NSF Workshop on Task-Based Information Search Systems

NSF Workshop on Task-Based Search
University of North Carolina at Chapel Hill
School of Information and Library Science
March 14-15, 2013

Diane Kelly (dianek@email.unc.edu), PI
Jaime Arguello, Co-PI
Robert Capra, Co-PI

Anita Crescenzi, Project Assistant

NSF Workshop Report on Task-Based Information Search Systems

Grant No.: IIS-1301958
Award Dates: January 1 – December 31, 2013
Award Amount: $49,996.00
NSF Award Abstract: http://www.nsf.gov/awardsearch/showAward?AWD_ID=1301958
Project Website: http://ils.unc.edu/taskbasedsearch

Participants

Eugene Agichtein, Emory University
Jae-wook Ahn, Drexel University
Jaime Arguello (organizer), University of North Carolina
Nick Belkin, Rutgers University
Pia Borlund, Royal School of Library and Information Science, Denmark
Katriina Byström, University of Boras, Sweden
Rob Capra (organizer), University of North Carolina
Ben Carterette, University of Delaware
Arjen de Vries, Delft University of Technology, The Netherlands
Fernando Diaz, Microsoft Research
Abdigani Diriye, Carnegie Mellon University
Susan Dumais, Microsoft Research
Luanne Freund, University of British Columbia, Canada
Gene Golovchinsky, FxPAL
Jaap Kamps, University of Amsterdam, The Netherlands
Diane Kelly (organizer), University of North Carolina
Bill Kules, Catholic University of America
Birger Larsen, Royal School of Library and Information Science, Denmark
Christina Lioma, University of Copenhagen, Denmark
Jingjing Liu, University of South Carolina
Gary Marchionini, University of North Carolina
Catherine Smith, Kent State University
Mark Smucker, University of Waterloo, Canada
Simone Stumpf, City University London, UK
Jaime Teevan, Microsoft Research
Elaine Toms, Sheffield University, UK
Pertti Vakkari, University of Tampere, Finland
Ryen White, Microsoft Research
Barbara Wildemuth, University of North Carolina

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1301958.

Any options, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

The workshop organizers greatly appreciate:
Maria Zemankova, our NSF Program Manager, for continued support and encouragement
Anita Crescenzi, our graduate assistant, for helping to organize and manage the workshop
Kay Lackey and the UNC SILS office staff for helping with logistics and financial matters
Wan-Ching Wu and Kathy Brennan, for helping with workshop logistics
The UNC School of Information and Library Science, for hosting the workshop

Table of Contents

Executive Summary

When people seek information, they typically do so in order to resolve some underlying need or task, such as finding a bus schedule to plan travel, finding a recipe to make a dish for a potluck dinner, or finding the homepage of an author of a recently read book to see what other books she has published. While contemporary search engines are good at helping people resolve these types of look-up tasks, they are not as useful in helping people engaged in more complex tasks whose resolution might require multiple search sessions and multiple search strategies. Instead, search engines are optimized for particular types of tasks (e.g., look-up tasks and commerce tasks such as travel and shopping), for particular types of search behaviors (i.e., enter a query, review snippets, make a transaction) and for particular types of searchers (i.e., those who want to quickly find a single piece of information). Search engines are not optimized for tasks that require sustained interaction and engagement with information, the use of multiple, diverse search approaches to finding information or for searchers who want to cultivate a deeper, internalized understanding of a problem or topic. Contemporary search environments are tailored to support a small set of basic search tasks and provide searchers with few options to search and interact with information, and little to help them synthesize and integrate information across sessions.

This report defines research challenges related to the development of task-based information search systems that were elicited during a NSF-sponsored workshop held at the University of North Carolina at Chapel Hill in March 2013. This workshop gathered leading international researchers in information retrieval, human-computer interaction and information behavior to discuss research and challenges in incorporating models of tasks, task-types, and users' needs into systems/tools to support complex, multi-search and multi-session tasks. There are many challenges in creating such *task-based search systems* and the goal of this workshop was to enumerate, discuss, and document these issues into a research agenda that could help guide work in this field. Specifically, this workshop focused on the following topics:

1. Identification, elicitation, modeling and tracking of tasks, processes and states, including the identification of frameworks for conceptualizing task and relevance models;
2. Creation of task-specific and task-aware search environments, including the development of interfaces, tools, features, indexing techniques and search algorithms;
3. Development of methods and measures for studying user behavior and evaluating task-based search systems.

Major themes of the workshop included the development of domain-neutral modeling techniques to represent tasks, task properties and task-related search behaviors, interface support tools to assist with a variety of task-related information behaviors and the identification of techniques and tools to evaluate task-based search systems. The most critical need identified was the development of task models; this was viewed as essential for addressing the challenges of tools and evaluation measures.

This report provides a foundation for research on task-based search systems and also identifies many barriers, especially for academic researchers who do not have the tools to collect longitudinal, naturalistic data of people's online information seeking behaviors. Additional materials related to this workshop can be found at the workshop website: *http://ils.unc.edu/taskbasedsearch/*, including all the homework responses, the full text of the breakout group reports and a bibliography of research relevant to task-based search.

Workshop Dates:	March 14-15, 2013
Workshop Location:	Chapel Hill, NC, USA
Organized by:	Diane Kelly, Jaime Arguello, and Rob Capra

School of Information & Library Science
University of North Carolina
100 Manning Hall, CB#3360
Chapel Hill, NC 27599-3360

Need for Workshop and Related Efforts

A small group of researchers have been working for some time on task-based information seeking and retrieval. One of the first reviews of this approach was written in 2003 (Vakkari, 2003), and since that time a steady and growing stream of research has been published. While this research has generated several notable task models and documented how task type and task properties can impact search behavior, there has yet to be any concentrated efforts to further develop this area and combine the findings with system design and development. There has been little work developing different indexing, retrieval and/or ranking functions, or developing different interfaces and interaction techniques for different task types. Furthermore, there have been few efforts to generate evaluation methods and measures, and relevance models that are tailored to different tasks and consider multi-session search.

In 2006, Byström, Sundin and Limberg convened a group of researchers in Sweden to better understand task-based research, and in 2009, Anderson et al. organized a panel discussion at the annual conference of the *American Society for Information Science and Technology (ASIST)* exploring conceptual and methodological approaches to studying task in information science. One outcome of both workshops was a call for the development of more task models and the integration of such models into system design. While this meeting and panel generated some research in these areas, the audiences and outreach was small, so the ideas have not widely spread. This meeting and panel only involved people from one community, while the problem requires researchers and perspectives from several communities, including information retrieval, human-computer interaction and information behavior.

There have been several workshops focused on task, including Larsen, Lioma and de Vries's (2012), *Task-based and Aggregated Search Workshop* and the *Second Strategic Workshop on Information Retrieval in Lorne* (SWIRL) (Allan, Croft, Moffat & Sanderson, 2012). Task-based search was discussed within the context of several larger themes identified by SWIRL participants, and was also presented briefly as a mini-theme. A recent NII Shonan Meeting, which was held in Japan in October 2012, focused on whole-session evaluation of interactive information retrieval systems. While not specifically on the topic of task-based search, the ideas generated during this workshop are relevant to task-based evaluation, since the types of tasks with which we are concerned are those that take place across many sessions. A number of our invitees participated in one or many of these past events, which allowed us to build-on the findings from these previous meetings.

These recent meetings and discussions demonstrate that this was an opportune time to host a workshop focused on task-based search, and that this topic is recognized by many leading researchers in information retrieval as significant.

Structure of Workshop and Schedule

Prior to attending the workshop, participants were asked to submit short statements identifying significant challenges and important research papers, with annotations describing the significance of the papers. At the start of the workshop, three presentations by leading researchers were given on core area: task-based information seeking behavior presented by Pertti Vakkari; search engines and task presented by Susan Dumais; and interfaces and task presented by Gene Golovchinsky. Following these presentations, two additional shorter presentations were made Nick Belkin and Ben Carterette about recent efforts to address evaluation of search sessions. Belkin described a recent workshop at the Shonan Village in Japan about whole-session evaluation of interactive information retrieval, while Carterette reported on the National Institute of Standards and Technology's Text Retrieval Conference (TREC) Session Track.

During the second-half of Day 1 of the workshop, participants selected some of the most challenging issues related to task-based search and discussed and analyzed these issues in small groups of 4-6 people. On Day 2, these groups sub-divided into smaller groups to discuss more focused aspects of the broader issues discussed the prior day. Participants were asked to develop a research agenda, including the specification of a research study that might initiate the agenda. Group leaders were elected on each day and asked to take notes, share summaries of the discussion during common plenaries, and generate reports describing the discussion.

The detailed schedule of the workshop is presented below.

Thursday, March 14: Day One

Time	Agenda
8:00-8:30 a.m.	Breakfast on-site
8:30-9:00 a.m.	Introductions and Overview
9:00-9:40 a.m.	Core Area 1: Task-based Information Seeking (Pertti Vakkari)
9:40-10:20 a.m.	Core Area 2: Search Engines and Task (Sue Dumais)
10:20-10:50 a.m.	Break
10:50-11:30 a.m.	Core Area 3: Interfaces and Task (Gene Golovchinsky)
11:30-11:50 a.m.	Field Views: Session Workshop Report (Nick Belkin)
11:50-12:10 p.m.	Field Views: TREC Session Track (Ben Carterette)
12:10-1:10 p.m.	Lunch on-site
1:10-1:30 p.m.	Instructions for Break-out Groups
1:30-3:00 p.m.	Break-Out Session 1
3:00-3:30 p.m.	Break
3:30-4:45 p.m.	Break-Out Group Presentations
6:30-7:30 p.m.	*Reception at the WXYZ Lounge*
7:30-9:30 p.m.	*Walk to dinner at Elements*

Friday, March 15: Day Two

Time	Agenda
8:00-8:30 a.m.	Breakfast on-site
8:30-9:30 a.m.	Plenary Discussion: Refine and Prioritize Research Questions
9:30-10:30 a.m.	Break-Out Session 2
10:30-11:00 a.m.	Break
11:00-12:00 p.m.	Break-Out Session 2 continues
12:00-1:00 p.m.	Lunch on-site
1:10-2:30 p.m.	Reports from Break-out Sessions
2:30-3:30 p.m.	Plenary Discussion: Next Steps and Research Agenda
3:30 p.m.	Close

Challenges Identified by Participants

Prior to the workshop, attendees were asked to identify one or two outstanding research questions that need to be addressed in order for search systems to become more task-aware. In the following paragraphs, we highlight some of the key challenges identified. The responses from our attendees fell under four broad categories: (1) Task Modeling, (2) System Predictions, (3) Interactions and Auxiliary Tools, and (4) Evaluation.

Task Modeling

The most basic unanswered research question is: What is a task? While some tasks (such as buying a car) are clearly defined, other tasks (such as learning a new skill) are more difficult to identify because they evolve or are embedded within a larger goal hierarchy. As a first step, it seems necessary to more clearly define what a task is and to determine whether search systems should model tasks as flat structures, with clearly defined start and end points, or as hierarchical structures.

Tasks differ across a number of attributes or characteristics. Task characteristics can be a function of the task structure, the user, and/or the user's context. Example characteristics include the task complexity, salience, urgency, and difficulty. Several attendees proposed the need to develop a more comprehensive typology of tasks. Responses also highlighted the need to understand which task characteristics influence search behavior and which characteristics have little effect across users. Along the same lines, some responses called for further research on how different task types influence relevance, search strategy, search interactions, and search outcomes. A deeper understanding of these relationships would help determine which task characteristics have actual design implications for the search system (for a search engine's ranking algorithm and/or the interface).

As mentioned above, certain tasks have clearly defined start and end points and can be defined as a sequence of sub-tasks. Future research should also investigate how task stage influences search behavior. Do users employ different relevance criteria at different stages in the task? Do they desire different types of functionality from the system?

Finally, if we view tasks as being made-up of smaller sub-tasks or components, future work should also consider whether certain components generalize across tasks. To explain this point, an analogy was made to the "cut and paste" sub-task, which is widely used in desktop applications. Knowing which sub-tasks generalize across tasks could inform the development of auxiliary tools that could help users across a wide range of tasks.

System Predictions

Several responses focused on predictions a system could make about the task associated with a search. These include predicting task type and task stage. If task type and stage influence search behavior, then it seems possible to predict task type and/or stage from user interactions with the search engine.

Many tasks require multiple search sessions. In order to make predictions about the task type and/or stage, a system would need to keep a record of all the searches associated with a particular task. This would require the system to detect when a user is embarking in a new task and when a user has completed a task. Likewise, the system would need to maintain an inventory of a user's "open" tasks and match the current search session with an open task. All of these challenges would be exacerbated if we considered searches across multiple devices. For example, search sessions associated with the same task may look very different from a desktop computer vs. a mobile device.

Beyond predicting task type and stage, we could also imagine a system that tracks every user's search trail for each task and then retrieves search trails in response to a user's new task. The general idea would be to have the system return search trails that are relevant to the task.

In addition to citing different types of predictions a system could make, several responses focused on sources of evidence a system could use to inform its predictions. One response called for a tighter integration between desktop and search applications. Users often use desktop applications to complete work tasks. Desktop applications could potentially accumulate information about a user's current task, and, if the task results in information-seeking, the application could convey this information to the search engine. Another alternative would be to capitalize on the fact that users within the same information-use environment accomplish similar types of tasks. Thus, the search engine could potentially share evidence across users in same environment.

Finally, additional research is needed to determine what kinds of information about a user's task a search engine should try to predict and what kinds of information the search engine should elicit from the user directly. What are the appropriate mechanisms for eliciting task information and what are the appropriate times to elicit information?

Interactions and Auxiliary Tools

Ultimately, the goal of making predictions about a user's task is to customize the search experience. Several attendees proposed research in developing specialized interactions and auxiliary tools. Above, we mentioned the possibility that task type and/or task stage may affect how users judge the relevance or utility of search results. One avenue to explore would to develop and integrate ranking functions that focus on the document attributes that are most important for a particular task type or task stage. Another alternative would be to dynamically adjust the presentation of results to highlight the most important document attributes.

Users use a wide range of tools to accomplish their tasks. Future work should also consider developing auxiliary tools that help users integrate and make better use of the information found during their searches. As an example, for tasks that require comparing between different alternatives, the search engine could surface a spreadsheet application.

Evaluation

Evaluation is a critical component of IR. It is necessary for both tuning system parameters and for comparing between alternative solutions and interfaces. In terms of evaluation, the main research question is: How to do we evaluate systems in a more "realistic" way? Ideally, the evaluation should consider the search engine's ability to help the user accomplish a task from start to finish. Thus, we require evaluation methods and metrics that operate across multiple queries and search sessions.

Reponses from our participants made reference to two common methods for evaluating search systems: user studies and test-collection-based evaluation. Within the context of task-based search, user studies typically use simulated tasks in order to study search behavior and/or system performance for different task types. One of the main challenges, however, is that different studies use different simulated tasks. Future research should focus on developing simulated tasks that can be shared across research groups. The simulated tasks should have clearly defined characteristics and should be empirically validated. That is, they should be tested to ensure that they elicit similar behavior across similar users. A re-usable set of simulated tasks would have two important benefits: they would help make user studies reproducible and they would allow the simulated tasks to be validated across different populations.

Finally, a number of responses expressed the need to develop test-collection based evaluation methods for task-based search. Test-collection evaluation has many benefits: it allows us to measure small improvements in performance and to reproduce results. The main challenge is that it requires modeling a user's sustained interaction with the search engine. To this end, one avenue to explore is how to simulate users. This brings us back to one of our initial points. Further studies are needed to understand how different task characteristics affect search behavior. Results from these studies would allow us to conduct more realistic user simulations.

Core Area Presentations

To start the workshop, three speakers were invited to give overviews of three "core" areas related to task-based search. The purpose of these talks was to give attendees grounding for the workshop topic and help people understand task-based research efforts from different communities. The first core area presentation was given by Pertti Vakkari and was titled, "Task-Based Information Seeking." In this presentation, Vakkari reviewed prior research efforts to define work tasks and search tasks, described the difficulties involved in defining tasks, and outlined important dimensions of task complexity. He also discussed issues and challenges involved in evaluating task-based search. Next, Sue Dumais gave a presentation titled, "Task-Based Search: A Search Engine Perspective." Dumais motivated the importance of this work by noting that long search sessions are very common and that tasks often extend over devices and long periods of time. She showed examples of common tasks and discussed methods for automatically detecting tasks and sub-tasks. Dumais also outlined a number of ideas about how to support users' task-based searches, including query histories, richer snippets, integration of verticals, inline answers, customization, and support for richer sensemaking. Finally, Gene Golovchinsky presented on "Interactivity and Feedback." In his talk, Golovchinsky described how two types of feedback could be considered; from person to system, and from system to person. He outlined ideas about how to increase the use of relevance feedback, and discussed ways that systems can provide hints about potential actions to users such as showing which documents are new, which terms are effective, and ways to reformulate a query. Golovchinsky also presented examples of how interfaces can support users in interacting with the past (e.g., previous actions/results) or the future (e.g., reformulations, suggestions). He described how interface components could help users in query formulation through previews and other persuasive features that nudge people to take positive actions.

Breakout Group Discussions

During the workshop, breakout groups were formed to discuss sub-topics. Each breakout group prepared a summary report based on their discussions. Group reports are clustered according to the major topic of focus: (1) modeling tasks and behaviors; (2) creation of tools and support for task-based search; and (3) evaluation. In addition, one small group focused on identifying ways to move the general research agenda forward.

Modeling Tasks And Behaviors

The first topic that guided breakout groups was modeling tasks and behaviors. This topic focused on the identification, elicitation, modeling and tracking of tasks, processes and states, including the identification of frameworks for conceptualizing task and relevance models. Group members included Katriina Byström, Luanne Freund, Jingjing Liu, Gary Marchionini, Pertti Vakkari and Barbara Wildemuth. This group identified as their main focus the question of how to represent and model a broader conception of task-based searching that extends beyond discrete, transactional searches, with a specific view on the transition points from one kind of activity to another, and from one goal or task to another. The group noted that most current models of search tend to be low level and overly simplistic, and only offer evidence of transition probabilities within a single search session (e.g., between querying and viewing results, not between different tasks or systems). The group thus decided to focus on developing a framework that connects models of information seeking tasks and information search tasks through the transitions between these tasks, including the probabilities of transitions and transition triggers.

In guiding the development of their framework, the group considered existing models and frameworks and established three key steps to guide their work: (1) identification of model elements; (2) identification of model structure; and (3) identification of task-based factors that are likely to influence the model, such as task type, task stage, and prior knowledge and expertise of the searcher. They proposed that the main benefit of the model would be to provide support for searchers' moves and decisions (e.g., by recommending specific tools). The group went on to identify challenges associated with developing and empirically testing such a model, including complications related to domain-specific information behavior and limitations associated with data collection tools. The group divided on the following day and proposed two research studies. The first was a cross-cultural, ethnographic study of information seeking behavior situated in the context of team-based patient care in the medical domain. The second sub-group proposed a series of studies, also in the health care domain, which focused on individuals and the health acceptance model.

In addition to these groups, two other smaller groups, which were initially part of the larger tools group, proposed research agendas that addressed the first topic area of the workshop. The first group (Fernando Diaz, Catherine Smith, Simone Stumpf and Elaine Toms) focused on the identification of task primitives. The group motivated their proposed research by observing that task models are needed to support task-based search, but little is known about the fundamental characteristics and dependencies between task activities and searching, which they term *task-dependencies*, and the extent to which these dependencies vary across domain. To investigate task dependencies across domains, the group suggested a standardized framework for decomposing task structure, which they term *task-primitives,* which would facilitate the discovery and documentation of universal task dependencies. In order to arrive at these task-primitives, the group proposed to record, decompose and analyze *in situ* task activities. The research agenda proposed by another group (Jae-wook Ahn, Gene Golovchinsky and Birger Larsen) sought to map and understand which tools/components/widgets are most useful for which moves/activities/actions. As a starting point, the group proposed to identify general abstractions and patterns that underlie information seeking behavior which can be used to suggest tools and compare behaviors across systems.

Tools and Support

Three breakout groups focused on tools to support task-based search. The first breakout group included Fernando Diaz, Sue Dumais, Jaap Kamps, Cathy Smith, Simone Stumpf, Elaine Toms, and Arjen de Vries and focused on the topic of "tools to support workflow." They formed two sub-groups to address the topic – one that considered "bringing task into search," and another that discussed "bringing search into task." The group focused on support for complex tasks that require synthesizing data from multiple sources across multiple search sessions. They also focused on tasks that have a specific output goal, commonly in the form of an aggregated set of information (family tree, written report). Progress toward such outputs can be measured and there was consensus in the group that such tasks are not well-supported by current search systems, tools, or apps. In considering how to embed search into the work task, the group advocates for considering a broad view of information access methods and considering how such information access is embedded across all aspects of the work task. The group also considered how the work environment can impact search. Here, they suggest exploiting the structure and constraints of the task output to help support both the task-specific product needs and the information access process. Extracting structure and sub-tasks out of the work environment was identified as a major challenge. The sub-structure and dependencies among sub-tasks are important to account for, but users must also have flexibility to fluidly move among or skip components, and to backtrack. Collaborative work adds another level of complexity. The group felt that today's web/cloud-based computing is blurring the boundaries between search and work tasks and that the time is right to integrate search into work and daily life tasks.

The second breakout that considered "Tools and Support" included Eugene Agichtein, Jaime Arguello, Christina Lioma, and Ryen White. This breakout group considered how smartphones could recognize the task that a user was trying to accomplish and make recommendations about apps that could help. To do this, the smartphone would use both implicit behavioral information and other contextual signals (such as location and time of day). The group compared task-based app recommendation to mobile search, and considered several example tasks with multiple steps such as preparing dinner for guests, or planning a day of vacation activities. As with many of the tasks considered for task-based search, these tasks required multiple steps, synthesizing information across sources, and the use of different applications. The group outlined a system architecture with a state-space controller and predictor model that would use features from the user's context and recommendation training data based on explicit and implicit feedback. Several challenges unique to mobile app recommendation were identified. First, at what stage in the task-completion process should an app be recommended? Early in the process users might not recognize the relevance of the recommendation, but delaying too long could be too late to be useful. A second challenge concerned how the system could assist the user by transferring the task-state among the apps being used to complete the task. To evaluate the system, the group proposed research questions that have broader applicability to task-based search systems. First, at what stage are users most likely to accept a recommendation? And second, does knowledge of the sequence of steps influence likeliness to accept a recommendation? They hypothesized that users would be more likely to accept recommendations at the early stages of task completion, and in situations where the user is aware of the task complexity *a priori*.

The third "Tools and Support" breakout group included Abdigani Diriye, Rob Capra, and Jaime Teevan. This group focused on ways to produce and present session-level and task-based summaries of search results. They outlined ways that search results could be summarized by processing the result set and providing overviews, surfacing common themes and topics, and noting documents that are similar or different from previous searches. Moving up a level, they posited that session-level summaries could help searchers to gain a deeper sense of what content was encountered and help them understand similarities and differences across the information found in the session. At the task-level, they hypothesized that summarization could help users complete complex tasks more quickly, they could simplify and encourage task resumption, and they could help accelerate knowledge acquisition. To generate task-level summaries, they discussed how queries, documents viewed, and browser-level actions could be used as input signals. The group identified several research questions that are important to the development of effective summaries for result sets, sessions, and tasks: 1) What features and attributes make a good summary?, 2) How useful are summaries across different tasks?, and 3) What techniques can be used to construct effective summaries? To evaluate summaries, they propose that measures should focus on quality, utility and usefulness.

Evaluation

The final topic that guided the breakout group discussions was the development of methods and measures for studying user behavior and evaluating task-based search systems. Participants who investigated this topic were: Nick Belkin, Pia Borlund, Ben Carterette, Diane Kelly, Bill Kules and Mark Smucker. The group focused on two main issues: (1) the limitations of the traditional IR evaluation framework when applied to task-based searching and (2) the establishment of a framework to guide reporting practices to better facilitate cross-study comparisons and the sharing of research infrastructure such as search tasks and questionnaires.

The group began by discussing the standard construal of information retrieval as a tool to help people find documents, which has led to evaluation measures focused on the (topical) relevance of documents, returned in response to a specific query, at a specific point in time. Documents are judged as relevant or not relevant, and a user's interaction with a search system is reduced to being the rate and amount of relevant documents consumed. In task-based search situations, such a construal is inadequate because searching might take place over an extended period of time, the user might issue a number of queries and sometimes the same query might be entered more than once. Furthermore, as person learns through the IR interaction what constitutes relevant information is likely to change. It was proposed that the concept of relevance be replaced by usefulness to indicate that an information object might help a person resolve their underlying tasks even if it is not relevant in a strict sense and also to indicate that human judgments during information seeking are likely to be dynamic. In addition to evaluating documents, it was proposed that the information seeking and retrieval process at the stage-level could be the focus of evaluation, as well as the end result (as opposed to the lists of documents produced by the system along with way). It was also proposed that the scope of the evaluation might be considered since it can range from the entire session, which might span multiple episodes across a number of days, to the usefulness of a particular system feature meant to support a specific aspect of the interaction at a specific point in time.

The group also discussed difficulties associated with creating measures that could be used within and across studies of users and tasks. Given a specific type of task, it is highly desirable to have a recommended set of measures that have been validated and calibrated so that they can be used for cross-study comparisons. Measures are likely to vary according to task type, but as of yet, there is no mapping of measures to tasks, and more pressing, there are few measures that are suitable for complex information seeking tasks that take place across multiple sessions. Thus, new measures are needed at both the micro- and macro-levels.

One specific approach taken by this group was to focus on stage-based approaches to information search and consider evaluation measures at this level of analysis. Questions asked included whether there are common stages that users experience while working on tasks; specific intentions of users at different stages; and how users view and describe the usefulness of their experiences at different stages, as well as overall. One important thing that emerged from this discussion was that more needs to be discovered about how people experience and view task-based search. This suggests more exploratory types of approaches to research as starting points for new evaluation measures and approaches. The group suggested a naturalistic, longitudinal study of people using instrumented laptops that would not only log their interactions, but also periodically elicit information from them about their search tasks, goals and experiences. The group further proposed periodic meetings with these participants to review their search histories and interactions in order to better understand successful and unsuccessful moments, the natural history of their search tasks and the types of measures that might be useful to evaluate different types of tasks and different stages of information seeking.

The group further elaborated on the need for a meta-framework for task-based information search studies in order to increase cross-study comparison and interoperability. It was noted that the current variety of research and reporting practices make integration across studies difficult and prevents long-term historical analysis, including meta-analysis, of studies. It further limits replication and reuse of instruments, tasks and measures. The proposed meta-framework suggests the following aspects of each study be clearly reported: tasks (and methods of task construction), study design, measures, and methods of analysis. The meta-framework would also provide guidelines about reporting practices (e.g., measures of effect size should be reported) and a matrix of measures/tasks for best practice.

Conclusions

Users engage in information-seeking in order to accomplish a higher-level task. The grand vision behind the NSF-sponsored workshop on Task-based Information Search Systems is that search systems should be more task-aware. Search systems should be designed and evaluated based on their ability to assist users in accomplishing their higher-level task. Workshop attendees consisted of leading international researchers from information retrieval, human-computer interaction and information behavior. The consensus that emerged from the Workshop was that making search systems more task-aware requires work in different directions. From the user side, it requires further understanding how task characteristics and task stage influence search behavior. From the system side, it requires modeling and tracking a user's task over multiple queries, search sessions, and devices, and designing interactions that guide the user towards task completion. It requires developing evaluation methodologies that more directly measure a system's ability to help users complete the task at hand. Finally, to facilitate research from all perspectives, it requires techniques for modeling tasks, task properties and task-specific search behaviors.

The Workshop provided a stimulating environment for researchers from different backgrounds to share their views about the outstanding research questions in the area of task-based search. Our hope is that this summary inspires researchers and practitioners to work towards building systems that go beyond the query/result-set paradigm and into the task-aware paradigm.

Participants and Challenge Statements

Eugene Agichtein

Emory University, USA

Automatically identifying and naturally supporting long-running (multi-session or multi-day) search tasks. Aspects of the problem include:

- Building a taxonomy of complex search tasks, and important components of the task, e.g., a template for the kinds of things people find when planning a trip.
- Automatically detecting early on that a user is embarking on a (potentially) long search task (e.g., as in [1]).
- Identifying the type of a task by matching to the taxonomy in [1].
- Detecting whether the user has completed the task or may resume it later.
- Understanding the possible interfaces to help the searcher resume the task from the last state (e.g., by expanding on [2]).

References

1. Eugene Agichtein, Ryen W. White, Susan T. Dumais, and Paul N. Bennet. Search, interrupted: understanding and predicting search task continuation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 315–324, New York, NY, USA, 2012. ACM.
2. Debora Donato, Francesco Bonchi, Tom Chi, and Yoelle Maarek. Do you want to take notes?: identifying research missions in yahoo! search pad. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 321–330, New York, NY, USA, 2010. ACM.

Bio

Eugene Agichtein is an Associate Professor of Computer Science at Emory University, where he founded and leads the Emory Intelligent Information Access Laboratory (IR Lab). The active projects in IR Lab include mining searcher behavior and interactions data, modeling social content creation and sharing, and applications to medical informatics. Dr. Agichtein obtained a Ph.D. in Computer Science from Columbia University, and did a Postdoc at Microsoft Research. He has published extensively on web search, information retrieval, and web and data mining. Dr. Agichtein's work has been supported by DARPA, NIH, NSF, Yahoo!, Microsoft, and Google, and has been recently recognized with the A.P. Sloan Research Fellowship and the 'Best Paper' award at the SIGIR 2011 conference.

Jae-wook Ahn

Drexel University, USA

What are the limitations of visual user interfaces for task-based search and how can we overcome them? [1,2] shows when a transparent user model (or task model) can fail. Unlike [1,2] which implement an offline search system or a text-based transparent user model, [3] presents a 2-D visualization based approach, which can overcome some of the limitations of the past approaches.

What are the properties that should be considered when evaluating task-based search system user interfaces that emphasize transparency? [4] suggests a list of aims for explanatory recommender systems, which could be helpful for defining the aims of task-based search system user interfaces.

References

1. Annika Waern. User involvement in automatic filtering: An experimental study. *User Modeling and User-Adapted Interaction*, 14(2-3):201–237, 2004.
2. Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. Open user profiles for adaptive news systems: help or harm? In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 11–20, New York, NY, USA, 2007. ACM.
3. Jae-wook Ahn and Peter Brusilovsky. Adaptive Visualization for Exploratory Information Retrieval. To appear in *Information Processing and Management*.
4. Nava Tintarev and Judith Masthoff. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22:399–439, 2012.

Bio

Jae-wook Ahn is working as a data scientist at Drexel University since September 2012. He worked as a post-doctoral CRA-NSF Computing Innovation Fellow at the HCIL (Human-Computer Interaction Lab), University of Maryland at College Park. He got his Ph.D. in Information Science from University of Pittsburgh in 2010 and was a member of PAWS (Personalized Adaptive Web Systems) and TALER (Teaching and Learning Research) Labs. He has worked for various projects on exploratory user interfaces for personalized information retrieval, visualization, and recommendation systems. He also worked on dynamic network visualization of participatory social media for his postdoc project. He is currently working for the Digging into Data and Meaningful Concept Display projects, on analyzing metadata from heterogeneous digital libraries and visualizing concepts in museum search environments.

Jaime Arguello (organizer)
University of North Carolina, USA

Bio

Jaime Arguello is an Assistant Professor at the School of Information and Library Science at the University of North Carolina at Chapel Hill. His research interests are in information retrieval, with a focus on aggregated search—the task of automatically integrating multiple search engines in a single interface. His research has been published in conferences including SIGIR, CIKM, WWW, ECIR, IIiX, HLT-NAACL, CHI, ICWSM, TREC, and DG.O. Jaime is a recipient of the SIGIR 2009 Best Paper Award and the 2011 ECIR Best Student Paper Award. He teaches courses in Information Retrieval and Text Data Mining. He received his Ph.D. and M.S. in Computer Science from Carnegie Mellon University and his B.S. in Electrical Engineering from Washington University in St. Louis.

Nicholas Belkin

Rutgers University, USA

I think that the most fundamental problem in this respect is the ability to infer motivating task type from the searcher's past and current information-seeking behaviors. This implies having a typology of motivating search tasks to start with, which in and of itself is a significant research problem. I find it difficult to separate these two research problems, so consider them in this context as one.

References

1. Yuelin Li and Nicholas J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, 44(6):1822–1837, 2008. This paper proposes a scheme for classification of both motivating search task types, and information searching tasks. The major contribution here is a means for classification that is not just naming different tasks, but rather a principled scheme for characterizing different task types. This means that in experimental situations, task type can be manipulated according to different values of some facets of task.
2. Chang Liu, Michael Cole, Eun Baik, and Nicholas J. Belkin. Rutgers at the TREC 2012 Session Track. In *Proceedings of 21st Text Retrieval Conference*, TREC '12. 2012. Although this paper does not investigate the issue of predicting motivating task from information-seeking behaviors, it makes a step in this direction by considering how different motivating tasks influence interpretation of information-seeking behaviors for the purpose of identifying "useful" documents during a search session.

Bio

Nick Belkin is Professor II of Information Science at the School of Communication & Information at Rutgers University. His recent research has focused on personalization of interaction with information, with specific reference to the influence of the task which motivates information seeking on interpretation of searcher behaviors with respect to predicting document usefulness.

Pia Borlund

Royal School of Library and Information Science, Denmark

The research problem that I would like to address is in line with the fourth mentioned example: "The need to develop IR evaluation methods that operate across multiple queries and even multiple search sessions". To me the objective is to be able to evaluate the IR interaction of the user as realistically as possible, that is, to handle multiple queries and even multiple search sessions – or in other words, to understand and evaluate IIR as it takes place in real life, including multi-facetted information needs and multi-tasking/task-switching. E.g., see the papers by Belkin (2008; 2010) and Spink (2004).

Also I would like to bring attention to the need for focus on research on searching of work tasks. That is, information searching as part of work task solving, as briefly addressed in the paper by Borlund, Dreier & Byström (2012).

References

1. Nicholas J. Belkin. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42(1):47–54, 2008.
   This paper explicitly points out a number of issues we ought to address, not the least with reference to evaluation of IIR systems.
2. Nicholas J. Belkin. On the Evaluation of Interactive Information Retrieval Systems. In: Larsen, B., Schneider, J.W., & Åström (Eds.). *The Janus Faced Scholar: A Festschrift in Honour of Peter Ingwersen*. Copenhagen: Royal School of Library and Information Science. 13–21, 2010.
   This paper contributes with a new framework for IIR evaluation focusing on usefulness.
3. Pia Borlund, Sabine Dreier, and Katriina Byström. What does time spent on searching indicate? In *Proceedings of the 4th Information Interaction in Context Symposium*, IIIX '12, pages 184–193, New York, NY, USA, 2012. ACM.
   The reported information seeking work task study reminds us that information searching takes place also in information intensive work task performance settings. The impression we have, is that the majority of current IIR research centres on Internet searching and everyday-life information needs – including the two IIR studies reported in this paper. However, there remains a need for IIR research on information searching in relation to information intensive work task performance with respect to optimise information searching, the various platforms used for information searching, and understanding of the conditions under which work task performance takes place.
4. Amanda Spink. Multitasking information behavior and information task switching: An exploratory study. *Journal of Documentation*, 60(4), 336–351. 2004. Emerald.
   This paper is an early example of the addressing of multi-tasking and task switching, hence not including the seamless IT and information environment of today, which also have to be taken into account.

Bio

Pia Borlund is Professor at the Royal School of Library and Information Science, Denmark, and docent at the School of Information Sciences, University of Tampere, Finland. Her research focuses on methods for evaluation of systems that support interactive information retrieval

(IIR). Her interest in IIR systems evaluation, design and usage brings together three broad areas: interactive information retrieval, human-computer interaction, and information seeking (behaviour). Pia Borlund has conducted research on frameworks and guidelines for performance evaluation of interactive information retrieval systems centered on the context instrument 'simulated work task situation' by involvement of users. Her current research focuses on methodological issues, test design and recommendations for evaluation of user search interaction.

Katriina Byström
University of Boras, Sweden

I think the following issue appears as fruitful to be addressed in order to design search systems that are more task-aware: Contextualizing task properties and search behavior, and the relationship them between into relevant information practices/behaviour.

Taylor's (1991) article discusses how different professional groups are formed around information use environments that in themselves include traits for what information is valued and consequently sought for as well as through what channels and sources this information is searched/distributed. Byström & Lloyd (2012) pushes the idea further by suggesting that each information use environment creates pervasive information practices with time sensitive professional and local influences. Work tasks fit into these environments as concrete instances where explicit and tacit knowledge culminates, which is why they provide useful base to study information search behavior and understand the role of IR systems. For the field of task-based information search this may provide a possibility to explain search behavior and design/evaluate IR systems not only from a user-oriented perspective, but also acknowledging the sociocultural aspects of search.

References

1. Katriina Byström and Annemaree Lloyd. Practice theory and work task performance: How are they related and how can they contribute to a study of information practices. *Proceedings of the American Society for Information Science and Technology*, 49(1):1–5, 2012.
2. Robert S. Taylor. Information use environments. In B. Dervin (Ed.) *Progress in Communication Sciences*, 10: 217–225,1991

Bio

Katriina Byström of the Swedish School of Library and Information Science, University of Borås does research on task-based information seeking and retrieval in workplaces and on information architecture. She is an active member of the academic Library and Information Science community and the co-founder and associate editor of the international Journal of Information Architecture. Her latest research projects are: Better Search Engine focusing on work task based search support and Better Web, the development of digital information and communication milieus on the web. Professor Byström chairs the European Network for Workplace Information (ENWI).

Rob Capra (organizer)
University of North Carolina, USA

Bio

Robert Capra is an Assistant Professor in the School of Information and Library Science at the University of North Carolina at Chapel Hill. His research interests include human-computer interaction, personal information management, and digital information seeking behaviors, tools, and interfaces. He holds a Ph.D. in Computer Science from Virginia Tech and M.S. and B.S. degrees in Computer Science from Washington University in St. Louis. At Virginia Tech, he was part of the Center for Human-Computer Interaction where he investigated multi-platform interfaces, information re-finding, and interfaces for digital libraries. Prior to Virginia Tech, he worked in corporate research and development, spending five years in the Speech and Language Technologies group at SBC Communications (now merged with AT&T Labs) where he focused on voice user interfaces, speech recognition, and natural language processing.

Ben Carterette
University of Delaware, USA

Whole-session evaluation: being able to evaluate the utility of a search system over the course of a user's interaction with it, ideally from task commencement to task completion. I'm envisioning a "task-aware" system as being one that attempts to determine a user's task from their interactions and adapt accordingly; if nothing else, it seems like some kind of sessiony evaluation would be necessary for use in objective functions. For example, Liu et al. [1] use task type prediction to select a feedback model during the course of a section.

While there are probably many ways to do whole-session evaluation (user studies, log analysis, etc), I am particularly interested in batch-style evaluations with reusable test collections. Batch evaluations allow researchers and developers to quickly perform tests of many possible combinations of features, models, and inputs while maintaining high statistical power. Reusability allows them to go back to any point in that search space and reliably get the same performance.

Creating test collections for whole-session evaluation is a difficult problem. We have been attempting to tackle it through the TREC Session track for the last three years [2, 3], and while we are happy with what we have accomplished, we still have a long way to go. The main problem is that it is difficult to model the fact that user interactions at time t+1 can depend on what the system does at time t; if the same test collection is going to be used to evaluate n different systems, it has to be able to model up to n different possible user actions at each time step. A direction we are considering is to use user simulation; while it is not likely that we will be able to accurately simulate users, we may be able to produce interactions that are at least useful for improving task-aware search systems.

The two Session track papers describe our efforts towards creating test collections for session evaluation. The second paper on the 2012 track is more specifically related to task-aware search, as our topics were categorized into four different broad task types. The Liu et al. paper describes the participation of Rutgers in the track. They built different feedback models for different task types and showed substantial improvements on some task types. This suggests that such a test collection can actually be useful for training task-aware systems.

References

1. Chang Liu, Michael Cole, Eun Baik, and Nicholas J. Belkin. Rutgers at the TREC 2012 Session Track. In *Proceedings of 21st Text Retrieval Conference*, TREC '12. 2012.
2. Evangelos Kanoulas, Ben Carterette, Mark Hall, Paul Clough, and Mark Sanderson. Overview of the 2011 Session Track. In *Proceedings of 20th Text Retrieval Conference*, TREC '11. 2011.
3. Evangelos Kanoulas, Ben Carterette, Mark Hall, Paul Clough, and Mark Sanderson. Overview of the 2012 Session Track. In *Proceedings of 21st Text Retrieval Conference*, TREC '12. 2012.

The notion of "relevance", which is so important to batch-style system-based evaluation, strikes me as limited in its ability to capture what users need from systems in order to actually complete tasks. If we instead talk about "utility"---as in the utility of a document to aid task completion---we can model utility not just by relevance but also by other important criteria such as timeliness, readability, truthfulness and trustfulness, completeness, novelty, obtainability, and more. Test collections in which documents are judged for utility given a specific task and context would allow researchers and developers to build and train systems that are more aware of tasks and user needs.

This idea is not new; it goes back to the late 60s and especially a number of papers by Cooper through the 70s (Stefano Mizzaro's review of the concept of relevance briefly describes much of this work [1]). But it hasn't been applied much, possibly because there are so many dimensions on which one can discuss "utility" that only looking at one or two at a time is even feasible. A few recent TREC tracks have done this: the Contextual Suggestion track, the Web track's diversity task.

Mark Rorvig argued that utility can be sufficiently modeled with preference judgments [2]: give an assessor two documents and a context, and ask which document they would prefer in that context. These preference judgments capture utility without needing to enumerate and judge against every possible aspect of utility. We have been applying this idea to building large collections of preferences that capture novelty and diversity along with relevance and other aspects of utility [3, 4].

References

1. Stefano Mizzaro. Relevance: the whole history. Journal of the American Society for Information Science, 48(9):810–832, 1997.
2. Mark E. Rorvig. The simple scalability of documents. *Journal of the American Society for Information Science*, 41(8):590–598, 1990.
3. Ben Carterette, Paul N. Bennett, David Maxwell Chickering, and Susan T. Dumais. Here or there: preference judgments for relevance. In *Proceedings of the IR research, 30th European Conference on Advances in Information Retrieval*, ECIR'08, pages 16–27, Berlin, Heidelberg, 2008. Springer-Verlag.
4. Praveen Chandar and Ben Carterette. Using preference judgments for novel document retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 861–870, New York, NY, USA, 2012. ACM.

Bio

Ben Carterette is an Assistant Professor of Computer & Information Sciences at the University of Delaware. His research interests include all aspects of information retrieval experimentation, from experimental design to test collection construction to evaluation measures to statistical analysis of experiments, and especially how the needs of users can be better modeled at each stage in a batch-style evaluation. In addition to publishing in venues such as ACM TOIS, SIGIR, CIKM, ECIR, and ICTIR, he has co-coordinated five TREC tracks and co-organized five workshops on IR evaluation.

Fernando Diaz

Microsoft Research, USA

What are appropriate auxiliary tools for different types of search tasks? Previously studied tools include query and URL history. However, it may be that finer-grained specialization of tools may be helpful. For example, when a user is researching a product, supplying a simple spreadsheet for price or review information may be useful; when a user is planning a trip, decomposing an interface into trip subtasks (e.g. accommodation, plane tickets) may be useful.

Can we adaptively augment traditional search interfaces with these auxiliary tools?

References

1. Dan Morris, Meredith Ringel Morris, and Gina Venolia. Searchbar: a search-centric web history for task resumption and information refinding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1207–1216, New York, NY, USA, 2008. ACM.
2. Henry Allen Feild and James Allan. Task-aware search assistant. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 1015–1015, New York, NY, USA, 2012. ACM.
3. Debora Donato, Francesco Bonchi, Tom Chi, and Yoelle Maarek. Do you want to take notes?: identifying research missions in yahoo! search pad. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 321–330, New York, NY, USA, 2010. ACM.

Bio

Fernando Diaz is a researcher at Microsoft Research New York. His primary research interest is formal information retrieval models. Fernando's research experience includes distributed information retrieval approaches to web search, interactive and faceted retrieval, mining of temporal patterns from news and query logs, cross-lingual information retrieval, graph-based retrieval methods, and synthesizing information from multiple corpora. Fernando received his PhD from the University of Massachusetts Amherst in 2008. His work on federation won the best paper awards at the SIGIR 2010 and WSDM 2010 conferences. He is a co-organizer of the Temporal Summarization track and Web track at TREC 2013.

Abdigani Diriye

Carnegie Mellon University, USA

One of the challenges stifling work on task-aware systems is identifying and mapping out the kind of search support and features needed to help users during different search tasks. The challenge here is identifying the inherent search activities the user might be engaged in, and the set of features and functionality that would best support them.

References

1. Gene Golovchinsky, Abdigani Diriye, and Tony Dunnigan. The future is in the past: designing for exploratory search. In *Proceedings of the 4th Information Interaction in Context Symposium*, IIIX '12, pages 52–61, New York, NY, USA, 2012. ACM. The above paper provides a good introduction on how to design for more complex and exploratory search tasks and some of the factors that need to be kept in mind.

Bio

Abdigani Diriye is a postdoctoral researcher at Carnegie Mellon University's Human-Computer Interaction Institute. He received his PhD from University College London. His PhD studies examined the role search interfaces play during information-seeking, and how we can build more useful and usable search systems. Previously, he has worked in the area of collaborative, social and multimedia search whilst interning at FX Palo Alto Labs, Microsoft Research and the Knowledge Media Institute. Currently, his research focuses on ways we can leverage human- and machine-generated data to support people when searching and sensemaking on the Web.

Susan T. Dumais

Microsoft Research, USA

Identifying tasks using implicit interactions. This is especially important for tasks that extend across time and devices. The references below provide examples of techniques for identifying queries related to tasks, for predicting whether a task will be resumed, and looking at tasks over a longer time scale.

References

1. Alexander Kotov, Paul N. Bennett, Ryen W. White, Susan T. Dumais, and Jaime Teevan. Modeling and analysis of cross-session search tasks. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 5–14, New York, NY, USA, 2011. ACM.
2. Debora Donato, Francesco Bonchi, Tom Chi, and Yoelle Maarek. Do you want to take notes?: identifying research missions in yahoo! search pad. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 321–330, New York, NY, USA, 2010. ACM.

Thinking broadly about what support for search tasks looks like. The references below provide examples from simple "answers" seen in web search engines, to apps for specific tasks, to richer environments for exploratory search.

References

1. Lydia B. Chilton and Jaime Teevan. Addressing people's information needs directly in a web search result page. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 27–36, New York, NY, USA, 2011. ACM.
2. 50 ultimate travel apps ... so far

Bio

Susan Dumais is a Principal Researcher and manager of the Context, Learning and User Experience for Search (CLUES) Group at Microsoft Research. Prior to joining Microsoft Research, she was at Bellcore and Bell Labs, where she worked on Latent Semantic Indexing (a statistical method for concept-based retrieval), interfaces for combining search and navigation, and organizational impacts of new technology. Her current research focuses on user modeling and personalization, context and information retrieval, temporal dynamics of information, interactive retrieval, and novel evaluation methods. She has worked closely with several Microsoft groups (Bing, Windows Desktop Search, SharePoint Portal Server, and Office Online Help) on search-related innovations. Susan has published more than 200 articles in the fields of information science, human-computer interaction, and cognitive science, and holds several patents on novel retrieval algorithms and interfaces. Susan is also an adjunct professor in the Information School at the University of Washington. She is Past-Chair of ACM's Special Interest Group in Information Retrieval (SIGIR), and serves on several editorial boards, technical program committees, and government panels. She was elected to the CHI Academy in 2005, an ACM

Fellow in 2006, received the SIGIR Gerard Salton Award for Lifetime Achievement in 2009, and was elected to the National Academy of Engineering (NAE) in 2011.

Luanne Freund
University of British Columbia, Canada

We still do not know very much about how tasks influence search, or more specifically: what are the task-based requirements of IR systems? Much of the research on task-based IR has focused on behavioural analyses of searchers in different task contexts, which informs our understanding of task as a contextual variable that influences behaviour, but does not necessarily have design implications for search.

This is a multifaceted problem, as it involves the relationships between task characteristics, document characteristics and characteristics of retrieval systems. We have descriptive models of each of these components that can help up identify key characteristics, but we are lacking in theoretical and empirical models that identify the relationships between them that are most likely to influence search outcomes. The empirical studies that we do have are of limited value due to the lack of a standard nomenclature for tasks and the idiosyncratic operationalization of task characteristics in assigned search tasks.

References

1. Norbert Fuhr. Salton award lecture information retrieval as engineering science. *SIGIR Forum*, 46(2):19–28, 2012.
   My thinking about this problem has been influenced by the 2012 SIGIR Salton Award keynote delivered by Norbert Fuhr, in which he discusses the need for an engineering approach in IR that would allow us to predict the kinds of systems and features needed in response to particular domain and task scenarios. The paper points us towards to importance of developing theoretical models of task-based IR as well as conducting more carefully controlled and systematic empirical studies to test and further develop these models.
2. Robert Capra, Gary Marchionini, Jung Sun Oh, Fred Stutzman, and Yan Zhang. Effects of structure and interaction style on distinct search tasks. In *Proceedings of the 7th ACM/IEEE-CS joint Conference on Digital libraries*, JCDL '07, pages 442–451, New York, NY, USA, 2007. ACM.
   There is very little published research that predicts and tests for task-based effects of retrieval system features on retrieval outcomes rather than user behaviours. The Capra et al. (2007) study comes close, as it examines relationships between task types, interaction styles and information architecture.
3. Wildemuth, B.M., & Freund, L. (2009). Search tasks and their role in studies of search behaviors. Paper presented at HCIR 2009: Bridging Human-Computer Interaction and Information Retrieval, Washington, DC, October 23, 2009.
   This position paper identifies some of the issues with task characterization and operationalization in interactive IR studies.

Bio

Luanne Freund is an Assistant Professor at the iSchool at the University of British Columbia, Canada. She received her PhD from the University of Toronto in 2008. Luanne's research is focused on interactive information retrieval, human information behaviour, and the effects of task and document genre on search. Current projects include the Systematic Review of Imposed Search Tasks (http://ils.unc.edu/searchtasks), which investigates the use of assigned search tasks in experimental studies; E-Informing the Public (http://diigubc.ca/research/egovernment), which is focused on the design of task and genre enhanced search systems to support public access to e-government information; Next Generation Information Access - NGAIA (http://diigubc.ca/ngaia), which is focused on the problem of domain-specific information retrieval, and Access to News Media, which seeks to support information seeking in the online news domain. Her research is funded by the Canadian Social Sciences and Humanities Research Council and the Graphics, Animation and New Media Network of Centres of Excellence, Canada.

Gene Golovchinsky

FxPAL, USA

I think the biggest obstacle to the deployment of task-aware systems is lack of understanding when such systems may be useful. When it's clear that records of prior interaction can be used to inform subsequent system behavior, this information is already incorporated into systems. There are no significant technical difficulties to start down this road. The biggest challenge is one of perception: because Google doesn't do something, doesn't mean that that something isn't possible or desirable in other contexts.

Bio

Gene Golovchinsky is a Sr. Research Scientist at FX Palo Alto Laboratory, Inc. His research interests include the design of interfaces for interactive information seeking, collaborative search, HCIR, dynamic hypertext, the role of visualization in information retrieval, and pen-

Jaap Kamps

University of Amsterdam, The Netherlands

To build an information access tool that actively supports a searcher to articulate a whole search task, and to interactively explore the results of every stage of the process. There is a striking difference in how we ask a person for information, giving context and articulating what we want and why, and how we communicate with current search engines. Current search technology requires us to slice-and-dice our problem into several queries and sub-queries, and laboriously combine the answers post hoc to solve our tasks. Combining different sources requires opening multiple windows or tabs, and cutting-and-pasting information between them. Current search engines may have reached a local optimum for answering micro information needs with lighting speed. Supporting the overall task opens up new ways to significantly advance our information access tools, by develop tools that are adapted to our overall tasks rather than have searchers adapt their search tactics to the "things that work."

References

1. Ian Ruthven. Interactive information retrieval. ARIST, 42(1): 43-91 (2008). http://dx.doi.org/10.1002/aris.2008.1440420109
   Solid overview of how much we know about the interaction, also immediately highlighting how little we know about the mechanics of interaction during a process of performing a complex task.
2. Arjen P. de Vries, Wouter Alink, Roberto Cornacchia: Search by strategy. *ESAIR* 2010:27-28 http://doi.acm.org/10.1145/1871962.1871979
   Interesting new approach to formulate a complex query (or search strategy) for tasks of increasing complexity.

Can we make a retrieval system aware of the searcher's stage in the information seeking process, tailor the results to each stage, and guide the searcher through the overall process? A search session for a non-trivial search task consists of stages with different sub-goals (e.g., problem identification) and specific search tactics (e.g., reading introductory texts, familiarizing with terminology). Making a system aware of a searcher's information seeking stage has the potential to significantly improve the search experience. Searchers are stimulated to actively engage with the material, to get a grasp on the information need and articulate effective queries, to critically evaluate retrieved results, and to construct a comprehensive answer. This may be of particularly great help for those searchers having poor information or media literacy. This is of obvious importance in many situations: e.g., education, medical information, and search for topics "that matter." Some special domains, such as patent search and evidence based practices in medicine, have clearly prescribed a particular information seeking process in great detail. Here building a systems to support (and enforce) this process is of obvious value.

References

1. Marcia J. Bates: Where should the person stop and the information search interface start? Inf. Process. Manage. 26(5): 575-591 (1990) http://dx.doi.org/10.1016/0306-4573(90)90103-9
   There is a need for a new discussion on what role the system and user play, and how the interface supports the task progress as well as the information seeking process.

2. Forest Woody Horton: Understanding information literacy: a primer; an easy-to-read, non-technical overview explaining what information literacy means, designed for busy public policy-makers, business executives, civil society administrators and practicing, UNESCO, 2008 http://unesdoc.unesco.org/images/0015/001570/157020e.pdf
Information/Media literacy research has many relations and basically outlines what type of information seeking behavior should be promoted by the system.

Meta questions on how to foster collaboration between research groups in computer science and information science, and in academia and industry, so that we could work *together* on solving some of these challenges in the near future.

Bio

Jaap Kamps is an Associate Professor of information retrieval at the University of Amsterdam. His research interests span all facets of information storage and retrieval, a common element is the combination of textual information with additional structure, such as document structure, Web-link structure, and/or contextual information, such as metadata, anchors, tags, or clicks. He leads research projects on retrieval with structured information, and projects with cultural heritage institutions (museums, archives, libraries).

Diane Kelly (organizer)

University of North Carolina, USA

Bio

Diane Kelly is an Associate Professor at the School of Information and Library Science at the University of North Carolina at Chapel Hill. Her research interests are in interactive information search and retrieval, information search behavior and evaluation methods and metrics. Her research has been published in several conferences and journals including ACM SIGIR, ACM CHI, CIKM, IIiX, JCDL, Transactions on Information Systems, Information Processing and Management, JASIST, IEEE Computer and CACM. She teaches undergraduate and graduate courses on research design, interactive information retrieval and foundations of information science. She has served on the UNC Behavioral Institutional Review Board (IRB) since 2005. She received a Ph.D. in Information Science and a Graduate Certificate in Cognitive Science from Rutgers University and an undergraduate degree in Psychology from the University of Alabama.

Bill Kules

Catholic University of America, USA

Research Problem: Design of exploratory search tasks for search system evaluation
Evaluation is an essential part of developing search tools that are more task aware, particularly for exploratory search, which is a recognized challenge for information seeking systems and an area of active research and development.

For any user study, tasks must be carefully constructed to balance ecological validity with experimental control. For exploratory search, this is a particular challenge, because we are trying to induce search behaviors that are inherently open-ended. Individual searchers have to interpret the task, formulate their own queries and evaluate the results based on their understanding of the information need and their own knowledge and experience. At the same time, we wish to maintain some level of experimental control to permit comparisons between systems and longitudinally.

Borlund (2003) developed the concept of a simulated work task, which forms the basis for many user evaluations of search systems. Many studies have used the simulated work task as the basis for search tasks, but tasks are rarely comparable between or even within studies, limiting our ability to build up a corpus of results in a manner similar to the TREC studies. Recent work has started to formalize attributes of exploratory search tasks and provide suggestions for how to create and validate such tasks (Kules and Capra, 2012; Wildemuth and Freund, 2012). There are a number of open questions to be investigated. Three of them are:

1. What is an appropriate, parsimonious set of attributes to define exploratory search tasks?
2. How can we quantify (can we quantify) measures for these attributes?
3. Given that searchers individually interpret tasks and results, what comparisons does this allow us to make between systems and studies?

References

1. Pia Borlund. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(2), 2003.This paper developed the concept of a simulated work task. It has formed the basis for much user-focused systems evaluation.
2. Bill Kules and Robert Capra. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. Journal of the American Society for Information Science and Technology, 63(1):114–138, 2012.This paper extends the simulated work task concept to create exploratory search tasks. It incorporates 7 additional criteria and describes an iterative procedure to create and validate exploratory search scenarios for a semi-controlled laboratory study.
3. Barbara M. Wildemuth and Luanne Freund. Assigning search tasks designed to elicit exploratory search behaviors. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*, HCIR '12, pages 4:1–4:10, New York, NY, USA, 2012. ACM.This paper reviews and critiques 84 papers related to exploratory search task design and proposes a set of design recommendations.

Bio

Bill Kules is Associate Professor and Associate Dean for Research and Administration in the School of Library and Information Science (SLIS) at The Catholic University of America. His research interests include faceted search tools to support complex information seeking tasks like exploratory search, and online teaching/learning environments to support engagement and enjoyment. Dr. Kules was a co-organizer of the workshop on Human-Computer Interaction and Information Retrieval (HCIR), an annual meeting of researchers and practitioners whose work spans the fields of human-computer interaction and information retrieval. He was also an organizer of the 2005 Workshop on Exploratory Search Interfaces, a guest editor for the April 2006 CACM Special Section on Supporting Exploratory Search, and is a guest co-editor for a special issue (in press) on HCIR for the journal Information Processing & Management. Before joining SLIS Dr. Kules spent 20 years designing and implementing information systems for a variety of applications, including wireless telephony, customer service, banking, and a multimedia web sites. He earned his Ph.D. in Computer Science at the University of Maryland.

Birger Larsen

Royal School of Library and Information Science, Denmark

One way of progressing for systems to become more task aware is to facilitate research by considering if it is possible and fruitful to extend the Cranfield paradigm to support experiments with task based search. What are the demands on topics and relevance assessment to support task based experiments, and what additional procedures and performance measures are needed? Can the complexity be handled and what could be learned from such experiments?

References

1. Marianne Lykke, Birger Larsen, Haakon Lund, and Peter Ingwersen. Developing a test collection for the evaluation of integrated search. In *Proceedings of the 32nd European Conference on Advances in Information Retrieval*, ECIR'2010, pages 627–630, Berlin, Heidelberg, 2010. Springer-Verlag.This poster paper describes the iSearch test collection, where we put much more emphasis on obtaining through and structured descriptions of the work tasks and information needs. This may be one step towards task based search as it facilitates experiments with extended task descriptions.

Bio

Birger Larsen (born 1974) is Associate Professor at the Royal School of Library and Information Science in Copenhagen, Denmark since 2006. From 2010 he is leader of the 'Information Systems and Interaction Design' research group. He has a passion for research that involves the activities, processes and experiences arising in the meeting between users, information, and information systems in a given context - with the goal of optimising these to empower users in their task and problem solving. His main research interests include Information Retrieval (IR), structured documents in IR, XML IR and user interaction, domain specific search, understanding user intents and exploiting context in IR, as well as Informetrics/Bibliometrics, citation analysis and quantitative research evaluation.

Christina Lioma
University of Copenhagen, Denmark

One potentially interesting aspect of task-aware search is the ranking model that estimates the relevance of the retrieved results. Traditionally, ranking models are grounded on mathematical estimations, such as metric distance or probabilities, and often include empirically-tuned parameters. It is not uncommon to use the exact same ranking model in different search tasks. However, relevance should not necessarily always be treated uniformly across different tasks. Task-based ranking models could be considered, taking as a starting point advances in dynamic similarity measures, which are partly tuneable at query time manually by the user (Bustos and Skopal 2006), or which accommodate various different task-based similarity functions (Ciaccia and Patella 2009). These papers present the two examples of dynamic similarity measures mentioned above:

References

1. Benjamin Bustos and Tomáš Skopal. Dynamic similarity search in multi-metric spaces. In *Proceedings of the 8th ACM International workshop on Multimedia Information Retrieval*, MIR '06, pages 137–146, New York, NY, USA, 2006. ACM.
2. Paolo Ciaccia and Marco Patella. Principles of information filtering in metric spaces. In *Proceedings of the 2009 Second International Workshop on Similarity Search and Applications*, SISAP '09, pages 99–106, Washington, DC, USA, 2009. IEEE Computer Society.

Bio

Christina Lioma is an Assistant Professor and Freja research fellow at the Department of Computer Science, University of Copenhagen, Denmark. Her research focuses on the computational processing of language, mainly in the areas of information retrieval and computational linguistics.

Jingjing Liu

University of South Carolina, USA

For multi-session tasks, how can search systems perform better, at different stages, and for different task types (e.g., tasks with different structures, difficulty/complexity levels, life vs. scholarly tasks, actionable vs. informational tasks, etc.)?

Frequently seen in everyday life, multi-session tasks are usually complex and require multi-sessions to complete. While IR systems do a decent job with simple search tasks, there's much room for them to improve in multi-session tasks. How can systems be better designed to facilitate users' finding and re-finding of information in multi-session tasks? What system features will be supportive and preferred by users? Understanding multi-session task features, user behaviors, and system features are all important to address this question.

References

1. Jaime Arguello, Wan-Ching Wu, Diane Kelly, and Ashlee Edwards. Task complexity, vertical display and user interaction in aggregated search. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 435–444, New York, NY, USA, 2012. ACM.Arguello et al. (2012) addressed task features (complexity) and the system interface feature (vertical display) as well as their interaction with users in aggregated search. This could be very relevant to and beneficial in dealing with multi-session tasks.
2. Alexander Kotov, Paul N. Bennett, Ryen W. White, Susan T. Dumais, and Jaime Teevan. Modeling and analysis of cross-session search tasks. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 5–14, New York, NY, USA, 2011. ACM.Kotov et al. (2011) showed that it is possible to effectively model and analyze users' cross-session search behaviors. Two problems they dealt with were: 1) identifying related queries to a current one from previous sessions, and 2) given a multi-query task, predicting if the user will return to the task in the future. This research is helpful for search systems to determine task context and suggest queries for multi-session tasks.

What task features make a search difficult? And how can systems better support "difficult" tasks according to the reasons why they are difficult?

Byström, K. & Järvelin (1995) and Byström, K. (2002) explored the effect of task complexity (defined as the a prior determinability of information inputs, processing, and outputs) on people's information seeking and use in a work task environment. These studies found that with the increase of task complexity, increased the complexity of information needed, the needs for domain information and problem solving information, and the number of sources, but decreased the success. There is a strong link between information types acquired and sources used, and that task complexity has a direct relationship to source use.

Although it is not the same concept as task difficulty, according to Li & Belkin (2008), both represent the information seeker's perception that the information seeking is not easy. More qualitative studies like these are needed to understand what task features make IR system users feel "difficulty". These will help design systems can better support "difficult" tasks according to

the reasons why they are difficult.
References

1. Katriina Byström and Kalervo Järvelin. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2):191– 213, March 1995.
2. Katriina Byström. Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science*, 53(7):581–591, 2002.
3. Yuelin Li and Nicholas J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, 44(6):1822–1837, 2008.

Bio

Jingjing Liu is an Assistant Professor in the School of Library and Information Science at University of South Carolina. She received her Ph.D. in Information Science from Rutgers University in 2010. Her research focuses on the design and evaluation of information systems that support information seeking and use, and the accomplishment of work tasks and search tasks of various types. Her recent projects deal with personalization of information retrieval, examining people's search behaviors in various types of tasks and predicting document usefulness based on user behaviors, understanding and predicting search task difficulty, understanding users' knowledge change in the search process, and exploring the factors that affect people's information task performance. In her research, Jingjing has been gaining and using rich experience and expertise in task type control, design, and analysis of the effects of task types on various aspects of information retrieval. She has published in journals such as Information Processing and Management and Journal of Documentation, as well as conference proceedings such as ACM SIGIR and CIKM.

Gary Marchionini

University of North Carolina, USA

An overarching problem is two-fold: user context elicitation and use. By this I mean determining what and how information seekers learn over sessions and correspondingly how systems might assist this process.

A second, more specific problem is how to represent search history to users.

References

1. Robert Capra, Gary Marchionini, Javier Velasco-Martin, and Katrina Muller. Tools-at-hand and learning in multi-session, collaborative search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 951–960, New York, NY, USA, 2010. ACM.

Bio

Gary Marchionini is Dean and Cary C. Boshamer Professor in the School of Information and Library Science at the University of North Carolina. He teaches courses in human-information interaction, interface design and testing, and digital libraries. He founded the Interaction Design Laboratory at SILS. He has led projects (NSF funded) on user interfaces for video retrieval, statistical tables, and multi-session and collaborative search. He has more than 200 publications over his career.

Catherine Smith
Kent State University, USA

Research Problem 1: the need to study transitions between task-specific applications and search sub-tasks.

As Belkin (2009) stated, "… we might say that an ultimate goal of, and challenge for IR research is to arrange things such that a person never has to engage with a separate IR system at all (although I am quite willing to agree that there are certainly circumstances in which such engagement might be indeed desirable.)." In this view, the burden of acquiring useful task descriptions (useful to the retrieval system) might be handled by applications that support "parent-task" goals (with a parent-task defined as any task that invokes an information search sub-task). In order to exploit task-related data available from such an application, we need to study transitions between search sub-tasks and parent-tasks.

References

1. Nicholas J. Belkin. Really supporting information seeking: A position paper. Information Seeking Support Systems. Technical Report. 2009.

Following from above, as an example of transitions, one can imagine search sub-tasks interleaved with active reading, where reading is the parent-task. An application like the one described by Hinkley, Bi, Pahud, & Bixton (2012) might collect implicit and/or explicit task-related data, which it could pass to a search utility when search sub-tasks are invoked. We need to describe transitions, and investigate how transitions may be improved for the user. Toms, Villa, & McCay-Peet (2013) is an example of an experimental study along these lines. The authors state their objective as, "… to explore the boundaries of the work task and search process to examine how users integrate search with the larger task" (p. 16). The study used an active reading interface which was developed by the researchers, and was an integral component of a larger experimental retrieval system. Work on this problem would be further advanced by collaborations with HCI researchers designing task-specific applications. This is particularly important if we are to consider an architecture that enables coupling of task applications and a search utility.

References

1. Elaine G. Toms, Robert Villa, and Lori McCay-Peet. How is a search system used in work task completion? *Journal of Information Science*, 39(1):15–25, 2013.
2. Ken Hinckley, Xiaojun Bi, Michel Pahud, and Bill Buxton. Informal information gathering techniques for active reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1893–1896, New York, NY, USA, 2012. ACM.

Bio

Catherine L. Smith is an assistant professor at the Kent State University School of Library and Information Science. She studies how people use interactive search systems, using theories from cognitive psychology in exploring and explaining search behavior. Her current research activities focus on search expertise and how expertise is gained in formal instruction. In a second research area, she studies the effects of semantic priming in search interaction. Cathy's work is motivated by the idea that advanced systems should help people learn how to search when information needs are unfamiliar, uncommon, or complex.

Mark Smucker

University of Waterloo, Canada

In our work on time-biased gain (TBG), Charlie Clarke and I have written about how TBG has no presupposed notion of gain, or of user interfaces, or even of retrieval systems. What matters to time-biased gain is that we have some way of estimating the gain achieved by the user over time.

This workshop concerns itself with search tasks where the user wants to "cultivate a deeper understanding of a problem or topic" and where the task requires "sustained interaction and engagement with information". The notion that lengthy interactions with information are central to task-based search implies that not only will gain likely be spread out over a long time period, but that gain may not be simply accumulated on acquisition of relevant material. I wonder to what extent our notions of gain in search must change. Today we think of gain as finding relevant documents, but will that be the correct model of gain for task-based search?

Cooper (1973) discusses the notion that each document encountered in a search session should have some positive or negative utility. In Cooper's formulation of the problem, the retrieval system's job is to deliver documents and the user can report to us the utility of each document. If we see our IR systems as becoming more than tools for retrieval of documents, we may need new measures of gain. For example, if our IR systems became designed for supporting creative work, we might need a measure of gain similar to the creativity support index of Carrol and Latulipe (2009). Or, perhaps we need to start measuring and modeling negative utility along the lines of searcher frustration as done by Feild, Allan, and Jones (2010). Once we know how to measure gain for users, we will then be faced with the task of how to incorporate these notions of gain into our Cranfield-style evaluations of task-based search.

References

1. Erin A. Carroll and Celine Latulipe. The creativity support index. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 4009–4014, New York, NY, USA, 2009. ACM.
2. William S. Cooper. On selecting a measure of retrieval effectiveness. *Journal of the American Society for Information Science*, 24(2):87–100, 1973.
3. Henry Allen Feild and James Allan. Task-aware search assistant. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 1015–1015, New York, NY, USA, 2012. ACM.
4. Mark D. Smucker and Charles L.A. Clarke. Time-based calibration of effectiveness measures. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 95–104, New York, NY, USA, 2012. ACM.

Bio

Mark Smucker is an assistant professor in the Department of Management Sciences at the University of Waterloo, and is cross appointed with the David R. Cheriton School of Computer Science. Mark's research interests include the design, analysis, and evaluation of interactive information retrieval systems. Mark's recent work has focused on making information retrieval evaluation more predictive of actual human search performance. Mark has been a co-organizer of two TREC tracks, and he co-organized a SIGIR 2010 workshop on the evaluation of retrieval systems via the simulation of interaction. He is a recipient of the SIGIR Best Paper award (2012) as well a recipient of the University of Waterloo, Faculty of Engineering's Teaching Excellence Award (2012). Mark earned his computer science PhD from the University of Massachusetts Amherst in 2008. Prior to his PhD, Mark worked for seven years in industry on projects ranging from recommendation systems to the study of internet advertising behavior.

Simone Stumpf
City University London, UK

There are two areas for task-based search systems that I think are interesting to explore:

1. Understanding task-based search for non-text items. Not all search is for text-based items; users' search tasks also include images, music and videos. Research into searching for these items is limited and fragmented. Previously, there has been some work to understand how users search for images (Westman 2009), however there is a growing realization that more information is needed that take the context and background of the user into account to support them in their task-based search. More recently, there has been increasing interest in searching for and in videos (Smeaton 2007).

2. Providing better cues and "scent" in task-based search. Search engines results pages on the web have moved on from being just a collection of ranked items and they now provide subtle cues for the user to get to the information that they want via snippets, visual previews, etc. However, there are two issues surrounding this. Firstly, this functionality is usually not available to users on their personal storage systems and they may rely on cues of association (Chau et al. 2008). Secondly, there is a lack of understanding of the role these cues play in users' task-based search (Woodruff et al. 2001).

References

1. Allison Woodruff, Andrew Faulring, Ruth Rosenholtz, Julie Morrsion, and Peter Pirolli. Using thumbnails to search the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 198–205, New York, NY, USA, 2001. ACM.
2. Alan F. Smeaton. Techniques used and open challenges to the analysis, indexing and retrieval of digital video. *Information Systems*, 32(4):545– 559, 2007.
3. Stina Westman. Image Users' Needs and Searching Behaviour, pages 63–83. John Wiley & Sons, Ltd, 2009.
4. Allison Woodruff, Andrew Faulring, Ruth Rosenholtz, Julie Morrsion, and Peter Pirolli. Using thumbnails to search the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 198–205, New York, NY, USA, 2001. ACM.

Fundamentally - what is a "task"? There are so many different understandings of this term and it really matters, as any systems that are developed rest on a basic assumption of what is meant by "task".

References

1. Victor M. Gonzalez and Gloria Mark. Constant, constant, multi-tasking craziness: managing multiple working spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 113–120, New York, NY, USA, 2004. ACM.This paper is one take on what a task is but not necessarily for searching.

Bio

Simone Stumpf received a PhD in Computer Science in 2001 and a BSc in Computer Science with Cognitive Science in 1996, both from University College London. She joined City University London in 2009 and currently holds a Senior Lecturer position. Previously, she conducted research at Oregon State University and University College London. Her research centres on end-user interactions with intelligent systems and personal information management systems. She is a member of the End Users Shaping Effective Software (EUSES) consortium, an international collaboration to develop and investigate technologies that support end-users to directly influence software behaviour. Dr Stumpf also has professional experience as a User Experience Architect.

Jaime Teevan

Microsoft Research, USA

Defining task boundaries. While some tasks (like buying a car or planning a trip) are clearly defined, others (like planning summer activities or doing research) are much harder to identify because they evolve, change, are part of larger tasks, and consist of sub-tasks. It can be very hard for a person -- let alone a computer -- to clearly identify task boundaries, but clear task definition may be important for tools that want to support task-based search.

References

1. Rosie Jones and Kristina Lisa Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 699–708, New York, NY, USA, 2008. ACM.Jones and Klinkner [CIKM 2008] discuss the challenges with identifying tasks (which they call "search missions") versus sessions, and present one way to do so automatically.

Bio

Jaime Teevan is a Senior Researcher in the Context, Learning, and User Experience for Search (CLUES) Group at Microsoft Research, and an Affiliate Assistant Professor in the Information School at the University of Washington. She studies how people use digital information, particularly as related to their social and temporal context, and builds tools to help better support these information interactions. Jaime was named a Technology Review (TR35) 2009 Young Innovator for her research on personalized search. She co-authored the first book on collaborative Web search, and was Chair of the Web Search and Data Mining (WSDM) 2012 conference. Jaime also edited a book on Personal Information Management (PIM), edited a special issue of Communications of the ACM on the topic, and organized workshops on PIM and query log analysis. She has published numerous technical papers, including several best papers, and received a Ph.D. and S.M. from MIT and a B.S. in Computer Science from Yale University.

Elaine Toms

Sheffield University, UK

In this discipline, we seem trapped in the user-centred paradigm; not everything information-oriented is about the user and search behavior. A task may exist in isolation from the people who accomplish it. Tasks emerge out of an organizational environment and their resolution supports some organizational outcome. Tasks have clear objectives, but may have multiple outcomes, and multiple ways of reaching that outcome. One could conceive of the user as a convenient slave/robot/handmaiden to get the task done. The challenge is twofold:

1. Understanding the process: think Henry Ford and project the automobile assembly line a century later when the task has information components that have to be mixed, scrapped, stirred, and moulded although not in quite the same physical way. We do not know how similar typical "knowledge work" tasks are to, for example, the tasks that occur on an automobile assembly line, although there have been clues as demonstrated by the work on how people write papers and proposals.
2. Understanding which "sledge hammer, drill or screwdriver" the "slave" needs to get the job done; perhaps less like the automobile assembly line, much of knowledge work requires human intervention in the form or decision making that requires intense cognitive activity. What tools does the slave need to assist with the job?

In the context of knowledge work, what are those generic tasks that are shared by many contexts, that is, which ones are comparable to, for example, the "cut and paste" tasks of the desktop application work? which ones are context specific, for example comparable to the produce a slide show in a presentation software? Which ones require finding data and/or information? Which ones require using information? Which ones rely on the talents of the slave because the technology is still not sophisticated enough to do the task from beginning to end, and how do we assist the slave with more useful tools?

Does the approach used by Bartlett in decomposing a bioinformatics task fit with other types of "knowledge work" tasks? The Kulthau and Vakkari work on writing proposals and papers goes a long way toward decomposing task in an educational context (although they may not see it that way). What "cognitive protheses" do we need to develop to support task completion? An interesting and short note that defines this concept:
http://www.lpi.usra.edu/publications/reportsCB-1089/ford.pdf

Why have we never done a formal requirements analysis for any of our information solutions? Take even the digital library. Its design is based on past practices.

References

1. Joan C. Bartlett and Elaine G. Toms. Developing a protocol for bioinformatics analysis: An integrated information behavior and task analysis approach: Research articles. *Journal of the American Society for Information Science*, 56(5):469–482, 2005.

Bio

Elaine Toms is Professor of Information Science, and Head of the Information Retrieval Research Group, Information School, University of Sheffield, UK. Her research focuses on how people interact with information systems, including interacting with the content and with the tools to support use; and how to evaluate such systems in a user-centred way. Her current interest lies in the topic of this workshop: how to create more task-specific search tools and widgets to aid the process, and the integration of search into systems that support tasks.

Pertti Vakkari

University of Tampere, Finland

Understanding more in detail how larger tasks are related to search tasks and searching. By tasks I mean information intensive work tasks, which generate several search sessions. Empirical results hint that various aspects of search process like term selection, querying, relevance judgment and the information utilized vary between search sessions when task performance proceeds. In order to understand the role of various activities (stages) in the search process within and between sessions it is necessary to understand the whole search process and how it is associated with task performance. This is important 1) theoretically for understanding the phenomenon we are interested in, 2) for system design to better match the tools with human activities from the angle of both search tasks and work tasks, and 3) for creating evaluation procedures and metrics for task-based search.

References

1. Jingjing Liu and Nicholas J. Belkin. Searching vs. writing: Factors affecting information use task performance. *Proceedings of the American Society for Information Science and Technology*, 49(1):1–10, 2012.

In an experimental longitudinal setting Liu & Belkin studied the associations between newspaper article writing tasks and information searching and use in three points in time during the preparation of the article. The results in this and other articles from the same experiment are important because they have extended our understanding of how some features of tasks and task performance are related to various aspects of searching and utility assessments.

2. Pertti Vakkari and Saila Huuskonen. Search effort degrades search output but improves task outcome. *Journal of the American Society for Information Science and Technology*, 63(4):657–670, 2012.

In a field study Vakkari & Huuskonen examined how medical students' search effort for an assigned essay writing task was associated to precision and relative recall, and how this was associated to the quality of the essay. They found out that effort in the search process degraded precision, but improved task outcome. The poorer the precision, the better the quality of the essay. The findings concerning the whole process are important, because they suggest that traditional effectiveness measures in information retrieval are not sufficient for task-based searching. They should be complemented with evaluation measures for search process and task outcome.

Bio

Pertti Vakkari is professor of Information Studies at the School of Information Sciences, University of Tampere, Finland. He has been involved in collaborative European research projects. Vakkari has been a member in several program committees of conferences in information studies like ISIC, IIiX and ECIR. He is a member of the editorial board of journals "Journal of Documentation" and "Information Processing & Management". Vakkari's research interests include task-based information searching and use, the use of digital libraries, the evaluation of information search systems, fiction retrieval, and perceived outcomes of public libraries. His publications include several monographs and readers and over 100 articles. He has received ASIS&T SIGUSE Award for Outstanding Contributions to Information Behavior.

Arjen P. de Vries

Centrum Wiskunde & Informatica, Delft University of Technology, The Netherlands

How do we design and evaluate search systems (and their retrieval models) given that we know that relevance is not just topical?

References

1. William S. Cooper. Expected search length: A single measure of retrieval effectiveness based on the weak ordering action of retrieval systems. *American Documentation*, 19(1):30–41, 1968.The paper is useful because it takes an effort oriented view on the evaluation of systems (actually, developing also a measure that quantifies a system's success in terms of the reduction in effort that can be expected). I have been surprised about the fact that the measures proposed in this work are hardly ever used. (Last year's best SIGIR paper may bring it back into the picture, who knows.)
2. Kevyn Collins-Thompson, Paul N. Bennett, Ryen W. White, Sebastian de la Chica, and David Sontag. Personalizing web search results by reading level. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 403–412, New York, NY, USA, 2011. ACM. This paper considers readability in IR, demonstrating how it is a relevance criterion that matters.

How can we tailor the retrieval model to the task? What part can we automate in this tailoring process, and what part will remain the designer's task?

References

1. Stefano Ceri, Alessandro Bozzon, and Marco Brambilla. The anatomy of a multi-domain search infrastructure. In *Web Engineering, volume 6757 of Lecture Notes in Computer Science*, pages 1–12. Springer Berlin Heidelberg, 2011.The Search Computing project developed a powerful search environment, that allows to define searches that integrate multiple sources. The more traditional database approach that motivates this project does emphasize that we could perhaps benefit from more abstraction in defining how search will take place, instead of the usual focus on a specific instance.
2. Norbert Fuhr. Logical and conceptual models for the integration of information retrieval and database systems. In *East/West Database Workshop*, Klaenfurt, pages 206–218. Springer Verlag, 1994.Describes IR as generalization of the DB approach, and argues for a view that would into both the DB, IR and HCI aspects of retrieval in a unified manner.

Bio

Arjen P. de Vries is a tenured researcher at CWI leading the Information Access research group, and a full professor (0.2 fte) in the area of multimedia data management at the Technical University of Delft. De Vries studies the intersection of information retrieval and databases. In November 2009, De Vries co-founded Spinque, a CWI spin-off that provides integrated access to any type of data, customized for information specialist or end user, to produce effective and transparent search results.

Ryen White

Microsoft Research, USA

- Characterizing and supporting cross-session and/or cross-device search tasks, including "slow search" support that capitalizes on time between search episodes. Motivation: Complex tasks persist over time. People are using multiple devices more frequently. Need ways to support transitions between devices that capitalizes on the time that search engines may have – both in predicting whether a searcher will resume the task, deciding what action to take to help them (e.g., finding more/better results while the searcher is away from the search engine), and helping them restore their task state.

- Leveraging on-task behavior of the current user (personalization) and similar users (those in related cohorts). Motivation: On-task behavior is most relevant for personalization. Need ways to automatically identify search tasks and use this task-relevant information to adapt the search experience (results and UX) within the current session and beyond. Also potential benefit from using other searchers' on-task search behavior, especially for addressing the "cold start" problem associated with new users.

- Understanding and modeling the impact of task and user characteristics on information search behavior. Motivation: Attributes of the user (e.g., their domain knowledge), the search task (e.g., it's complexity), or their relationship (e.g., user familiarity with tasks of this type) affect search behavior. Better understanding these effects and developing user/task models that consider these effects can help design better systems and methodologies (including user simulations learned from sources such as logs) to evaluate these systems.

- Automatically identifying components of search tasks and guiding users through those stages. Motivation: Complex search tasks have multiple aspects. Automatically identifying those parts can help systems guide users through the stages in a useful sequence. Tours or trails could be shown to searchers as an alternative/complement to existing result lists. These tours can be manually created or determined algorithmically from sources such as search log data.

References

1. Eugene Agichtein, Ryen W. White, Susan T. Dumais, and Paul N. Bennet. Search, interrupted: understanding and predicting search task continuation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 315–324, New York, NY, USA, 2012. ACM.
2. Alexander Kotov, Paul N. Bennett, Ryen W. White, Susan T. Dumais, and Jaime Teevan. Modeling and analysis of cross-session search tasks. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 5–14, New York, NY, USA, 2011. ACM.
3. Yu Wang, Xiao Huang, and Ryen W. White. Characterizing and supporting cross-device search tasks. In *Proceedings of the sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 707–716, New York, NY, USA, 2013. ACM.
4. James Pitkow, Hinrich Schütze, Todd Cass, Rob Cooley, Don Turnbull, Andy Edmonds, Eytan Adar, and Thomas Breuel. Personalized search. *Communications of the ACM*,

45(9):50–55, September 2002.
5. Bin Tan, Xuehua Shen, and ChengXiang Zhai. Mining long-term search history to improve search accuracy. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, pages 718–723, New York, NY, USA, 2006. ACM.
6. Zhen Liao, Yang Song, Li-wei He, and Yalou Huang. Evaluating the effectiveness of search task trails. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 489–498, New York, NY, USA, 2012. ACM.
7. Ryen W. White, Wei Chu, Ahmed Hassan, Xiaodong He, Yang Song, and Hongning Wang. Enhancing personalized search by mining and modeling task behavior. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, New York, NY, USA, To Appear. ACM.
8. Diane Kelly and Colleen Cool. The effects of topic familiarity on information search behavior. In *Proceedings of the 2nd ACM/IEEE-CS joint Conference on Digital libraries*, JCDL '02, pages 74–75, New York, NY, USA, 2002. ACM.
9. Katriina Byström and Kalervo Järvelin. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2):191– 213, March 1995.
10. Yuelin Li and Nicholas J. Belkin. An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology*, 61(9):1771– 1789, 2010.
11. Ryen W. White, Susan T. Dumais, and Jaime Teevan. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, pages 132–141, New York, NY, USA, 2009. ACM.
12. Ryen W. White, Ian Ruthven, Joemon M. Jose, and C. J. Van Rijsbergen. Evaluating implicit feedback models using searcher simulations. *ACM Transactions of Information Systems*, 23(3):325–361, 2005.
13. Catherine Guinan and Alan F. Smeaton. Information retrieval from hypertext using dynamically planned guided tours. In *Proceedings of the ACM Conference on Hypertext*, ECHT '92, pages 122–130, New York, NY, USA, 1992. ACM.
14. Adish Singla, Ryen White, and Jeff Huang. Studying trail finding algorithms for enhanced web search. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 443–450, New York, NY, USA, 2010. ACM.
15. Ahmed Hassan and Ryen W. White. Task tours: helping users tackle complex search tasks. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 1885–1889, New York, NY, USA, 2012. ACM.
16. Randall H. Trigg. Guided tours and tabletops: tools for communicating in a hypertext environment. *ACM Transactions of Information Systems*, 6(4):398–414, 1988.
17. Alan Wexelblat and Pattie Maes. Footprints: history-rich tools for information foraging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 270–277, New York, NY, USA, 1999. ACM.

Bio

Ryen White is a Senior Researcher at Microsoft Research. His research interests lie in understanding search interaction and in developing tools to help people search more effectively.

He received his Ph.D. in Interactive Information Retrieval from the Department of Computing Science, University of Glasgow, United Kingdom, in 2004. Ryen has published over 150 conference papers and journal articles in Web search, log analysis, and user studies of search systems. He has received six best-paper awards, including two at the ACM SIGIR conference (2007, 2010), one at the ACM SIGCHI conference (2011), and one in JASIST (2010). His doctoral research received the British Computer Society's Distinguished Dissertation Award for the best Computer Science Ph.D. dissertation in the United Kingdom in 2004/2005. Ryen has co-organized numerous workshops on information seeking, in particular exploratory search, including an NSF-sponsored invitational workshop, and has guest co-edited special issues in these areas for a variety of outlets, including Communications of the ACM and IEEE Computer. Since 2008, he has co-organized the annual HCIR Symposium. Ryen has served as area chair for many top conferences such as SIGIR, WSDM, WWW, and CIKM, and currently serves on the editorial board of ACM TOIS, ACM TWEB, and the Information Retrieval Journal. In addition to academic impact, his research has been shipped in many Microsoft products, including Bing, Xbox, Internet Explorer, and Lync.

Barbara Wildemuth
University of North Carolina, USA

Many studies have used one or more attributes of search tasks as an independent variable and examined various search behaviors (e.g., search terms selected, search strategy formulation and re-formulation, or browsing behavior) as the dependent variable. Many of these have found some type of effect, but not all of them have. Which task attributes are most worthwhile to incorporate in future studies of this type? Are there any that consistently show no effect on search behaviors or outcomes?

References

1. Yuelin Li and Nicholas J. Belkin. An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology*, 61(9):1771– 1789, 2010.Li's taxonomy of task attributes contributed to this empirical study; it's useful because it demonstrates a strong (though not perfect, I think) conceptual foundation for an empirical study.
2. Elaine G. Toms, Luanne Freund, Richard Kopak, and Joan C. Bartlett. The effect of task domain on search. In *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative research*, CASCON '03. IBM Press, 2003. This is an older study, but it makes me think that we should try to pick some of the low-hanging fruit first. It may be relatively easy to detect the domain in which a person is searching; can we then tune the search engine to better support the searcher? Many interactive IR studies use (search) task complexity or difficulty as an independent variable. Yet these concepts are rarely defined clearly and have been operationalized in a variety of ways. So that the results of future studies can be compared with each other, we need to come to some agreement on the definitions of search task difficulty and search task complexity. In addition, in many studies, it's not clear whether the focus is on the search task or the work task, so we may also need to come to some agreement on definitions of work task difficulty and work task complexity.

Many interactive IR studies use (search) task complexity or difficulty as an independent variable. Yet these concepts are rarely defined clearly and have been operationalized in a variety of ways. So that the results of future studies can be compared with each other, we need to come to some agreement on the definitions of search task difficulty and search task complexity. In addition, in many studies, it's not clear whether the focus is on the search task or the work task, so we may also need to come to some agreement on definitions of work task difficulty and work task complexity.

References

1. Donald J. Campbell. Task complexity: A review and analysis. *The Academy of Management Review*, 13(1):40–52, 1988.This is a classic review, but outside our field. Campbell treats task complexity as 1) primarily a psychological experience, 2) an interaction between task and person characteristics, and 3) a function of objective task characteristics. Applied to our work today, this framework may provide a basis for our own definitions of task complexity.

2. Katriina Byström and Kalervo Järvelin. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2): 191– 213, March 1995.
3. Jingjing Liu, Chang Liu, Michael Cole, Nicholas J. Belkin, and Xiangmin Zhang. 2012. Exploring and predicting search task difficulty. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12. ACM, New York, NY, USA, 1313-1322.The first of this pair focuses on work tasks and the second focuses on search tasks. While we tend to focus on search tasks (as interactive IR researchers), it may be equally important (or more important) to attend to the complexity or difficulty of work tasks.

Other research topics

The role of simulated work task situations/scenarios: Do we all agree that use of simulated work tasks is "best practice" in developing search tasks for experimental studies? Across all types of tasks?

Bio

Barbara Wildemuth is Professor and Associate Dean in the School of Information and Library Science at the University of North Carolina at Chapel Hill. Her research focuses on people's use of information and information technologies. In particular, her studies have included investigations of medical students' searching of online databases, law students' use of a Web-based database of legal resources, and the effects of different interface designs on the effectiveness of database use. Her most recent work includes a methodological study of the search tasks assigned in interactive information retrieval experiments, being conducted in collaboration with Luanne Freund (UBC) and Elaine Toms (Sheffield). Her recent book, "Applications of Social Research Methods to Questions in Information and Library Science," has been adopted as a text in a number of ILS schools in the United States and abroad. Her teaching responsibilities include courses in various aspects of research methods, human information interactions, and information ethics.

Reports from Workshop Breakout Groups

MODELING TASKS AND BEHAVIORS

Group: Katriina Byström, Luanne Freund, Jingjing Liu, Gary Marchionini, Pertti Vakkari , Barbara Wildemuth

Search systems have been designed to support discrete, transactional searches despite wide recognition that search behavior, and information behavior more generally, is often embedded in and motivated by work tasks that prompt search processes that are often lengthy, iterative, and intermittent, and are characterized by distinct stages, shifting goals and multitasking. Furthermore, searching does not happen in isolation: ubiquitous access to networks, online content and search technology has created an environment in which searches are interwoven with other kinds of information seeking behaviors, such as reading, learning, communicating and acting in the real world. Our group discussion on the first day of the workshop focused on the question of how to represent and model this broader conception of task-based searching, with a focus on the transition points from one kind of activity to another, and from one goal or task to another. We were interested in modeling the way that search fits into people's lives.

Our discussion began by acknowledging that human information behavior occurs at many different task levels, including work tasks, information seeking tasks and information retrieval tasks, as articulated by Byström and Hansen (2005), and that existing models and conceptions of these different levels do not fit together well. Models of searching tend to be low level and overly simplistic, offering some guidance as to the probability of state transitions within a single search session, e.g., between querying and viewing results, but not considering task-switching, multi-tasking, human collaboration activities, system switching and the interplay of other kinds of information behaviors with search. On the other hand, models of information behavior tend to be underspecified with respect to search tasks, often treating the use of search systems simply as one of many possible sources of information (e.g., Leckie, Pettigrew & Sylvain, 1996).

This led us to focus on how we might establish a framework that connects models of information seeking tasks and information search tasks by focusing on the transitions between them, the probabilities of transitions taking place, and the triggers. We discussed the work task level and recognized that modeling at this level would be the most challenging, due to variation in work tasks across domains that would reduce the likelihood of strong general patterns. Rather, we consider the work task as an overarching problem or project: the motivating or embedding task for information seeking and searching. While essential in terms of establishing the information seeking goals and providing a basis for evaluating outcomes, we decided not to attempt to model the work task further than that. We used gardening as an example of an everyday life work task to frame our discussion: specifically, a back-yard landscaping/fencing project with a well-defined goal (to keep deer out of the yard), but an uncertain path and unknown set of inputs and constraints. We imagined the project and its requirements as evolving over time and prompting multiple cycles of information seeking and searching tasks as well as cognitive and physical tasks, such as selecting and purchasing supplies and doing landscaping work.

We considered a number of existing models and frameworks and discussed aspects of them that could be relevant to our goals:

- The Information horizons approach (Sonnenwald & Wildemuth, 2001), which considers the wide range of information seeking sources and strategies available to an individual when carrying out a task
- Early cognitive models of information retrieval interactions (e.g., Ingwersen, 1982, 1996; Belkin, 1980, 1990), as well as more recent cognitive models (e.g., Hung, Johnson, Kaufman & Mendonca, 2008)
- Information seeking and interaction models that take into account process and iteration:
  - Marchionini's (1995) Information Seeking Process Model
  - Toms' (2002) Information Interaction Model
- Models that represent tasks longitudinally and incorporate the concept of task stages:
  - Vakkari's (2001) Task-Based Information Seeking framework
  - Kuhlthau's (1991) Information Seeking Process Model
- Theories and models that focus on human activity and articulate different task/activity levels:
  - Activity theory (Kaptelinin & Nardi, 2006)
  - Goals, Operators, Methods and Selection Rules (GOMS) (Card, Moran & Newell, 1983)
- Models that identify the key elements of information seeking activities
  - Järvelin & Ingwersen's (2005) Information Seeking Model

In discussing our design scenario, we established three key steps for developing a working model. Our initial thoughts on each are summarized below.

Identify the elements of the model

We consider the elements of the model to be the searchers' options for task-based information seeking activity at any point in time. Some of the options we identified are shown in Figure 1.

Identify the structure of the model

We believe that this model of the process should focus on transitions and changes that occur over time and at different task stages. To represent its structure, we chose a simple state transition model with all elements at play at each point in time. The process is represented as the state of the user at ordered points in time (T1, T2, T3). The model could be developed further by incorporating hierarchical relationships and adding more depth to the hierarchies.

Identify task-based factors that are likely to influence the model

Such factors might include:

- task type, at any of the levels (for example: Work Task: administrative vs. managerial tasks or routine vs. complex tasks; Information seeking task: collaborative vs. individual; Search task: lookup vs. exploratory)
- task stage, at any level
- prior knowledge and expertise of the person completing the task; this factor will affect the extent to which tasks are habituated, conscious.

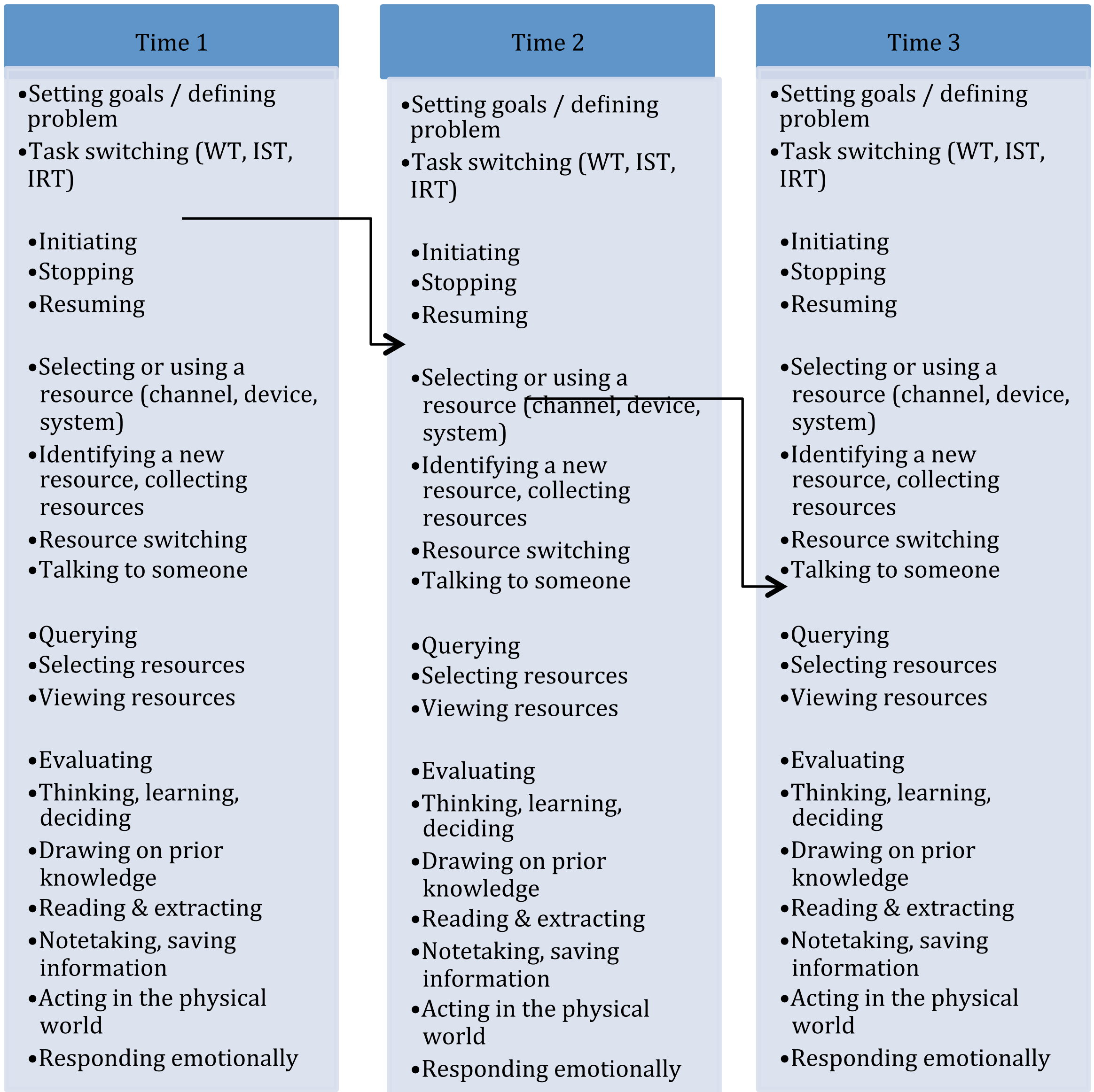


Figure 1: Simple state transition model of task-based information seeking

We expect that the main benefit of developing this model would be to provide support for searchers' moves and decisions (e.g., recommend tools, provide a means of maintaining and preserving state, and moving content back and forth) and to identify ways to integrate search into broader life activities. It may hold particular value for conceptualizing mobile search applications, as mobile search is likely to be tightly integrated with other types of information seeking activities. However, we also identified a number of challenges in developing and testing this model empirically. Because information behavior is domain dependent, the model can only function at a high level and therefore may be too abstract to offer significant value in understanding human behavior and/or informing design. Another difficulty arises out of the limitations of data collection methods with respect to tasks performed over time. Relatively simple methods such as transaction logging or interviews will always be incomplete, as the underlying intentionality of the searcher is either inferred or may be misrepresented through self-reporting. For this reason, more costly and intrusive naturalistic studies are needed. Evaluation is another challenge, as it will be necessary to assess the extent to which people are able to perform tasks which extend beyond the boundaries of a single system. This will require longitudinal data collection and the availability of benchmarks for comparison.

A number of types of studies could be designed to help develop and test this model. Any given study might focus on different levels of task or activity, and the part of the model not under consideration could be black boxed for purposes of that study. In this way, the model could serve as a general framework connecting many studies with different aims and focusing on different components of the model. In particular, studies focusing on compelling, real life information use environments would provide an effective grounding for the development and testing of this model. We discussed a range of data collection methods, targeting those that could function across different systems and types of activities and that would capture naturalistic behaviors. These included: diaries, transaction logs, photos, video recording, email and chat.

Our discussions on day two of the workshop carried on from these initial discussions to consider specific research proposals. These are presented in the sections to follow.

Information-Related Tasks of Patient Care Teams

Katriina Byström, Luanne Freund, Barbara Wildemuth

Building on the discussion on day 1 of the workshop, we focused on the goal of understanding how information search tasks are part of and complement other information seeking activities. We developed plans for an ethnographic study focusing on an information environment with a rich interplay of information related activities, including both interpersonal interactions among groups of professionals and non-professionals and online searching in a range of information systems. We chose the medical domain and focused our plans on team-based patient care as a study environment.

Research questions:

We identified two primary research questions to be addressed by the study:

- What information seeking tasks occur within this activity, and with which specific work tasks are they associated?
    - Selection of particular information resources
    - Transitions from one resource to another
    - Searches of existing information systems
    - Use of information
- Under what circumstances are transitions made from/to information seeking/use tasks and other tasks?

We conceptualized the problem at a high level, as one of identifying and distinguishing between different levels and types of tasks and examining the transition points between them. For example, within the team-based patient health care context, a common Work Task would be to develop a patient treatment plan. Within that Work Task would be a number of Information Seeking and Use Tasks as well as other types of tasks, such as communicating the plan to the patient and documenting the plan. Within each Information Seeking and Use Task there would be a number of Information Retrieval Tasks as well as other types of tasks (see Figure 2). The challenge in this project is to focus on the interplay and shifts between these different types of task, rather than isolating a particular kind of task, such as Information Retrieval tasks, and studying those in isolation.

Study setting and participants

The study will focus on team-based care of aging patients. In order to increase the research frame and external validity, we propose to conduct a cross-cultural parallel study in three locations: Canada, Norway and the United States. Working with a designated hospital in each location, researchers will follow the work of healthcare teams as they work with 2 to 3 patients over an extended period of time.

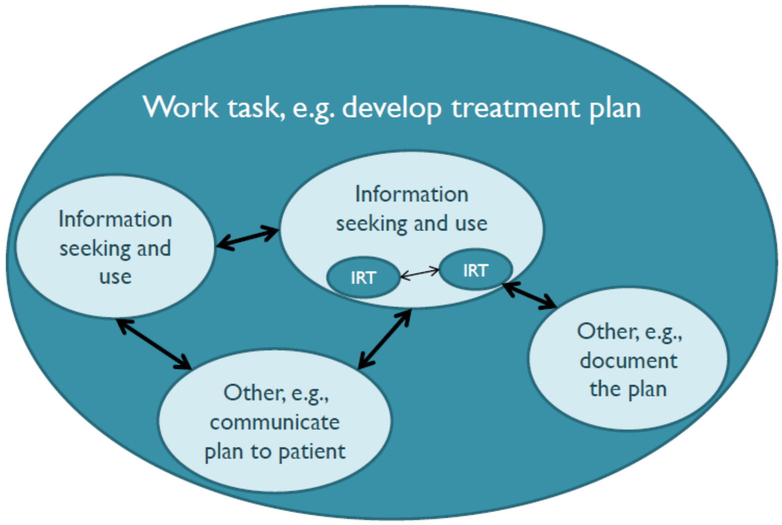


Figure 2: Conceptualization of the problem: task-based information seeking and retrieval

Data collection and analysis

Multiple methods of data collection will be used in order to capture the full range of activities of team members, as well as their information needs, goals and strategies. Particular attention will be paid to documenting transition points between information-related tasks, and the situations and triggers that prompted the transitions. The following data collection methods will be used:

- Direct observation of team activities, documented in field memos
- Interviews with all team members
- Additional notes/diaries from team members
- If possible, logging of activity on PDAs or other mobile devices, as well as the hospital's electronic medical records system

A range of qualitative data analysis methods will be used:

- Inducing concepts of interest directly from the data
  - Incorporating findings/concepts/models from prior studies
- Multiple levels of coding
  - Overall activities
  - Transitions between states
  - Specific search behaviors
- Constant comparative method
  - Comparing raw data with codes
  - Comparing codes across multiple data sources
  - Comparing codes with the categories in which they are grouped
  - Comparing categories and their definitions

Challenges and impact

A study such as this, which aims to capture a holistic perspective on a complex, real world, professional environment, is not without its challenges. The main challenges include the need to ensure patient privacy and maintain clear ethical boundaries, the difficulty of gaining access to treatment teams and the cooperation of all team members, and the diligence required to successfully observe and document information behaviors and activity shifts that could occur at any time.

Despite the obvious challenges, we believe it is important to take this broader perspective on information behavior. Such a perspective will lead to better understanding of the role of information seeking and search tasks within the broader context of specific work tasks. On a more practical level, this research has the potential to improve our understanding of how information resources are linked in the ecology of information use in this domain. Given the critical importance of coordination and collaboration in team-based health care, results can have a genuine impact in optimizing workflows, and informing the design of better tools and practices.

Health information seeking in everyday life: Stages, behaviors, and assistance

Gary Marchionini, Pertti Vakkari, Jingjing Liu

In day 1's group discussion, Group 2 discussed that information seeking should be put in a larger context of work task accomplishment instead of considering search tasks only. We came up with two general research questions: 1) What are the elements in and structure of a task-based information seeking activity model? 2) What are the transitional probabilities in this model?

Based on day 2's discussion, Group 2b proposed a research project that aims at exploring the above questions, as well as providing assistance to the information seekers based on the information seeking activity model. Considering the likelihood of variations in a user information seeking behavior model across domains, we thought it both practical and reasonable to start exploring these questions within a specific domain. We chose the health domain because a quantitative approach to building a model in the health information seeking area is needed, despite a rich literature mainly using qualitative methods.

The specific research questions for the proposed research are:

- What are users' information seeking behaviors (both cognitive and emotional) in different stages of a health acceptance model?
- Can we predict the health information seekers' stage in the health acceptance model based on their behaviors, for providing assistance/support for their future information seeking based on what is needed in the stage?
- What kinds of assistance can be provided to the health information seekers?
- What is the effectiveness of the assistance?

To answer these research questions, we plan to conduct a series of three studies.

First, we will conduct a qualitative study to collect data on what people diagnosed with a specific health condition do to seek information and emotional support. We plan to examine various health related discussion websites/forums/boards to collect the questions asked and answers/discussion provided. Content analysis will be conducted to explore the following aspects:

- What questions do people ask?
- What information do they want?
- What sources do they go to?
- What is the stage of their health condition?

Second, we will conduct a quantitative study to collect data on people's health information seeking activities. We plan to recruit people diagnosed with a certain illness. We will observe and record their activities related to seeking information and emotional support on this illness, in both digital devices (including the computer, smart phones, ipad, etc.) and non-digital devices (such as face-to-face communication). We will also collect data on other activities that they perform, including all other tasks that are not directly related to this specific health condition. These will help build users' information activity model, with transitional probabilities.

We will look at both individual and group information seeking behaviors. Some people that already participate in online group discussions will be considered as group information seekers.

Others that do not seek information through a group will be considered as individual information seekers.

We will use multiple methods to collect data. As mentioned above, some activities are supported through digital devices, and we will collect logs through these devices. For activities supported by non-digital devices, we will ask participants to record and report such activities, at intervals, through diaries, journals, questionnaires, and focus group discussions. We will also ask about their stage diagnosed with this condition.

Data analysis will be two-phased. Phase 1 will look at the behavioral patterns at different stages. Phase 2 will attempt to predict the stage based on users' behaviors. The purpose of including both descriptive and predictive approaches to the data analysis is that, if and once we are able to detect the stage that someone is in based on his/her behaviors, we will be able to provide suggestions/assistance of what sources to go to in order to obtain information and/or emotional support.

Third, we will conduct a controlled lab experiment to evaluate the effectiveness of the suggestions/assistance that are provided to the health information seekers based on their stage as indicated by their behaviors. An interface will be built that can provide suggestions/assistance to information seekers. In the evaluation study, a controlled group of participants will not be provided any assistance, and an experimental group of participants will be provided assistance. We will compare these two groups in terms of their health information seeking behaviors, their perceptions of the assistance's usefulness, and other aspects of their interactions during information seeking.

There are challenges in various aspects of these proposed studies. Participant recruitment, collection of people's information seeking behaviors (especially non-device supported activities), and data integration are all challenging. The impact of this research is to provide assistance to people's health information seeking and emotional support.

References

Belkin, N.J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science, 5,* 133-143.

Belkin, N.J. (1990). The cognitive viewpoint in information science. *Journal of Information Science, 16*(1), 11-15.

Byström, K., & Hansen, P. (2005). Conceptual framework for tasks in information studies. *Journal of the American Society for Information Science & Technology, 56*(10), 1050-1061.

Card, S.K., Moran, T.P., & Newell, A. (1983), *The Psychology of Human-Computer Interaction*, London: Lawrence Erlbaum Associates.

Hung, P.W., Johnson, S.B, Kaufman, D.R. & Mendonca, E.A. (2008). A multi-level model of information seeking in the clinical domain. *Journal of Biomedical Information*, 41(2), 357-370.

Ingwersen, P. (1982). Search procedures in the library--Analyzed from the cognitive point of view. *Journal of Documentation, 38*(3), 165-191.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation, 52*(1), 3-50.

Järvelin, K., & Ingwersen, P. (2005). *The Turn: Integration of Information Seeking and Retrieval in Context.* Dordrecht: Springer.

Kaptelinin, V. & Nardi, B.A.. (2006) *Acting with Technology: Activity Theory and Interaction Design.* MIT Press.

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.

Leckie, G.J., Pettigrew, K.E., & Sylvain, C. (1996). Modeling the information-seeking of professionals: A general model derived from research on engineers, health care professionals, and lawyers. *Library Quarterly*, *66*(2), 161-193.

Marchionini, G. (1995). *Information Seeking in Electronic Environments.* New York: Cambridge University Press.

Sonnenwald, D.H. & Wildemuth, B.M. (2001). A research method to investigate information seeking using the concept of information horizons: An example from a study of lower socio-economic student's information seeking behavior. *New Review of Information Behaviour Research*, *2*, 65–86.

Toms, E.G. (2002). Information interaction: providing a framework for information architecture. *Journal of the American Society for Information Science & Technology, 53*(10), 855-862.

Vakkari, P. (2001). A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, 57(1), 44-60.

Bringing search into task: Identifying task primitives

Fernando Diaz, Catherine Smith, Simone Stumpf, and Elaine Toms

Problem / Motivation / Goals

In work on interactive information retrieval, search is often studied in isolation from the task activities that prompt searching and which are dependent on the products of searching. *In situ* log studies describe search behavior that occurs in natural *task contexts* (e.g. planning a wedding), however these studies often lack information about that context, as well as data on the task activities that occur outside of the search system (e.g. recording information in a word processor). In experimental studies, research subjects are generally given a task context (e.g. pretend you are a journalist) and assigned specific task goals (e.g. find background information on the sequester). In most studies in this setting, task activities that might precede and follow search are hypothetical and little is learned about the dependencies between these activities and search. Longitudinal studies conducted in specialized domain settings (e.g. chemical engineering) have examined *in situ* task and search activities however, few studies have attempted to systematically extend this work across domains. In summary, little is known about the fundamental characteristics of dependencies between task activities and searching (hereafter called *task dependencies*). This knowledge is essential to the development of the task models required for support integrated search and task activities.

The investigation of task dependencies across domains requires a standardized framework for decomposing task structure to a sufficient level of detail. Because a suitable framework does not yet exist, part of the problem is to develop and test an initial framework. In applying the framework, the goal is to discover and describe task dependencies that occur universally across domains and task contexts; we term this type of dependency a *task-primitive*.

We propose to examine the question of whether there exists a set of task-primitives and to address the following research questions specifically.

Research questions

- Where in the flow of sub-task activities is search invoked?
- Can a set of task primitives be identified across diverse contexts and task activities?
- Can common dependencies between primitives be identified?
- How can descriptions of task primitives be standardized for research purposes?
- What might be the components of a generalized task activity model?

Approach

Our approach is to record, decompose, and analyze *in situ* task activities at a level of specificity sufficient to reveal task primitives. We will study two diverse task contexts, and will decompose task structure using the framework summarized below.

Data collection overview

Because we seek to understand naturally occurring task dependencies, a longitudinal design is required. Data will be collected using multiple methods and from multiple sources, including initial and periodic interviews about the task under study, collection of comprehensive log data from multiple devices, and participant diaries and annotations. Interviews will focus on expected and experienced task structure. Log data will record transitions between search systems and systems used in completing the task, as well as selected transactions within select systems. Periodically throughout data collection, participants will be asked to record annotations on specific aspects of task activities. Interviews, diaries, and annotations will be transcribed and coded. Coded data will be integrated with log data for analysis.

Selecting task contexts

We define a *task context* as a high-level multi-part goal. Task contexts differ in many dimensions. For the proposed study we focus on the specificity of the goal, level of collaboration, time urgency, duration, and geographic scale. Examples of task contexts include crisis response, travel planning, buying a house, selecting a school, developing a product, debugging code, managing a medical condition, handling a family legal matter, etc..

For the proposed study we have selected two task contexts that vary on several dimensions. The first, a small group assignment to be completed in an online course, involves collaboration with time urgency and limited duration. The second task context, personal management of an asthma condition, is primarily individual with a regular ongoing repeating time urgency. Both have specific articulated goals and limited geographic scale. The differences between the contexts provides the contrast required to address the question of whether task-primitives can be found in diverse contexts.

The framework and its application

In order to standardize the decomposition of task structure, we will use a three-phase iterative analytical framework. The framework sets aside questions of task detection or classification of tasks by type. Below, we step briefly through the application of the framework for analysis.

Phase 1: Task context decomposition

The table below outlines the two task contexts and examples of possible associated *task activities*, which we define as high-level multi-part sub-tasks of the task context. As defined in this initial version of the framework (we expect that the framework will evolve as we learn about task dependencies), task activities express the steps or composition of the larger context. For the asthma study the activities repeat regularly, while for the group project they have a finite time limit. It is expected that task activities will emerge during iterative application of the framework.

	TASK CONTEXT	
Phase 1: *Task activity decomposition*	Group Project in Online Graduate Course	Asthma Management
TASK ACTIVITIES (hypothetical)	• Select topic from list • Plan project • Assign duties • Conduct research • Share and analyze • Prepare paper	• Measure peak flow rate • Observe symptoms • Observe triggers • Adjust medication

Phase 2: Task activity analysis

The table below provides an example analysis of integrated data from one hypothetical task activity in the group project task context. The goal of analysis is to identify activity sequences that comprise task activities, and with those activities, to identify potential task primitives. It is expected that primitives will emerge during analysis of the integrated data. Analysis will be conducted using an combination of algorithmic and human processes.

In the example below participant Joe has used his laptop to search for information on the assigned project topic, looking a library catalog and a database service. He's recorded and saved notes using his word processor. Four days later Joe uses his cellphone to search the web, and browses a Wikipedia page and a YouTube video, all related to the project topic. The next day he sends an instant message to Jill, his partner on the project and she replies. She then uses her laptop to search a database service and saves PDFs of several articles to a remote server. As can be seen in the rightmost column, each sequence has been identified as a potential task primitive.

Phase 2: TASK ACTIVITY ANALYSIS (example task activity: Select Topic)						Pot. TASK PRIMITIVE
time	*participant*	*device*	*tool used*	*information input*	*information output*	
AM 9/8	Joe	laptop	Library catalog	topic descriptions	– bibliographic records	explore
		laptop	Database service	topic descriptions	– bibliographic records – article abstracts – full text document displays	
		laptop	Word processor	notes on topics	– text file	
PM 9/12	Joe	phone	Search engine	topic terms	– results pages	
			Wikipedia	links	– text pages	
			YouTube	topic terms	– video	
9/13 AM	Joe	phone	message	comment on progress	– message to Jill	contact
9/13 AM	Jill	phone	message	message from Joe	– comment on plan	
			message	comment on plan	– message to Joe	
		laptop	Database service	topic descriptions	– bibliographic records – article abstracts	gather
		laptop	remote server	full text documents	– saved PDF files	store

Phase 3. Analysis across task contexts and activities

In the third phase of analysis, potential task primitives will be aligned across task contexts and task activities, with the objective of identifying and describing any global characteristics for each primitive, and any features that may be task dependent.

Phase 3	TASK ACTIVITIES					
	Group Project			Asthma Management		
Potential TASK PRIMITIVES	select topic	…etc…	prepare paper	measure peak flow rate	…etc…	adjust medication
explore						
gather						
store						
route						
filter						
evaluate						
share						
compare						
integrate						
extract						
separate						
find trends						
contact						
….						

Challenges / Resources Required / Caveats

In order to capture task activities on all types of devices, log recording methods are needed. If these methods are not available, knowledge of task activities will be limited to those that occur on personal computers. A second challenge is participant recruiting and retention for long-term ubiquitous data collection. Because participation requires the disclosure of information likely to be perceived as personal and private, adequate incentives will be necessary. Also, because the data to be collected has a large scope and will be voluminous, integration and analysis will be complex and will require resources.

Plan for future

The proposed study is limited to two task contexts in two different domains. Task contexts vary on many other dimensions. Assuming the analytical framework proves to be a useful approach, further research in different domains and contexts will be needed to validate and extend the task primitives found in the first study.

Impact

Interactive search is ubiquitous in everyday life. It is integral to and useful for all types of life goals. At present, search is not well integrated with the tasks it supports. This makes information intensive tasks burdensome, particularly for populations that lack experience with the many applications often required to support task completion. By identifying and describing task primitives we may be able to determine those which have the broadest application and thus prioritize research and development efforts.

Task Modeling

Jae-wook Ahn, Gene Golovchinsky & Birger Larsen

Motivation and goals

An important aspect of bring “task into search” is to expand current search engines and systems with generic tools to support the user’s process of working intuitively on their (work) tasks. Currently the main burden of keeping track of the search process and the possible alternatives available is on the user. The overall purpose of the proposed approach is to map and understand which tools/components/widgets are most useful for which moves/activities/actions in a task-based search scenario. Many information seeking studies have identified specific moves, actions and activities in particular scenarios (e.g. Kumpulainen and Järvelin (2010) in the area of molecular medicine). The aim is not to deal with the specific scenarios, but rather to come up with a generic list of tools, some of which can be implemented in a given search scenario depending on the specific tasks in that scenario.

The goal is therefore to identify abstractions that underlie information seeking behaviours to compare systems, people and tasks. This includes identifying sufficiently abstract ways to describe and compare across these. The hypothesis is that common diversified patterns across systems can be identified, learned and understood and therefore used to inform design. The focus will be on essential features rather than surface features. The analysis is exploratory, and the aim is not to come up with a metamodel of information behaviour or information seeking - instead the goal is to identify stable patterns that can inform design.

Method and analysis

The first step is to logs as much user activity and contextual information as practically possible, e.g. querying, interaction with results lists, document inspection, query reformulation, information use etc. The type of task is not of prime importance, but details about the task and setting should be recorded for richer analysis.

Once activity data have been obtained a main component in the analysis is to carry out a state/transition analysis to identify typical patterns (e.g. using a Hidden Markov Model analysis). The motivation behind using such an approach is to get away from specific details and build more general models. The aim is to label higher level patterns (e.g. inspired by Bates’ tactics (Bates, 1979)) to facilitate identification of commonalities and differences in patterns across users, tasks, systems, experiments. Such knowledge can for instance help to detect if users show expected behaviour with a given system, and can inform surveys and detailed studies for system redesign and automatic critical incident identification leading to e.g. system simplifications or added features. For instance interpretation of patterns may help to identify if behaviour is efficient or shows examples of workaround because of system deficiencies.

Challenges

A main challenge will be to obtain sufficient amounts of activity data across different systems, users and tasks. Automatic recording of these, e.g. using automated observation techniques such as the browser plugins developed by Grzywaczewski et al. (2012), may aid in obtaining a critical mass of sufficiently detailed activity data. Another challenge is how to facilitate analysis and interpretation of patterns with large amounts of user data. Here visualisation tools may be helpful, e.g. to visualise temporal events. Also, although we do not focus on task types hidden task dependencies may affect results.

Impact

The main impact of this approach is that it will allow discovering and comparing commonalities and differences in patterns across users, tasks, systems and experiments, and thus build more general models to understand the relation between moves/activities/actions and tools/components/widgets. In addition, this approach can also facilitate the creation and sharing of comparable data where the overall patterns are collected, published and shard rather than the raw logs. This can overcome some of the privacy issues associated with raw log data and thus enable work across research groups.

Next Steps

A possible first step is to analyse already existing data sets, e.g. from information seeking and interactive IR studies. This will allow testing the overall approach and identifying challenges. Based on this a number of pilot studies can be designed and carried out, before moving to full scale data collection and analysis.

References

Bates, M.J. (1979: Information Search Tactics. Journal of the American Society for Information Science, 30, p. 205-214.

Grzywaczewski, A., Iqbal, R., Halloran, J., Iqbal, K. and James A. (2012): Copy and Paste as an Indicator of Relevance: Towards the Development of Intelligent Recommender Systems for Software Engineers. In: Larsen, B., Lioma, C. and de Vries, A. (eds.): *Proceedings of the ECIR 2012 Workshop on Task-Based and Aggregated Search (TBAS2012), held in Barcelona, Spain 1 April 2012*, p. 14-18. Available online at: http://itlab.dbit.dk/~tbas2012/.

Kumpulainen, S. & Järvelin, K. (2010): Information Interaction in Molecular Medicine: Integrated Use of Multiple Channels. In: Belkin, N. & al. (Eds.), *Proceedings of the Information Interaction in Context conference (IIiX 2010), New Brunswick, NJ, August 2010*. New York, NY: ACM, p. 95-104. (Best paper award, IIiX'10)

EVALUATION

Nick Belkin, Pia Borlund, Ben Carterette, Diane Kelly, Bill Kules, and Mark Smucker

Our group considered the problem of evaluation. This area has been extensively examined, and we did not attempt to enumerate the variety of definitions and challenges that have been proposed. Instead, we focused on a few themes that seem common throughout the literature. Traditionally, we have considered IR as a tool for finding documents, and evaluation consisted of assessing how good the documents are. This approach has been questioned and alternative models and measures have been proposed. For example, in addition to evaluating documents, we can evaluate both the information seeking and retrieval process, as well as the end results. We can consider evaluation at multiple levels, too. The scope of an evaluation can range (at least) from the entire session (with respect to the motivating task), to the usefulness of a particular system support (with respect to its own design intention or the higher level task). The evaluation can also be framed in terms of how well the system (or component) supports the person in optimizing the search process.

We identified several evaluation challenges as particularly apropos for task-based search, including ecological validity, trustworthiness of results and generalizability, domain-specific search. Creating measures that can be used within and across studies of users is an important challenge. Given a task-type, it is highly desirable to have a recommended set of measures that have been validated and calibrated so that they can be used to compare results between studies. This could be represented as a matrix of task-type by recommended measure. Other challenges include:

- Better ways to get people to opt into exploratory research of usefulness
- What constitutes a realistic environment in the lab
- How to minimize demand effects
- What data is needed to support simulation evaluation – build models & measures - characterizing the space of ways to get to an outcome

Usefulness of Task-Based Searches

Nick Belkin, Ben Carterette, and Mark Smucker

Problem / Motivation / Goals

In traditional information retrieval evaluation frameworks, we consider documents to be relevant or non-relevant and a user's interaction with a search system is reduced to being the rate and amount of relevant documents retrieved by a user. In task-based search, we do not know how users value their usage of search engines in support of their tasks. For example, we would like to, but do not know the answers to the following questions:

- Are there common stages that users experience while working on tasks?
- What are the intentions of users at these various stages of their task?
- How do users measure / describe the usefulness of their experience in the various stages (if stages exist, otherwise steps)?
- How do users measure / describe the overall usefulness of their interaction with the search engine?

Without the basic knowledge of how users value task-based searches, we are unable to properly evaluate retrieval systems' support for task-based search tasks.

Approaches

We propose two approaches to help us better understand how users value search systems for task-based search tasks. The first is a broad approach that would ask study participants to recall tasks and then discuss the usefulness of their actions. We envision using extensive instrumentation to collect detailed usage data from participants, e.g. eye-tracking laptops. The study would be conducted over a long time period of at least 3 months with weekly interviews conducted. In addition to instrumented laptops, we would ask participants to maintain diaries of their tasks and also to have the user flag relevant activity for later review with the researchers. The idea behind this broad approach is that recorded behavioral data often does not reveal users' motivations and actual tasks.

The second approach would be a controlled lab study. In this study, we would have assigned tasks that differ in the dimensions of expected behavior. Again, we would use a highly instrumented system. By using stimulated recall with the participants, we could carefully measure micro-actions without interruption of the activity. This second approach would logically follow after the conclusion of the first.

Nature of the Interview

The "broad approach" above calls for participants to use a highly instrumented laptop, but even with a collection of data, we still need some way to identify episodes and critical incidents to review with the participants. Some possible aids that could be used to identify task-based searches include:

- Elicitation question: This is a question designed to help participants recall useful incidents to examine with the researcher.
- Various aids such as diary entries, flagged points in time, and replayed screenshots / video: Each of these involve either the participant making note of interesting events when they occur or helping the participant recall the event from a replay of activity.

Likewise, the researcher could attempt to identify task-based searches from the record data, and it would be good to compare participant identified vs. experimenter identified episodes.

Once a task-based search episode is identified, the researcher would likely proceed through a series of questions with the participant including:

- What was your intention with this search?
- Is this task a one-off or a regular task?
- This looks where you started, is this right or was it earlier?

For the various events occurring in an episode, the questions would include:

- Can you remember what you wanted to do?
- Do you feel you were successful here?
- What makes you feel that way?
- How do feel about this search in the whole?

Once the quantitative behavior data and qualitative interview data has been collected, it would be necessary to analyze the interview content and tie it to the behavior data.

Analysis of Data

The interviews would produce a large amount of verbal data. In this data, the hope is that the participants would use words that are indicative of measures of usefulness. Likewise, we anticipate that these searches will go through various stages that can be generalized to most searches. With stages identified, we would then look for sequences of behavior that could be used to identify a stage. Likewise, we would attempt to classify the various identified tasks according to their characteristics.

Challenges / Resources Required / Caveats

We have outlined a large study involving the deployment of expensive laptops. Assuming the technology works, in a short period of time a large amount of data will be produced, and the amount of data may overwhelm attempts to understand it. Likewise, the verbal data is less reliable measure of value than we would like.

Impact

Today evaluation of search is tied to a notion that documents deliver the only value to the user. We intuitively know that there is value to search that goes beyond a set of relevant documents. By focusing on what users find useful, we have the potential for improving the evaluation of retrieval systems, and as our ability to measure effectiveness improves, so does our ability to build better systems.

Plan for future

As outline above, our proposal is to conduct an exploratory study to identify intentions and measures of usefulness in task-based search. Following this study, we will need a controlled experimental study to test the hypotheses generated by this study. The controlled study will try to determine if, for given tasks, do users have the same intentions and notions of usefulness that we identified in the naturalistic study? Assuming success with the exploratory and lab studies, we would then work on creation and validation of explicit measures for support of motivating task types and search intentions.

Meta-framework for comparable task-based evaluation

Pia Borlund, Diane Kelly, Bill Kules

On day 2, a small subgroup examined the challenge of comparability of evaluations. The current variety of research and reporting practices makes integration of multiple studies difficult. This is holding back meta-evaluations. Some modest changes in research practices could increase the comparability and interoperability of studies, through the clear and consistent description of studies and their results. As a start, we identified a meta-framework for studies that includes six elements: tasks, study design, measures, analysis, reporting and a matrix of measures. For each element, we explored actions that could be taken to document best practices for the design of each element of the evaluation. As shown in the table, reporting practices for users, tasks, etc. are well-established. Practices for tasks, study design and analysis methods could be synthesized from existing literature. The design of measures and a matrix for recommending which measures to use by task-type will require new research.

Elements	Actions
Tasks	Synthesis (being done)
Study Design	Synthesis (to-do)
Measures	R&D Needed
Analysis	Synthesis (to-do)
Reporting practices • Users • Tasks • …	Now
Matrix	Future

The potential impacts of such a framework include facilitation of future meta-studies, as well as longer-term historical analysis (e.g. 50+ years). A framework will also encourage better reflection on specific practices and their limitations. It could also yield educational benefits to researchers and students and enhance research integrity. Unintended consequences could include mimicry (that is, encouraging copying rather than learning) and overall rigidity that slows innovation. Researchers may also be resistant to change their practices.

Nevertheless, such a framework could be beneficial. Several possible steps could be taken, starting with proposals for best practices on reporting tasks. Design and analysis elements could be synthesized from the existing literature. More research and development will be required for the analysis element. The entire framework will need to be evaluated. This could be initially accomplished by applying it to analyze existing studies.

TOOLS AND SUPPORT

Tools to support the workflow

Fernando Diaz, Sue Dumais, Jaap Kamps, Cathy Smith, Simone Stumpf, Elaine Toms, Arjen de Vries, and Maria Zemankova

Problem

The breakout group focused on the following two problems: First, how to embed search into the work task? Second, how can the work environment inform search? This generated lots of interesting discussion on a variety of explicit tasks, such as genealogy, term paper writing, environmental managing, wedding planning, trip planning, school selection, etc.

The discussion of tasks was rather unstructured, yet four broad observations surfaced several times. First, there seemed to be a family resemblance between the tasks discussed: each of them being a complex task requiring the combination of a variety of sources, and occurring in a session or episode that takes hours or days and weeks to finish. Second, most of these tasks have as goal a specific output that is essentially an information aggregation (family tree, written report, plan), which creates an obvious direct measure of (task) progress and success. Third, the discussion was problem driven: there was substantial criticism about how these tasks are now supported by current web search, verticals, apps, or dedicated software. Fourth, in addition to their general usefulness to support a task at hand, there is clear importance of such tools for those with poorer information, media, or search literacy.

Approach

Concerning the first problem of how to embed search into the work task. This requires a broad definition of information access methods rather than a narrowly defined search task. It also requires a holistic approach, starting from the whole task (work environment) and including all activities (such as reading, writing, thinking) beyond the narrow search interaction. Complex tasks have internal structure: each tasks has many subtasks with their own subgoals, rather than singling out the search (sub)task, there are information access activities embedded in each of the (sub)tasks.

Concerning the second problem of how can the work environment inform search. Here the task structure and constraints on the task output are a valuable source to exploit. This allows for a separation between supporting task specific aspects (supporting the product) and supporting the search or information access process (supporting the process). This will require lots of data interchange, making search aware of the tasks context and all previous activities, which raises ethical and architectural questions (locally stored, centralized cloud server, peer-to-peer exchange based on a need-to-know basis).

Challenges

Various challenges were identified: How to extract (search) tasks out of the work environment: how to define the primitives (what are 'atomic' task aspects)? Tasks are distinct from vanilla search because of 'clear' goal/success states, and task context or even templates for the `ideal' information seeking process (e.g., evidence based medicine): how to incorporate these? How to balance generic tools and task-specific tools unique to a case at hand? Task and subtasks have intricate dependencies: how to take them into account? We need more than a note-taking/scratch pad! There is a need for flexibility in the workflow through the different steps, not only a rigid waterfall model with stages, but can we skip components/steps (look ahead) and backtrack? What are crucial features of the user profiles and contexts (in particular in collaborative task)? How can we infer or extract the needed information? Many tasks are collaborative, even those done by a single person may have different roles, how can we exploit the support tools in collaborative situations.

Impact

There have been systems that support professional tasks for many years, but often offering a specific additional search tool, rather than being embedded into the work task as ambient search support. In today's always online world, the distinction between search tasks and work tasks is increasingly blurred, with all of our work stored in the clouds (including a full transaction history) and all our context and interactions logged and shared between all on our devices. Hence this is the time to integrate search into our tasks (both in work and in daily life).

Next steps

A sensible approach is to focus on the overarching tools that piece together results from subtasks (which could be web searches, intranet, …). This would balance generic tools and task-specific tools: with support for the task process (decomposition and aggregation) being fairly generic, and specific subtasks potentially requiring dedicated task specific modules. In the evaluation, it is essential to look at a broader range of factors than only search task success (i.e., "whether an answer is found") but also on how well the tool supported finding the answers, and how this helped in achieving the overall task's goals (in particular the created information product).

Abdigani Diriye, Rob Capra & Jaime Teevan

Introduction

Summarizing information is a key technique in information retrieval. It allows us to convey the gist of an information object in the most concise form possible to users. We observe that much of the work in summarization has focused on individual documents, while broader applications of this technique have been underexplored. Work by Capra & Marchionini [1] and Golovchinsky & Diriye [2] are examples of summarization research that pushes outside the box. Capra & Marchionini [1] applied summarization to document collections, and Golovchinsky and Diriye [2] bootstrapped a summary of the search results to the query box. In our subgroup we explored ways we can apply and extend summarization to different contexts.

Beyond Document Summarization

Summarization has been studied and applied to documents to construct a one to three sentence surrogate of a document. While approaches in this area have been well explored, we find other novel applications such as search results, search tasks, and sessions to be underexplored yet promising avenues for investigation.

Search Results

Producing a search result summary would involve processing the search results on the SERP (Search Engine Result Page) and giving a gist of what they, as a set, entail. This can involve providing an overview of the search results on the page, surfacing the common themes and topics on the SERP, and teasing out what documents are similar or different to previous search results.

Sessions

One level up from search result summaries lie session-level summaries, which involves summarization across multiple SERPs or documents to give users a gist of the information encountered over a period of time. Summarizing information at the session-level can:

- Help searchers get a deeper sense of the content encountered
- Help searchers understand what information is different, similar or substantiate information found in a previous query

Tasks

Tasks span multiple sessions, and summarization at this level can result in faster task completion time, accelerate knowledge acquisition and, encourage task resumption. Queries, documents, and browser-level actions like pages bookmarked and links clicked, for example, can be used to generate task summaries.

Research Questions

From our discussion, a number of research questions have been raised, including:

- What makes a good search result/session/task summary?
- How useful are summaries across different tasks?
- How can such summaries be constructed?

Evaluation

To evaluate summaries for search results, sessions and tasks, a number of metrics can be employed that focus on their quality, utility and usefulness. Low-level metrics like user preference, time-to-click, click-back, relevance judgements, dwelltime, etc. can give us a good idea.

[1] R. G. Capra and G. Marchionini. 2008. The relation browser tool for faceted exploratory search. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries* (JCDL '08).

[2] G. Golovchinsky, A Diriye. 2011. Session-based search with Querium. In *Proceedings of the HCIR 2011 Workshop, Mountain View, CA, USA*.

Mobile Phone App

Eugene Agichtein, Jaime Arguello, Christina Lioma, and Ryen White

Day 1 – Research Topic Discussion

Mobile apps (short for "mobile applications") are software tools that are specifically designed to accomplish a particular higher-level task from a mobile device. For example, Shazam[1] is an app that automatically identifies a song currently playing in the user's physical space and then gives the user the option to purchase the song/album online. As the usage of smartphones and tablets continues to grow, people are increasingly in situations where they are depending on their mobile devices to assist with task completion. There is an opportunity to provide a service that will recommend a relevant application (or set of applications) to users based on information about their current task gathered implicitly from behavioral data and other contextual signals such as user location and time of day.

Our group discussion focused on the task of *app recommendation*. Specifically, we focus on apps that help users accomplish tasks that require several subtasks, such as planning an evening out or purchasing a product. The goal for an app recommendation system would be to automatically recognize the task the user is trying to accomplish (using information derived from the user's context) and to recommend apps that might assist the user in accomplishing this task. Given that mobile devices can capture rich information about a user's geo-temporal context (and compare this with historical information from the user and other users), we focused our discussion to the mobile domain. However, the same idea could also generalize to the desktop domain, where operating systems such as Microsoft Windows are also moving toward an application model. Example tasks in the mobile domain include preparing dinner for guests (deciding what to cook, what ingredients to buy and where, and learning how to prepare each item) or planning a day of activities in a new city (deciding on places to visit throughout the day given a set of budgetary constraints). For these tasks, the user may need to collect and synthesize information from multiple sources and leverage different applications to accomplish each of the subtasks.

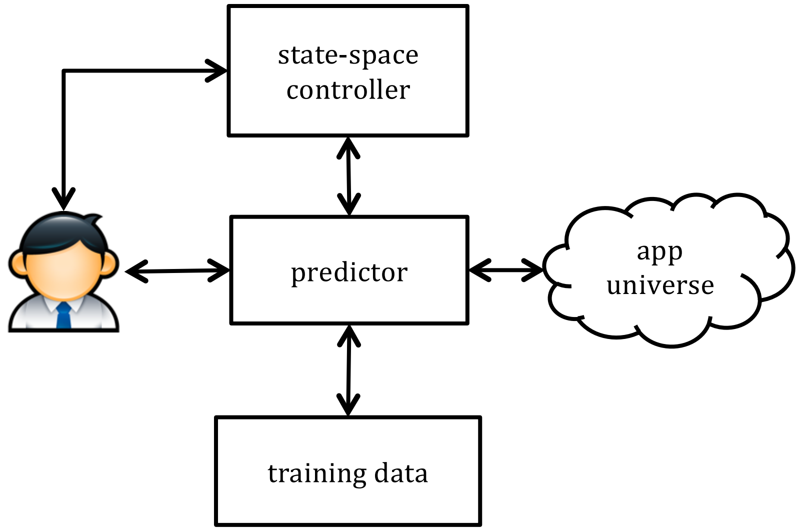
Our discussion of app recommendation was driven by the following questions:

1. What would the system architecture look like?
2. What types of contextual signals or features might the system use to predict app relevance?
3. What are some of the major challenges in developing such a system?
4. What are some of the related subfields of IR?

The basic system architecture is shown below.

[1] http://www.shazaam.com

[2] Estimating recall is challenging when relevance is derived from user interactions with



We envision the following inputs to the system (the predictor): a universe of retrievable mobile apps (either installed on the user's mobile device or available online), some representation of the user's context (possibly enhanced by the state-space controller), and a set of positive/negative recommendation instances to be used as training data. Each training instance would consist of a context-app pair and a positive/negative relevance label.

The predictor would be a machine learned model trained to predict app relevance as a function of a set of input features. Possible features could be derived from the user's context (including location, time, speed and direction of travel, proximity of friends, etc.), from meta-data associated with the app, and from previous user interactions with the system. From this information, the predictor would need to create a representation of the task that could be used to find related applications. Training labels could be derived from either explicit or implicit user feedback. The training data could be initially gathered from third-party judges who could make a determination based on the appropriateness of the application for a context provided to them, but could then shift to the utilization of online signals based on users' reaction when the recommendations are made (e.g., did they accept or decline the invitation to use the application when it was offered to them?). The goal of the state-space controller would be to derive higher-level contextual features that might narrow the space of possible tasks the user is trying to accomplish. These higher-level features could then be incorporated into the predictor's feature representation. An example higher-level feature might be whether the user is currently in an outlier location. The state-space controller could make such a prediction using the current time and location information and comparing this with the user's historical profile. Knowing that the user is an unexpected location may suggest that the user is trying to accomplish a task that requires finding a business or landmark, or perhaps in addition to their inferred task, they may also need help in situating themselves within their physical environment.

The following are some of the major challenges we identified. At the core, we view app recommendation as a search problem. Given a user who is performing a particular task, the goal for the system is to retrieve one or several apps likely to be relevant to the task. Thus, some of the same challenges associated with any search system apply: modeling the information need (the task in this case), finding features that are correlated with app relevance, estimating recall[2], and personalizing results in light of data sparseness. However, we believe that other challenges are unique to app recommendation. Our group discussion focused on two unique challenges. First, what is the optimal stage within the task-completion lifecycle to recommend a potentially relevant app? Recommending an app too soon may interrupt the user's cognitive flow. The user may not yet fully understand the steps and processes required to complete the task and may therefore not recognize a relevant app recommendation. On the other hand, recommending an app too *late* may render the recommendation useless. The user may have already committed to accomplishing the task using one or several familiar (and potentially suboptimal) tools and may lose much momentum in switching. Second, how should the system transfer task-state. Recognizing that the user is working on a task implies that the user has already started working on the task. Thus, it seems necessary to be able to transfer information about the already-completed portion of the task to the new app. Otherwise, the user may incur too heavy of a cost in switching. The latter challenge requires some degree of interoperability between apps, perhaps even in the form of a standard for information sharing.

Several areas within IR seem to be related to app recommendation. *Contextual suggestion* is the task of recommending things to do for a user that is in a particular geo-temporal context and has been investigated at TREC 2012 and 2013. This is a simpler problem than app recommendation as the user's task is consistently "entertain me" and the retrievable units are webpages associated with local businesses and landmarks. However, some of the same challenges apply. The system has to isolate recommendations that are appropriate for the user's context and must incorporate prior user feedback in order to personalize results. Other related areas include multimedia retrieval, which involves retrieving items that are inherently associated with text. There is also work recent work on task and user modeling whereby models of users' active search interests and intentions are inferred from their search behavior and used to predict future activity or select relevant search results.

Another area related to app recommendation is *mobile search*. The advanced functionalities of mobile computing devices (built-in cameras, scanners, position aware micro-sensors such as GPS, digital compasses and accelerometers) coupled with the availability of fast internet connections through 3G networks create opportunities for developing mobile web IR systems that are capable of responding to user needs 'on the go'. Limitations in screen real estate, processing speed, and lower user tolerance to bad design, provide motivations for exploring novel mobile user needs and interaction paradigms, where IR and HCI know-how must be considered alongside. Several spatially aware mobile applications that can search for information in the proximity of a point of interest exist (e.g. AroundMe for iPhones[3] and GeoVector[4]). App recommendation in mobile devices is a challenging research subarea in that domain.

Day 2 – Preliminary Study Design

During our Day 2 discussion, we set out to design a user study to investigate how two factors affect users' willingness to accept an app recommendation: task completion stage and prior

---

[2] Estimating recall is challenging when relevance is derived from user interactions with search results.

[3] http://www.aroundme.com/

[4] http://www.geovector.com/

knowledge about the task. More specifically, the study is motivated by the following three research questions:

- RQ1: Are users more likely to accept an app recommendation during different stages of task completion?
- RQ2: Are you users more likely to accept an app recommendation when they have greater prior understanding of the sequence of steps required to complete the task?
- RQ3: Is there an interaction effect between these two variables? That is, does prior knowledge of the task steps moderate the effect of task completion stage on a user's willingness to accept a task recommendation?

The study would have two experimental variables: (1) task completion (early, middle, late) and (2) prior knowledge (yes, no). The study would also have four outcome measures: (1) app recommendation acceptance (Did the user accept the app recommendation and use it to complete the task?), (2) post-task satisfaction, (3) task completion time, and (4) effort (measured as the number of actions required to complete the task).

The study could have a within-subject design. Each participant would complete six tasks using a mobile device. For three of the tasks, the participant would have prior knowledge of the steps required to complete the task. For the other three, the participant would have no prior knowledge. For two of the tasks, the app would be recommended in the first stage, second stage, and third stage, respectively. Participants would have the option of completing the task using the recommended app, or to complete the task using a set of baseline mobile tools familiar to the participant.

Our main hypothesis is that users are more likely to accept an app recommendation in the early stages of task completion (but not the first stage since they may be happily focused on their task and not in need of assistance). We also hypothesize that prior knowledge of the stages required to complete the task moderates the likelihood that they will accept the app recommendation early. In other words, we conjecture that a user may be more likely to accept the recommendation if they have prior knowledge of the task stages and are more aware its complexity *a priori*.

The following are some possible example tasks:

cook:

1. pick cuisine (early)
2. get recipes (middle)
3. find ingredients (late)

restaurant:

1. find restaurant (early)
2. book table (middle)
3. get directions (late)

travel:

1. choose destination (early)
2. book travel (middle)
3. find things to do (late)

Bibliography

Eugene Agichtein, Ryen W. White, Susan T. Dumais, and Paul N. Bennet. Search, interrupted: understanding and predicting search task continuation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 315–324, New York, NY, USA, 2012. ACM.

Jae-wook Ahn and Peter Brusilovsky. Adaptive Visualization for Exploratory Information Retrieval. To appear in *Information Processing and Management*.

Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. Open user profiles for adaptive news systems: help or harm? In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 11–20, New York, NY, USA, 2007. ACM.

Jaime Arguello, Wan-Ching Wu, Diane Kelly, and Ashlee Edwards. Task complexity, vertical display and user interaction in aggregated search. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 435–444, New York, NY, USA, 2012. ACM.

Joan C. Bartlett and Elaine G. Toms. Developing a protocol for bioinformatics analysis: An integrated information behavior and task analysis approach: Research articles. *Journal of the American Society for Information Science*, 56(5), 469–482, 2005.

Marcia J. Bates: Where should the person stop and the information search interface start? Inf. Process. Manage. 26(5): 575-591 (1990) http://dx.doi.org/10.1016/0306-4573(90)90103-9

Nicholas J. Belkin. Really supporting information seeking: A position paper. Information Seeking Support Systems. Technical Report. 2009.

Nicholas J. Belkin. On the Evaluation of Interactive Information Retrieval Systems. In: Larsen, B., Schneider, J.W., & Åström (Eds.). *The Janus Faced Scholar: A Festschrift in Honour of Peter Ingwersen*. Copenhagen: Royal School of Library and Information Science. 13–21, 2010.

Nicholas J. Belkin. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42(1), 47–54, 2008.

Pia Borlund. The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8(2), 2003.

Pia Borlund, Sabine Dreier, and Katriina Byström. What does time spent on searching indicate? In *Proceedings of the 4th Information Interaction in Context Symposium*, IIIX '12, pages 184–193, New York, NY, USA, 2012. ACM.

Oliver Brdiczka, Norman Makoto Su, and James Bo Begole. Temporal task footprinting: identifying routine tasks by their temporal patterns. In *Proceedings of the 15th International Conference on Intelligent user interfaces*, IUI '10, pages 281–284, New York, NY, USA, 2010. ACM.

Benjamin Bustos and Tomáš Skopal. Dynamic similarity search in multi-metric spaces. In *Proceedings of the 8th ACM International workshop on Multimedia Information Retrieval*, MIR '06, pages 137–146, New York, NY, USA, 2006. ACM.

Katriina Byström. Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science*, 53(7), 581–591, 2002.

Katriina Byström and Kalervo Järvelin. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2), 191– 213, March 1995.

Katriina Byström and Annemaree Lloyd. Practice theory and work task performance: How are they related and how can they contribute to a study of information practices. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–5, 2012.

Donald J. Campbell. Task complexity: A review and analysis. *The Academy of Management Review*, 13(1), 40–52, 1988.

Robert Capra, Gary Marchionini, Jung Sun Oh, Fred Stutzman, and Yan Zhang. Effects of structure and interaction style on distinct search tasks. In *Proceedings of the 7th ACM/IEEE-CS joint Conference on Digital libraries*, JCDL '07, pages 442–451, New York, NY, USA, 2007. ACM.

Robert Capra, Gary Marchionini, Javier Velasco-Martin, and Katrina Muller. Tools-at-hand and learning in multi-session, collaborative search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 951–960, New York, NY, USA, 2010. ACM.

Erin A. Carroll and Celine Latulipe. The creativity support index. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 4009–4014, New York, NY, USA, 2009. ACM.

Ben Carterette, Paul N. Bennett, David Maxwell Chickering, and Susan T. Dumais. Here or there: preference judgments for relevance. In *Proceedings of the IR research, 30th European Conference on Advances in Information Retrieval*, ECIR'08, pages 16–27, Berlin, Heidelberg, 2008. Springer-Verlag.

Stefano Ceri, Alessandro Bozzon, and Marco Brambilla. The anatomy of a multi-domain search infrastructure. In *Web Engineering, volume 6757 of Lecture Notes in Computer Science*, pages 1–12. Springer Berlin Heidelberg, 2011.

Praveen Chandar and Ben Carterette. Using preference judgments for novel document retrieval. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 861–870, New York, NY, USA, 2012. ACM.

Duen Horng Chau, Brad Myers, and Andrew Faulring. What to do when search fails: finding information by association. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 999–1008, New York, NY, USA, 2008. ACM.

Lydia B. Chilton and Jaime Teevan. Addressing people's information needs directly in a web search result page. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 27–36, New York, NY, USA, 2011. ACM.

Paolo Ciaccia and Marco Patella. Principles of information filtering in metric spaces. In *Proceedings of the 2009 Second International Workshop on Similarity Search and Applications*, SISAP '09, pages 99–106, Washington, DC, USA, 2009. IEEE Computer Society.

Kevyn Collins-Thompson, Paul N. Bennett, Ryen W. White, Sebastian de la Chica, and David Sontag. Personalizing web search results by reading level. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 403–412, New York, NY, USA, 2011. ACM.

William S. Cooper. Expected search length: A single measure of retrieval effectiveness based on the weak ordering action of retrieval systems. *American Documentation*, 19(1), 30–41, 1968.

William S. Cooper. On selecting a measure of retrieval effectiveness. *Journal of the American Society for Information Science*, 24(2), 87–100, 1973.

Arjen P. de Vries, Wouter Alink, Roberto Cornacchia: Search by strategy. *ESAIR* 2010:27-28 http://doi.acm.org/10.1145/1871962.1871979

Debora Donato, Francesco Bonchi, Tom Chi, and Yoelle Maarek. Do you want to take notes?: identifying research missions in yahoo! search pad. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 321–330, New York, NY, USA, 2010. ACM.

Henry A. Feild, James Allan, and Rosie Jones. Predicting searcher frustration. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 34–41, New York, NY, USA, 2010. ACM.

Henry Allen Feild and James Allan. Task-aware search assistant. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 1015–1015, New York, NY, USA, 2012. ACM.

Norbert Fuhr. Logical and conceptual models for the integration of information retrieval and database systems. In *East/West Database Workshop*, Klaenfurt, pages 206–218. Springer Verlag, 1994.

Norbert Fuhr. Salton award lecture information retrieval as engineering science. *SIGIR Forum*, 46(2), 19–28, 2012.

Gene Golovchinsky, Abdigani Diriye, and Tony Dunnigan. The future is in the past: designing for exploratory search. In *Proceedings of the 4th Information Interaction in Context Symposium*, IIIX '12, pages 52–61, New York, NY, USA, 2012. ACM.

Victor M. Gonzalez and Gloria Mark. Constant, constant, multi-tasking craziness: managing multiple working spheres. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 113–120, New York, NY, USA, 2004. ACM.

Catherine Guinan and Alan F. Smeaton. Information retrieval from hypertext using dynamically planned guided tours. In *Proceedings of the ACM Conference on Hypertext*, ECHT '92, pages 122–130, New York, NY, USA, 1992. ACM.

Ahmed Hassan and Ryen W. White. Task tours: helping users tackle complex search tasks. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pages 1885–1889, New York, NY, USA, 2012. ACM.

Ken Hinckley, Xiaojun Bi, Michel Pahud, and Bill Buxton. Informal information gathering techniques for active reading. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1893–1896, New York, NY, USA, 2012. ACM.

Forest Woody Horton: Understanding information literacy: a primer; an easy-to-read, non-technical overview explaining what information literacy means, designed for busy public policy-makers, business executives, civil society administrators and practicing, UNESCO, 2008

Rosie Jones and Kristina Lisa Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 699–708, New York, NY, USA, 2008. ACM.

Evangelos Kanoulas, Ben Carterette, Mark Hall, Paul Clough, and Mark Sanderson. Overview of the 2011 Session Track. In *Proceedings of 20th Text Retrieval Conference*, TREC '11. 2011.

Evangelos Kanoulas, Ben Carterette, Mark Hall, Paul Clough, and Mark Sanderson. Overview of the 2012 Session Track. In *Proceedings of 21st Text Retrieval Conference*, TREC '12. 2012.

Diane Kelly and Colleen Cool. The effects of topic familiarity on information search behavior. In *Proceedings of the 2nd ACM/IEEE-CS joint Conference on Digital libraries*, JCDL '02, pages 74–75, New York, NY, USA, 2002. ACM.

Alexander Kotov, Paul N. Bennett, Ryen W. White, Susan T. Dumais, and Jaime Teevan. Modeling and analysis of cross-session search tasks. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 5–14, New York, NY, USA, 2011. ACM.

Bill Kules and Robert Capra. Influence of training and stage of search on gaze behavior in a library catalog faceted search interface. Journal of the American Society for Information Science and Technology, 63(1), 114–138, 2012.

Yuelin Li and Nicholas J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, 44(6), 1822–1837, 2008.

Yuelin Li and Nicholas J. Belkin. An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology*, 61(9), 1771– 1789, 2010.

Zhen Liao, Yang Song, Li-wei He, and Yalou Huang. Evaluating the effectiveness of search task trails. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 489–498, New York, NY, USA, 2012. ACM.

Jingjing Liu and Nicholas J. Belkin. Searching vs. writing: Factors affecting information use task performance. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–10, 2012.

Jingjing Liu, Chang Liu, Michael Cole, Nicholas J. Belkin, and Xiangmin Zhang. Exploring and predicting search task difficulty. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12. ACM, New York, NY, USA, 1313-1322. 2012.

Chang Liu, Michael Cole, Eun Baik, and Nicholas J. Belkin. Rutgers at the TREC 2012 Session Track. In *Proceedings of 21st Text Retrieval Conference*, TREC '12. 2012.

Marianne Lykke, Birger Larsen, Haakon Lund, and Peter Ingwersen. Developing a test collection for the evaluation of integrated search. In *Proceedings of the 32nd European Conference on Advances in Information Retrieval*, ECIR'2010, pages 627–630, Berlin, Heidelberg, 2010. Springer-Verlag.

Stefano Mizzaro. Relevance: the whole history. Journal of the American Society for Information Science, 48(9), 810–832, 1997.

Dan Morris, Meredith Ringel Morris, and Gina Venolia. Searchbar: a search-centric web history for task resumption and information refinding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1207–1216, New York, NY, USA, 2008. ACM.

James Pitkow, Hinrich Schütze, Todd Cass, Rob Cooley, Don Turnbull, Andy Edmonds, Eytan Adar, and Thomas Breuel. Personalized search. *Communications of the ACM*, 45(9), 50–55, September 2002.

Mark E. Rorvig. The simple scalability of documents. *Journal of the American Society for Information Science*, 41(8), 590–598, 1990.

Ian Ruthven. Interactive information retrieval. ARIST, 42(1): 43-91 (2008). http://dx.doi.org/10.1002/aris.2008.1440420109

Adish Singla, Ryen White, and Jeff Huang. Studying trail finding algorithms for enhanced web search. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 443–450, New York, NY, USA, 2010. ACM.

Alan F. Smeaton. Techniques used and open challenges to the analysis, indexing and retrieval of digital video. *Information Systems*, 32(4), 545– 559, 2007.

Mark D. Smucker and Charles L.A. Clarke. Time-based calibration of effectiveness measures. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 95–104, New York, NY, USA, 2012. ACM.

Amanda Spink. Multitasking information behavior and information task switching: An exploratory study. *Journal of Documentation*, 60(4), 336–351. 2004. Emerald.

Bin Tan, Xuehua Shen, and ChengXiang Zhai. Mining long-term search history to improve search accuracy. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, pages 718–723, New York, NY, USA, 2006. ACM.

Robert S. Taylor. Information use environments. In B. Dervin (Ed.) *Progress in Communication Sciences*, 10, 217–225,1991

Nava Tintarev and Judith Masthoff. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22, 399–439, 2012.

Elaine G. Toms, Luanne Freund, Richard Kopak, and Joan C. Bartlett. The effect of task domain on search. In *Proceedings of the 2003 Conference of the Centre for Advanced Studies on Collaborative research*, CASCON '03. IBM Press, 2003.

Elaine G. Toms, Robert Villa, and Lori McCay-Peet. How is a search system used in work task completion? *Journal of Information Science*, 39(1), 15–25, 2013.

Randall H. Trigg. Guided tours and tabletops: tools for communicating in a hypertext environment. *ACM Transactions of Information Systems*, 6(4), 398–414, 1988.

Pertti Vakkari and Saila Huuskonen. Search effort degrades search output but improves task outcome. *Journal of the American Society for Information Science and Technology*, 63(4), 657–670, 2012.

Annika Waern. User involvement in automatic filtering: An experimental study. *User Modeling and User-Adapted Interaction*, 14(2-3), 201–237, 2004.

Yu Wang, Xiao Huang, and Ryen W. White. Characterizing and supporting cross-device search tasks. In *Proceedings of the sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 707–716, New York, NY, USA, 2013. ACM.

Stina Westman. Image Users' Needs and Searching Behaviour, pages 63–83. John Wiley & Sons, Ltd, 2009.

Alan Wexelblat and Pattie Maes. Footprints: history-rich tools for information foraging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ‘99, pages 270–277, New York, NY, USA, 1999. ACM.

Ryen W. White, Wei Chu, Ahmed Hassan, Xiaodong He, Yang Song, and Hongning Wang. Enhancing personalized search by mining and modeling task behavior. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW ‘13, New York, NY, USA, To Appear. ACM.

Ryen W. White, Susan T. Dumais, and Jaime Teevan. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM ‘09, pages 132–141, New York, NY, USA, 2009. ACM.

Ryen W. White, Ian Ruthven, Joemon M. Jose, and C. J. Van Rijsbergen. Evaluating implicit feedback models using searcher simulations. *ACM Transactions of Information Systems*, 23(3), 325–361, 2005.

Wildemuth, B.M., & Freund, L. (2009). Search tasks and their role in studies of search behaviors. Paper presented at HCIR 2009: Bridging Human-Computer Interaction and Information Retrieval, Washington, DC, October 23, 2009.

Barbara M. Wildemuth and Luanne Freund. Assigning search tasks designed to elicit exploratory search behaviors. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval*, HCIR ‘12, pages 1–10, New York, NY, USA, 2012. ACM.

Allison Woodruff, Andrew Faulring, Ruth Rosenholtz, Julie Morrsion, and Peter Pirolli. Using thumbnails to search the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ‘01, pages 198–205, New York, NY, USA, 2001. ACM.

Kaidi Zhao, Bing Liu, Thomas M. Tirpak, and Weimin Xiao. Opportunity map: a visualization framework for fast identification of actionable knowledge. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM ‘05, pages 60–67, New York, NY, USA, 2005. ACM.