleton, K. (2015a). Developing a multiple-document-processing performance assessment for epistemic literacy. Presented at the The 5th International Learning Analytics & Knowledge Conference (LAK15): Scaling Up: Big Data to Big Impact, Poughkeepsie, NY, USA (Forthcoming). Retrieved from <http://lak15.solaresearch.org/>

16. Knight, S., & Mercer, N. (2014). The role of exploratory talk in classroom search engine tasks. *Technology, Pedagogy and Education*. <http://doi.org/10.1080/01475939.2014.931884>
17. Kobayashi, K. (2014). Students' consideration of source information during the reading of multiple texts and its effect on intertextual conflict resolution. *Instructional Science*, 42(2), 183–205.
18. Kuiper, E., Volman, M., & Terwel, J. (2009). Developing Web literacy in collaborative inquiry activities. *Computers & Education*, 52(3), 668–680. <http://doi.org/10.1016/j.compedu.2008.11.010>
19. Large, A., Beheshti, J., & Rahman, T. (2002). Gender differences in collaborative Web searching behavior: an elementary school study. *Information Processing & Management*, 38(3), 427–443. [http://doi.org/10.1016/S0306-4573\(01\)00034-6](http://doi.org/10.1016/S0306-4573(01)00034-6)
20. Lazonder, A. W. (2005). Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes. *British Journal of Educational Technology*, 36(3), 465–475. <http://doi.org/10.1111/j.1467-8535.2005.00478.x>
21. Liechti, O., & Sumi, Y. (2002). Editorial: Awareness and the WWW. *International Journal of Human-Computer Studies*, 56(1), 1–5.
22. Lin, C.-C., & Tsai, C.-C. (2007). A Navigation Flow Map Method of Representing Students' Searching Behaviors and Strategies on the Web, with Relation to Searching Outcomes. *CyberPsychology & Behavior*, 10(5), 689–695. <http://doi.org/10.1089/cpb.2007.9969>
23. Lin, C., & Tsai, C. (2008). Exploring the Structural Relationships between High School Students' Scientific Epistemological Views and their Utilization of Information Commitments toward Online Science Information. *International Journal of Science Education*, 30(15), 2001–2022. <http://doi.org/10.1080/09500690701613733>
24. Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the Web to learn about a controversial topic: are students epistemically active? *Instructional Science*, 38(6), 607–633. <http://doi.org/10.1007/s11251-008-9089-y>
25. Mokhtari, H. (2014). A Quantitative Survey on the Influence of Students' Epistemic Beliefs on Their General Information Seeking Behavior. *The Journal of Academic Librarianship*, in press. <http://doi.org/10.1016/j.acalib.2014.04.001>
26. Park, Y., & Lim, H. (2012). Meaning of Epistemological Belief Through Online Communication: Exploratory Study. *Procedia - Social and Behavioral Sciences*, 46, 3254–3258. <http://doi.org/10.1016/j.sbspro.2012.06.046>
27. Puustinen, M., & Rouet, J.-F. (2009). Learning with new technologies: Help seeking and information searching revisited. *Computers & Education*, 53(4), 1014–1019. <http://doi.org/10.1016/j.compedu.2008.07.002>

28. Robson, A., & Robinson, L. (2013). Building on models of information behaviour: linking information seeking and communication. *Journal of Documentation*, 69(2), 169–193. <http://doi.org/10.1108/00220411311300039>
29. Salmerón, L., Kammerer, Y., & García-Carrión, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior*, 29(6), 2161–2171. <http://doi.org/10.1016/j.chb.2013.04.034>
30. Scherr, R. E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction*, 27(2), 147–174. <http://doi.org/10.1080/07370000902797379>
31. Sundin, O., & Francke, H. (2009). In search of credibility: pupils' information practices in learning environments. *Information Research*, 14(4). Retrieved from <http://www.informationr.net/ir/14-4/paper418.html>
32. Teng, Y.-T. (2010). *The relationship between epistemic beliefs and knowledge contribution to online communities of practice* (PhD). University of Illinois at Urbana-Champaign. Retrieved from <https://www.ideals.illinois.edu/handle/2142/16126>
33. Tsai, M.-J., Hsu, C.-Y., & Tsai, C.-C. (2012). Investigation of High School Students' Online Science Information Searching Performance: The Role of Implicit and Explicit Strategies. *Journal of Science Education and Technology*, 21(2), 246–254. <http://doi.org/10.1007/s10956-011-9307-2>
34. Tsai, P.-S., Tsai, C.-C., & Hwang, G.-J. (2011). The correlates of Taiwan teachers' epistemological beliefs concerning Internet environments, online search strategies, and search outcomes. *The Internet and Higher Education*, 14(1), 54–63. <http://doi.org/10.1016/j.iheduc.2010.03.003>
35. Tseng, J. C. R., Hwang, G.-J., Tsai, P.-S., & Tsai, C.-C. (2009). META-ANALYZER: A WEB-BASED LEARNING ENVIRONMENT FOR ANALYZING STUDENT INFORMATION SEARCHING BEHAVIORS. *Computing, Information and Control*, 5(3), 567–579.
36. Walraven, A., Brand-Gruwel, S., & Boshuizen, H. (2008). Information-problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior*, 24(3), 623–648. <http://doi.org/10.1016/j.chb.2007.01.030>
37. Wopereis, I. G. J. H., & van Merriënboer, J. J. G. (2011). Evaluating text-based information on the World Wide Web. *Learning and Instruction*, 21(2), 232–237. <http://doi.org/10.1016/j.learninstruc.2010.02.003>
38. Wu, Y., & Tsai, C. (2005). Information commitments: evaluative standards and information searching strategies in web-based learning environments. *Journal of Computer Assisted Learning*, 21(5), 374–385. <http://doi.org/10.1111/j.1365-2729.2005.00144.x>
39. Wu, Y.-T., & Tsai, C.-C. (2005). Developing the Information Commitment Survey for assessing college and graduate students' evaluative standards for Web information and their searching strategies in Web-based learning environments. In *World*

*SCIS 2015 Workshop Report*

*Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (Vol. 2005, pp. 2500–2512). Retrieved from

[http://www.editlib.org/INDEX.CFM?fuseaction=Reader.ViewPresentation&paper\\_id=21573&paperfile\\_id=34292](http://www.editlib.org/INDEX.CFM?fuseaction=Reader.ViewPresentation&paper_id=21573&paperfile_id=34292)

40. Wu, Y.-T., & Tsai, C.-C. (2007). Developing an Information Commitment Survey for assessing students' web information searching strategies and evaluative standards for web materials. *JOURNAL OF EDUCATIONAL TECHNOLOGY AND SOCIETY*, 10(2), 120.
41. Yue, Z., Han, S., & He, D. (2012). Search Tactics in Collaborative Exploratory Web Search. In *Proceedings of HCIR 2012*. Cambridge, MA. Retrieved from [http://crystal.exp.sis.pitt.edu:8080/ips/paper/hcir2012\\_cs.pdf](http://crystal.exp.sis.pitt.edu:8080/ips/paper/hcir2012_cs.pdf)
42. Yue, Z., Han, S., & He, D. (n.d.). An Investigation of the Query Behavior in Task-based Collaborative Exploratory Web Search. Retrieved from <https://asis.org/asist2013/proceedings/submissions/papers/120paper.pdf>

**Some more general refs (not comprehensive)**

1. Amershi, S., & Morris, M. R. (2008). CoSearch: a system for co-located collaborative web search. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1647–1656). Retrieved from <http://research.microsoft.com/en-us/um/people/merrie/papers/cosearch.pdf>
2. Amershi, S., & Morris, M. R. (2009). Co-located collaborative web search. ACM Press. <http://doi.org/10.1145/1520340.1520547>
3. Capra, R., Chen, A. T., Hawthorne, K., Arguello, J., Shaw, L., & Marchionini, G. (2012). Design and evaluation of a system to support collaborative search. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–10.
4. Capra, R., Chen, A. T., McArthur, E., & Davis, N. (2013). Searcher Actions and Strategies in Asynchronous Collaborative Search. Presented at the ASIST. Retrieved from <https://asis.org/asist2013/proceedings/submissions/papers/91paper.pdf>
5. Capra, R., Marchionini, G., Velasco-Martin, J., & Muller, K. (2010). Tools-at-hand and learning in multi-session, collaborative search (p. 951). ACM Press. <http://doi.org/10.1145/1753326.1753468>
6. Chi, E. H., Pirolli, P., & Lam, S. K. (2007). Aspects of augmented social cognition: Social information foraging and social search. In *Proceedings of the 2nd international conference on Online communities and social computing* (pp. 60–69). Retrieved from <http://dl.acm.org/citation.cfm?id=1784305>
7. Crescenzi, A., & Capra, R. (2013). Collaboration in Search Processes and Results Sharing. In *Collaborative Information Seeking Workshop*. San Antonio, Texas, USA. Retrieved from <http://collab.infoseeking.org/resources/papers/cis2013/CrescenziPaper.pdf>

8. Ellis, D., Wilson, T. D., Ford, N., Foster, A., Lam, H. M., Burton, R., & Spink, A. (2002). Information seeking and mediated searching. Part 5. User–intermediary interaction. *Journal of the American Society for Information Science and Technology*, 53(11), 883–893. <http://doi.org/10.1002/asi.10133>
9. Evans, B. M., & Chi, E. H. (2010). An elaborated model of social search. *Information Processing & Management*, 46(6), 656–678. <http://doi.org/10.1016/j.ipm.2009.10.012>
10. Evans, B. M., Kairam, S., & Pirolli, P. (2009). Do your friends make you smarter?: An analysis of social strategies in online information seeking. *Information Processing & Management*, 46(6), 679–692. <http://doi.org/10.1016/j.ipm.2009.12.001>
11. Foster, J. (2009). Understanding interaction in information seeking and use as a discourse: a dialogic approach. *Journal of Documentation*, 65(1), 83–105. <http://doi.org/10.1108/00220410910926130>
12. Hansen, P., & Järvelin, K. (2005). Collaborative Information Retrieval in an information-intensive domain. *Information Processing & Management*, 41(5), 1101–1119. <http://doi.org/10.1016/j.ipm.2004.04.016>
13. Hansen, P., Shah, C., & Klas, C.-P. (Eds.). (Forthcoming). *Collaborative Information Seeking: Best practices, New Domains, New Thoughts*. Springer. Retrieved from <http://collab.infoseeking.org/cisbook2015/>
14. Hearst, M. A. (2009a). Chapter 3: Models of the Information Seeking Process. In *Search User Interfaces* (1st ed.). Cambridge University Press.
15. Hearst, M. A. (2009b). *Search User Interfaces* (1st ed.). Cambridge University Press.
16. Hertzum, M. (2008). Collaborative information seeking: The combined activity of information seeking and collaborative grounding. *Information Processing & Management*, 44(2), 957–962. <http://doi.org/10.1016/j.ipm.2007.03.007>
17. Karunakaran, A., Spence, P. R., & Reddy, M. C. (2010). Towards a Model of Collaborative Information Behavior. In *Workshop on Collaborative Information Retrieval*. Savannah, GA. Retrieved from <http://www.google.co.uk/url?sa=t&rct=j&q=+Towards+a+Model+of+Collaborative+Information+Behavior&source=web&cd=1&ved=0CCgQFjAA&url=http%3A%2F%2Fworkshops.fxpal.com%2Fcscw2010cis%2Fsubmissions%2Ftmp10.pdf&ei=t8s7T-bWC4rH0QXizdRs&usg=AFQjCNFHjJUptg9R3LMdZTMmY6goknb85A&sig2=C9oJgmkmNdL-9ussTxVcyQ&cad=rja>
18. Knight, S. A., & Spink, A. (2004). Toward a Web Search Information Behavior Model. In A. Spink & B. J. Jansen (Eds.), *Web search: public searching on the Web*. Springer.
19. Knight, S. A., & Spink, A. (2004). Toward a Web Search Information Behavior Model. In A. Spink & B. J. Jansen (Eds.), *Web search: public searching on the Web*. Springer.
20. Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5), 361–371. [http://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5<361::AID-ASI6>3.0.CO;2-#](http://doi.org/10.1002/(SICI)1097-4571(199106)42:5<361::AID-ASI6>3.0.CO;2-#)

21. Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4), 41. <http://doi.org/10.1145/1121949.1121979>
22. Marchionini, G., & White, R. (2007). Find What You Need, Understand What You Find. *International Journal of Human-Computer Interaction*, 23, 205–237. <http://doi.org/10.1080/10447310701702352>
23. Morris, M. R. (2008). A survey of collaborative web search practices (p. 1657). ACM Press. <http://doi.org/10.1145/1357054.1357312>
24. Morris, M. R. (2013). Collaborative Search Revisited. Presented at the CSCW, San Antonio, Texas, USA. Retrieved from [http://research.microsoft.com/en-us/um/people/merrie/papers/morris\\_cscw2013\\_collab\\_search\\_revisited.pdf](http://research.microsoft.com/en-us/um/people/merrie/papers/morris_cscw2013_collab_search_revisited.pdf)
25. Newman, K., Knight, S., Elbeshausen, S., & Hansen, P. ((Accepted)). Situating CIS – The importance of Context in Collaborative Information Seeking. In P. Hansen, C. Shah, & C.-P. Klas (Eds.), *Collaborative Information Seeking: Best practices, New Domains, New Thoughts*. Springer. Retrieved from <http://collab.infoseeking.org/cisbook2015/>
26. Shah, C. (2012). *Information Retrieval: Collaborative Information Seeking: the Art and Science of Making the Whole Greater Than the Sum of All*. Springer.
27. Shah, C. (2013). Effects of awareness on coordination in collaborative information seeking. *Journal of the American Society for Information Science and Technology*, n/a–n/a. <http://doi.org/10.1002/asi.22819>
28. Shah, C., & González-Ibáñez, R. (2010). Exploring information seeking processes in collaborative search tasks. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1–7. <http://doi.org/10.1002/meet.14504701211>
29. Shah, C., & Marchionini, G. (2010). Awareness in collaborative information seeking. *Journal of the American Society for Information Science and Technology*, 61(10), 1970–1986. <http://doi.org/10.1002/asi.21379>
30. Spence, P. R., & Reddy, M. (2012). Beyond practices: A field study of the contextual factors impacting collaborative information seeking. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–10.
31. Twidale, M. B., Nichols, D. M., & Paice, C. D. (1997). Browsing is a collaborative process. *Information Processing & Management*, 33(6), 761–783.

## **Group #5 – Exploring Search Trails as Learning**

Prepared by: Rob Capra, Simon Knight, Chris Leeder, Sandra Toze, Michael Twidale, and Soo Young Rieh

This group discussed the potential of studying search trails as a method of student learning of search skills and domain knowledge. Two main questions emerged: what are the potential uses of search trails, and what are benefits to others?

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. Search trails can be explored either through server log data or through visualization of the trail. One example of search trail visualization is the Trailblazer plugin (in beta at [www.trailblazer.io](http://www.trailblazer.io)) that displays a search trail as a network of nodes, representing both queries and pages visited as nodes connected by lines. These trails can be branching, iterative, recursive, and complex.

The group discussion identified possible benefits of exposing students to search trails. Viewing previous search trails could demonstrate to students that search is a complex process with skills that can be learned and improved. Exposure to multiple search trails on the same topic could help address the common misconception that there is one "right answer" to questions and demonstrate that different search trails produce different results. Instead of satisficing with whatever result they find first, or simply starting over when they do not find an answer, search trails could give students a sense of search as an iterated, continuing process. This could motivate students to learn about the search process itself rather than focusing solely on an answer. Students could gain an understanding of the fact that if you learn to search better, you can be more efficient, searching faster, and finding better results.

Another potential benefit is that students could learn why experts continue past an initial result, demonstrating how experts evaluate credibility and trustworthiness of sources. The implicit knowledge that experts possess about both search and knowledge domains could be made explicit by reviewing their trails.

Visualization of search trails could help clarify students' conceptual understanding of the process of search and encourage reflection on their search skills. This could help develop metacognition that could be applied in future searches. Such metacognition could also be used to understand a concept space, and assist in sensemaking.

Search trails could be used as an explicit resource for teaching, focusing what people learn from reviewing trails, both as learning to search and learning domain knowledge.

Some previous work in this area was discussed. Ryan White and his colleagues [2, 3] have conducted large-scale analysis of user search data and looked at the information content of “scenic routes” rather than simply jumping to an answer. Rob Capra and his colleagues [1] conducted a lab study and found that people made use of search trails in different ways for different task types. For complex tasks, participants described using the trails to find new information, and for less complex tasks they described using the trails to confirm information had already found. Dan Russell (Google Search Education) encourages users to share their trail in finding solutions to challenging questions.

Some suggestions for potential research studies in this area include:

- Show partially completed trails and ask students how they would get to a result
- Record search sessions and use think aloud to understand how and when subjects used a search trail
- Curate an annotated set of good and bad trails
- Build a set of alternative trails on one concept or topic

These potential studies can balance individual vs. social search. 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. This research would measure the outcomes by evaluating what students learned from exposure to the search trails, and how much their search and/or domain knowledge increased.

Some issues were identified with using search trails to investigate student learning. Learning is notoriously difficult to measure, and it is difficult to control for prior domain knowledge that will influence learning outcomes. Search tasks in these studies cannot be simply fact finding but involve topics that require searching beyond just answer finding. It may also be difficult to identify what are “good” search trails to use as examples, given different skill levels and domain knowledge.

These potential benefits and issues should be explored through further research. There has been limited work investigating the area of using search trails to investigate student learning, and there are great opportunities for new research.

*References*

R. Capra, J. Arguello, A. Crescenzi, E. Vardell. Differences in the Use of Search Assistance for Tasks of Varying Complexity. To appear, SIGIR 2015.

R. W. White, M. Bilenko, and S. Cucerzan. Studying the use of popular destinations to enhance web search interaction. In SIGIR, pages 159-166. ACM, 2007.

R. W. White and J. Huang. Assessing the scenic route: Measuring the value of search trails in web logs. In SIGIR, pages 587-594. ACM, 2010.

## **Group #6 – How to Evaluate SCIS Systems**

Prepared by: Roberto González-Ibáñez, Aiko Takazawa, and Yinglong Zhang

Group 6 started with a general discussion about different research problems involving social and collaborative information seeking (breakout session 1). This discussion was followed with a brainstorming on how to evaluate knowledge change as a result of collaborative information seeking (breakout session 2). Note that in the two breakout sessions different people participated in the discussion.

Below a list of the research questions that were formulated during the first breakout session and a study design to approach a particular question involving CIS and learning.

### **Research Problems**

1. How do you tease out the influence in collaborative encounters? How much influence do
2. Individual people, system components, have on encounters? Can we tease out the effect of that influence?
3. What elements of collaboration are ephemeral and can't be captured by systems?
4. Can we design for spontaneous collaboration systems for serendipitous encounters?
5. Does being able to tease out influence even matter in collaborative information seeking?
6. What cues do people use in assessing credibility or relevance (of the system? of other people?) in CIS?
7. What are the contexts or scenarios where we can create tools for support of CIS? With systems what are contexts where we CANNOT support CIS? (Are explicit tools always possible or even desirable?)
8. How can lightweight, flexible tools for collaborative info behavior behaviors? Does everything have to be a heavyweight "system"?
9. What role does distance (geographical, cultural, affective, experiential, etc.) play in collab info behavior?
10. As researchers, what practical issues in CIB can we tackle in the next 13 years?
11. What affordances should be built into CIB systems?

### **Study Design (CIS and learning)**

**Objective:** study relation between CIS and learning - in particular knowledge change.

**Research questions:**

1. To what extent learning occurs as a result of information seeking?
2. To what extent, if any, collaboration in information seeking contributes to learning?

**Methodological approach:** Mixed Method approach

**Study type:** Lab study (or bringing the lab to the classroom)

**Specific research methods:**

- 1) Interviews (Individual and Group)
- 2) Observational Research
- 3) Surveys/Standard Instruments
- 4) Log data

**Study Design:** Pretest-Posttest design

**Test instrumentation:** It could be an exam

**Conditions:** Individuals (as baseline) and Groups

**Variables to control:** Space, Time, Common ground, Prior knowledge, Education level (institutional), Group size, Resources/Tools, Task difficulty/complexity

**Task:** Research on a topic (e.g.: first order equations), define it, explain it, and then apply it.

**Target population:** Students

**Experimental design:**

- 1) Individual Pretest
- 2) Perform Search Task
- 3) Individual Posttest
- 4) Presentation (define-explain-apply)
- 5) Follow-up Interview