Information Seeking Support Systems

An invitational workshop sponsored by the National Science Foundation

June 26-27, 2008
Chapel Hill, NC USA

http://ils.unc.edu/ISSS/

Organizers:

Gary Marchionini, UNC-Chapel Hill, USA Co-Chair
Ryen White, Microsoft, USA Co-Chair
Nick Belkin, Rutgers University, USA
Gene Golovchinsky, FxPAL, USA
Diane Kelly, UNC-Chapel Hill, USA
Peter Pirolli, Xerox PARC, USA
mc schraefel, U. Southhampton, UK

Report: May 2009

NSF Workshop Report on Information Seeking Support Systems

Participants

James Allan, Univ. of Massachusetts
Peter Anick, Yahoo
Anne Aula, Google
Nick Belkin, Rutgers Univ.
Peter Brusilovsky, Univ. of Pittsburgh
Robert Capra, UNC
Ed Chi, PARC
Susan Dumais, Microsoft
Efthimis Efthimiadis, Univ. of Washington
Gary Geisler, Univ. of Texas
Gene Golovchinsky, FxPAL
Steve Griffin, NSF
Brad Hemminger, UNC
Diane Kelly, UNC
Bill Kules, Catholic Univ. of America
Michael Levi, U.S. Bureau of Labor Statistics
Gary Marchionini, UNC
David Millen, IBM
Jan Pedersen, Yahoo
Peter Pirolli, PARC
Edie Rasmussen, Univ. of British Columbia
Dan Rose, A9
Dan Russell, Google
m.c. Schraefel, Univ. of Southampton
Ben Shneiderman, Univ. of Maryland
Elaine Toms, Dalhousie
Daniel Tunkelang, Endeca
Ryen White, Microsoft

Acknowledgements

This workshop was supported by a grant from the National Science Foundation (#115-0836616)

The workshop organizers greatly appreciate…

Maria Zemankova, our National Science Foundation Program Manager, and Steve Griffin who participated in the workshop
Rachael Clemens, for administering and coordinating logistics the workshop
The School of Information and Library Science and UNC Chapel Hill for hosting the workshop and providing meeting rooms.

Table of Contents

Executive Summary

This report defines research and development challenges related to Information Seeking Support Systems. It includes papers and examples from a NSF-sponsored workshop held in Chapel Hill, North Carolina in June 2008 and subsequent papers published as a special issue of *Computer* in March 2009.

The adoption of web search technologies have led people to expect immediate and easy access to information in all aspects of life. Web-based search services have become fundamental components of the cyberinfrastructure that supports economic growth and social advance. Although we have built a base set of expectations for search, human needs for information go beyond search to filtering, assessing, sensemaking, synthesizing, and using information to meet our needs and create new knowledge. Our nation and our world depend on citizens who are able to seek, assess, understand, and use diverse kinds of information. Much of the information we need is complex with different components held in disparate electronic sources and many of our efforts to gather, assess, and use this information are done in collaboration with others. Additionally, much of the information we need is not discretely anticipated, but rather emerges as seeking and reflection continues over time. Information seeking in the digital age is a kind of problem solving activity that demands agile and symbiotic coordination of human and cyber resources; in short, a fundamental kind of computationally-augmented thinking. Computation has expanded our ability to do scalable what if thinking that leverages the best capabilities of humans and machines to abstract, synthesize, and iterate intellectual actions, and today's search engines are the primitives on the technical side of information seeking. We must rise to the challenge to move information seeking from search engine support that provides discrete items in response to simple queries to tools and services that support reflective and interactive search over time and in collaboration. Tools and services that serve this more comprehensive set of requirements are termed **information seeking support systems**. This report demonstrates some of the challenges such support present and points to a sample of the work that is underway to move us beyond search.

Three kinds of challenges are defined and preliminary steps toward meeting the challenges are presented in this report: robust models of human-information interaction; new tools, techniques, and services to support the full range of information seeking activities; and techniques and methods to evaluate information seeking across communities, platforms, sources, and time. Special attention is given to collaborative information seeking and the need for industry-academic collaboration.

Much broader and intensive efforts on the part of the academy, government, and industry are required if we are to meet the grand challenges of usable and ubiquitous information seeking support systems that empower people to solve problems, create new knowledge, and increase participation in efforts to improve the global human condition. Preliminary efforts as illustrated in this report provide promising directions, however, sustained efforts are urgently needed to support research that leads to understanding information seeking as computationally augmented learning and problem solving, better seamless and ubiquitous systems for supporting information seeking, methods for training people to practice effective and efficient information seeking, and techniques and measures for assessing the tools and practices.

Information Seeking Support Systems

Gary Marchionini and Ryen W. White

The emergence of the World Wide Web (WWW) and various search engines that index its content has made information retrieval a daily activity for most people. Whether at work, school, or play, people have come to expect instant access to information on any topic at any place and time. Retrieval is sufficient when the need is well-defined in the searcher's mind, however, when information is sought for learning, decision making, and other complex mental activities that take place over time, retrieval is necessary but not sufficient. What are required are tools and support services that aid people in managing, analyzing, and sharing sets of retrieved information. The information needs of people grappling with chronic illness, work teams creating services or products, learners studying topics over time, families making long-term plans, scientists investigating complex phenomena, and hobbyists tracking developments over a lifetime are well-served at only the most superficial levels by today's web-based search engines. So called *information seeking support systems* aim to address this shortcoming by providing search solutions that empower users to go beyond single-session lookup tasks. It is during complex and exploratory search scenarios that information seekers require support from systems that extend beyond the provision of search results.

There is increasing attention by the research community and the search engine companies to design and implement systems that meet the broader requirements of these information seekers. This attention is manifested in R&D workshops, papers, and prototypes that use a variety of terms (exploratory search, interactive search, human-computer information retrieval[1]) to describe this more comprehensive information seeking problem and possible solutions. In this report we provide an overview of progress to date and a précis for what is to come as the R&D community moves beyond search alone and focuses on information seeking support systems that help users find, understand, and use information as part of a holistic process that takes place over time and often in collaboration.

Information seeking is a fundamental human activity that provides many of the 'raw materials' for planned behavior, decision making, and the production of new information products. People search for

---

-[1] Exploratory Search Interfaces: Human-Computer Interaction Laboratory 2005 Symposium-Open House, College Park, Maryland, June 2, 2005 http://www.umiacs.umd.edu/~ryen/xsi
-Evaluating Exploratory Search Systems: ACM SIGIR 2006 Conference Workshop, Seattle, Washington, August 10, 2006 http://research.microsoft.com/~ryenw/eess
-Exploratory Search and HCI: Designing and Evaluating Interfaces to Support Exploratory Search Interaction: ACM SIGCHI 2007 Conference Workshop, San Jose, CA, April 29, 2007 http://research.microsoft.com/~ryenw/esi
-Web Information-Seeking and Interaction: ACM SIGIR 2007 Conference Workshop, Amsterdam, July 27, 2007 http://research.microsoft.com/~ryenw/wisi
- Workshop on Human-Computer Interaction and Information Retrieval: HCIR '07, MIT CSAIL, Cambridge, MA, October 23, 2007 http://projects.csail.mit.edu/hcir/web
-Second Workshop on Human-Computer Interaction and Information Retrieval: HCIR 08, Microsoft Research, Redmond, WA http://research.microsoft.com/en-us/um/people/ryenw/hcir2008/
-Special issues of ACM Communications of the ACM (April 2006); Information Processing & Management (2007), and IEEE Computer (March 2009).

information objects that embody ideas, use cognitive effort to understand what they find, and employ additional effort to use these understandings to create problem solutions. Marchionini (1995) characterizes the information seeking process with seven subprocesses (recognize need, accept problem, formulate the problem, express the need, examine results, reformulate the problem, and transition to use) that recur according to the information seeker's ongoing sense making and reflective monitoring of progress. Pirolli and Card's information foraging theory (1999) models these processes as highly adaptive to the information environment. These views of information seeking admit ranges of needs from known item instances such as locating the phone number to order a pizza to life-long efforts to conduct research in a rapidly changing domain. They also admit the possibility that people reject the problem at any iteration, either through avoidance of traumatic information or due to feelings of information overload as more information is discovered.

Additionally, the combination of our natural inclinations toward gathering information to inform behavior and the computational tools that underlie the WWW broadens the trend toward computational thinking in everyday life. The WWW and associated search tools have accelerated our capabilities for information seeking, brought this activity front and center as a skill that all literate individuals must have, and raised expectations dramatically about access and use of information in all aspects of our lives.

In addition to having transformational impact on the roles of major information services such as newspapers and directory services such as yellow pages, this ready access to facts, images, and documents realizes an important element of computing as augmentation of the intellect envisioned by Doug Engelbart 40 years ago. Today's search engines combined with mobile access to the web are augmenting our memories, presumably freeing us to focus more mental effort on interpreting and using information to learn, make decisions, and create new knowledge.

The easier access to information becomes, the greater become our expectations for ubiquitous access in all kinds of situations. The snowballing effect of ubiquitous access to information, coupled with the expected growth in the range of search task types being attempted, brings new challenges to information seeking theory and the design of information seeking support systems. There are three mutually interdependent aspects of this challenge: more robust models of human-information interaction; new tools, techniques, and services to meet the expanding expectations in every-increasingly comprehensive information problem spaces; and better techniques and methods to evaluate information seeking across communities, platforms, sources, and time. This dynamic view of information seeking, where people, information sources, and systems change continually and search becomes a seamless component of computationally augmented life rather than a discrete activity requires new models, tools, and measures. The papers in this report represent some of the current approaches to these emergent requirements

The workshop was organized around the three challenges: theoretical models of ISSS, tools and techniques for supporting information seeking, and evaluating information seeking processes. Additionally, we aimed to address a translational science approach to ISSS research and development by balancing participation between academic and corporate researchers and discussing ways that corporate, government, and academic researchers might collaborate. Before the workshop, participants were asked to submit short position papers and an agenda was organized with breakout sessions for each of the three challenges. Based on discussions at the workshop and the position papers, a special issue of *IEEE Computer* was proposed and papers solicited. The special issue appeared in March 2009 and included papers or sidebars that included 20 different authors, 16 of whom were workshop participants. IEEE agreed that rights for the published papers and sidebars would be shared with authors and could be included as part of this workshop report. Discussions during and after the workshop and the papers submitted to the special journal issue illustrated the growing importance of search done in collaboration and the influence of Web 2.0 technologies on information seeking. This report is therefore an aggregation of the *Computer* papers and the position papers submitted for the workshop.

The models section includes Pirolli's *Computer* paper that calls for rational and predictive models of information seeking that operate across time and space. Adopting utility-based models from ecology (foraging) and psychology (sensemaking), he presents a model that predicts human behavior for specific information seeking conditions. Six participant position papers are also included in this section that present different theoretical views on information seeking as a process and key challenges to research and development.

The section on Social Search includes two papers and one sidebar from the *Computer* issue (Chi; Golovchinsky, Qvarfordt, & Pickens; and Millen) as well as Millen's position paper. These contributions provide examples of social and collaborative search, and new directions for leveraging the collective search experience and social network technologies during information seeking.

The section on Tools and Techniques includes a broad range of systems and ideas for supporting information seeking. First, White provides a summary of the kinds of themes that emerged during the workshop. schraefel's *Computer* paper puts today's tools in the historical context of pre-web computation and presents examples of faceted search tools that tightly couple selections and results and add visual representations that support highly interactive information seeking. Two *Computer* issue sidebars and eight position papers are also included, illustrating the range of systems under development and the rich possibilities for new kinds of systems and services.

The section on Evaluation reflects the extensive discussion at the workshop about the complexity of human behavior over time and the tight coupling of tools and human acts during information seeking that makes it difficult to separate user and system effects, thus making our classical metrics such as time, mouseclicks, or precision too granular to alone assess the effects of our tools for effectiveness, efficiency, and satisfaction. The section includes the *Computer* issue paper by Kelly, Dumais, and Pederson who provide an overview of evaluation strategies that aim to get inside the black box between query and result. They raise the possibility of distributed evaluation in a living laboratory that allows the Web populace to be both users and investigators in service to the greater good of all. Three *Computer* issue sidebars and one position paper are also included that address specific aspects of the evaluation challenge.

Russell's paper on industry-academic collaboration from the *Computer* issue highlights the importance of collaborations among corporate and academic scientists. For example, search engine companies have enormous amounts of proprietary data at their disposal and academic teams are equipped to create and apply novel approaches to analyzing and interpreting these data. These collaborations are a meta example of the way the web leverages social systems (e.g. hyperlinks between documents, tags and annotation, usage-based recommendations) and the results of such collaborations promise to propel research and development ahead.

The report concludes with a brief call to action that encourages researchers to take up the challenges defined by the workshop and academic, corporate, and government institutions to advocate and support this research and development. Increasing the power of search systems and the improving availability of information can create an informed citizenry and help drive the information economy. When we consider information rather than material resources, the tragedy of the commons becomes the treasures of the commons. We may all be consumers of information, but we will never deplete this resource that is the fruit of human imagination and intellectual effort. The development and widespread adoption of WWW-based information and associated search engines has opened the door to new kinds of systems that are more powerful and integral to the full range of information seeking activities. Research and development on these information seeking support systems will help support our increasingly-complex information needs and desire for anytime, anywhere information access and use.

References

Clark, A. (1997). *Being there: Putting brain, body, and the world together again*. Cambridge, MA: MIT Press.

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

Pirolli, P. and Card, S.K. (1999). Information foraging. *Psychological Review*, 106 (4), 643-675.

Biographies

Gary Marchionini (march@ils.unc.edu) is the Cary C. Boshamer Professor in the School of Information and Library Science at the University of North Carolina at Chapel Hill. His research interests include human information interaction theory and the design of systems that support information seeking. He received a PhD in mathematics education from Wayne State University.

Ryen White (ryenw@microsoft.com) is a researcher in the Text Mining, Search, and Navigation Group at Microsoft Research. His research interests include exploratory search and the evaluation of search systems with humans and simulations. He received a PhD in Computer Science from the University of Glasgow, United Kingdom.

Powers of 10: Modeling Complex Information Seeking Systems at Multiple Scales

Peter Pirolli
Palo Alto Research Center

Introduction

These are exciting times for scientists, designers, and engineers in the field of information seeking support systems (ISSSs).[1] New technology continues to fuel a staggering growth of available information, which in turn affects our lives by providing resources for adapting to an increasingly complex world, as well as new ways of being entertained. We are witnessing an efflorescence of new ways of interacting with--and producing--rich content. National efforts, such as the U.S. Cyberinfrastucture initiative[2] are aimed at producing even more fertile platforms for information. This evolving domain is a great opportunity for science, because there is so much new territory to explore and explain, so many new ideas about how to do so, and so much potential for having an impact on innovative engineering and imaginative design.

This paper discusses two of the exciting opportunities facing science and engineering in this field. The first is that we are moving from prescientific conceptual frameworks about information seeking to more rigorous scientific theories and predictive models. Progress in cognitive science and human computer interaction is moving towards a coherent set of theories and models to address the complexity of modern-day information seeking at a range of scales from the long-term social down to individual moment-by-moment interaction. The second opportunity is the expansion of the kinds of things we study and develop. Information seeking in the current world involves much more than isolated solitary users working with a single tool to retrieve some document or fact. The information environment has become a place to explore and learn over longer periods of time. It has become much more social. People use many tools and systems fluidly for many purposes. Information is no longer just passive text, but includes rich media that often is seeking users as much as the other way around.

[1] The definition of the ISSS field that I assume in this article is "The study of ISSSs and the phenomena that surround them."

[2] See the National Science special report: http://www.nsf.gov/news/special_reports/cyber/index.jsp

Classic Conceptual Frameworks

There has been a long tradition of researchers developing *conceptual frameworks* (often called *conceptual models*) of ISSSs that provide a general pool of concepts, processes, assumptions, methods, and heuristics that orient researchers to a particular way of viewing and describing the world. For instance, one classic conceptual framework for information retrieval is depicted in Figure 1 [11]. A user has an information need that they must reformulate as a query to a system, which in turn retrieves a document whose system representation best matches, the query. This classic information retrieval model motivated a wave of research crested with the search engine technology that is pervasive on the Web, which addresses a large class of everyday information needs for hundreds of millions of users.

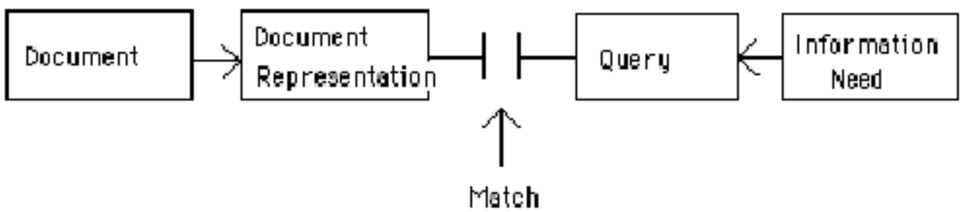


FIGURE 1: The Classic Information Retrieval Model

However, it also became clear throughout the 1980s and 1990s that information seeking systems and behavior must include more than just query-based search. For instance, browsing in hypermedia systems, such as the Web, were not captured by Figure 1. The classic model did not address more complex sensemaking activities in which large amounts of information about a situation or topic are collected and deliberated upon in order to form an understanding that becomes the basis for problem solving and action. Researchers studying real user behavior and systems began to expand on the classic conceptual framework in Figure 1. For instance, analysis of real-world information seeking [8] lead to the identification of an expanded set of general processes and factors involved in information seeking behavior. Conceptually, an information seeker is seen as situated in some environmental context with an embedding task or goal that triggers a need for more knowledge, Information systems provide interactive access to information (that must be shaped into knowledge) to meet those needs and achieve those tasks. The information provided by the system enhances the ability of people to find the right information for their purposes. This expanded conceptual framework leads to the identification of many factors that can shape information seeking, including factors associated with

- *Information seekers*, including prior knowledge, skills, and other individual differences

- *Tasks*, which can vary greatly in terms of complexity, structure, and constraints, which drive and shape information-seeking and problem-solving behaviors, and provide criteria by which information is evaluated.
- *Domains*, which are the general bodies of knowledge that can vary in complexity, kinds of content, rate of growth, change, and organization, among other things.
- *Contexts*, which can include physical, social, and temporal factors and constraints.
- *Search systems*, which can vary enormously in how domain information is represented and how they present and interact with users.

Information seeking as *berrypicking* [3] also become an influential metaphor and conceptual framework. Users often start with some vague information need and iteratively seek and select bits of information that cause the information need and behavior to evolve over time; there is no straight line of behavior to a single best query and retrieval set. In real libraries, users were observed to employ a wide variety of information navigation strategies, such as footnote chasing, citation chaining, reviewing a journal series, browsing entire areas at different levels of generality, and browsing and summarizing works by author. Today, all of these techniques seem mundane, which is a testament to how real-world information seeking strategies have come to be better supported by specific system features and user interface designs.

Theories and Models Across Ten Powers of Ten of Time Scale

So, conceptual frameworks have proven to be productive. They provide conceptual tools for analyzing and describing observed behavior, which in turn can suggest new functions for information seeking systems. They provide a common ground for summarizing findings and accumulating results, for formulating hypotheses and analyses, and for contrasting and debating different ways of charactering ISSS systems.

However conceptual models are not scientific *theories* or *models* that provide a basis for making predictions about the effects of design and engineering decisions on information seeking support systems. Scientific theories are constructed within frameworks by providing additional assumptions that allow researchers to make predictions that can be tested. Typically, this is achieved by specifying a *model* for a specific situation or class of situations that makes precise predictions that can be fit to observation and measurement.

Two exciting challenges are (1) developing truly predictive and explanatory scientific theories and models and (2) developing them in a way that addresses the full complexity of information seeking behavior at multiple time scales permitting, for instance, the prediction of how minute changes at the micro-scale of individual users interacting can percolate upwards to emergent macro-scale phenomena such as the evolution of wikis and tagging systems. Predictive models can provide a basis for understanding and control over ISSS systems and the behavior they support. In practical terms, it means that designer and engineers can explore and explain the effects of different ISSS design decisions before the heavy investment of resources for implementation and testing. This exploration of design space will become more efficient and innovative.

Time Scales of Human-Information Interaction Behavior

To get a sense of what may be possible we can consider the hierarchical organization of human behavior and the phenomena that emerge at different time scales of analysis (Table 1) Newell and Card [9] argued that human behavior (including information seeking) could be viewed as the result of a hierarchically organized set of systems rooted in physics and biology at one end of the spectrum and large-scale social and cultural phenomena at the other end. This framework (see Table 1) has been useful in cognitive science [1] and human computer interaction [9]. The basic time scale of operation of each system level in this hierarchy increases by approximately a factor of 10 as one moves up the hierarchy (Table 1). Some behaviors, such as information seeking on Web, can be modeled at multiple time scales. However, the most exciting work on developing models that map out this territory for different kinds of ISSSs and phenomena has yet to begin.

Table 1
Time-scales on which human behavior occurs. Different bands are different phenomenological worlds.

Scale (seconds)	Time Unit	Band
10^7	months	
10^6	weeks	**Social**
10^5	days	
10^4	hours	
10^3	10 minutes	**Rational**
10^2	minutes	
10^1	10 seconds	
10^0	1 second	**Cognitive**
10^{-1}	100 ms	
10^{-2}	1 ms	**Biological**

Within the time frames in Table, there are layered sets of bands at which behaviors are largely dominated by different kinds of factors:

- The *biological band* (approximately milliseconds to tens of milliseconds) phenomena are mainly determined by biochemical, biophysical, and especially neural processes, such as the time it takes for a neuron to fire.
- The *psychological band* of activity (approximately hundreds of milliseconds to tens of seconds) where elementary psychological machinery for perception, cognition, and action play a major part in shaping behavior. This has traditionally been the domain of cognitive psychology.
- The *rational band* of phenomena (approximately tens of seconds to minutes to days) , where it is the structure of the task and other environmental and contextual constraints come to dominate the linking of actions to goals. Longer-term goals are typically realized by task structures that are hierarchically composed of many shorter-term goals. There is a tendency for individuals to approximate a principle of rationality: based on their perceptions of the ongoing situation and their current knowledge they prefer actions that will move them towards their goals.
- The *social band* (approximately days to weeks to months and beyond) involves systems that are social, involving many individuals in communication and interaction. At this level, factors such as communication, coordination, cooperation, trust, reputation, etc. play a role as well as the structure and dynamics of social networks.

A Simple Example: Navigation Choices on the Web

Consider information-seeking on the Web where accurate models [6] at the rational, psychological, and social bands have been developed for simple tasks such as finding products, event dates, particular documents etc. These have even been incorporated into an automated Web usability evaluator called Bloodhound [4]. Analysis at the rational band of Web information seeking involves a method called *rational analysis* [2]. Rational analysis focuses on the task environment that is the aim of performance, the information environment that structures access to valuable knowledge, and the adaptive fit of the human-information interaction system to the demands of these environments. The recipe for this method is:

1. Specify the information-seeking goals of the user.
2. Develop a formal model of the task environment and information environment (e.g., the information architecture of the Web)
3. Make minimal assumptions about cognitive costs.
4. Derive the rational (optimal) behavior of the user considering (1) – (3)
5. Test the rational predictions against data.
6. Iterate.

In the case of the Web tasks discussed above, one focus of rational analysis concerned how users choose the most cost-effective and useful browsing actions to take based on the relation of a user's information need to the perceived cues associated with Web links. Such user interface cues have been called *information scent* because users follow these cues on trails through information architecture. To an approximation, link choices can be modelled by a Random Utility Model [10]: The probability that a user will choose a particular link L, having a perceived utility V_L, from a set of links C on a Web page, given a user information goal, G is,

$$\Pr(L \mid G, C) = \frac{e^{\mu V_L}}{\sum_{k \in C} e^{\mu V_k}} \quad (1)$$

where μ is a scaling parameter.

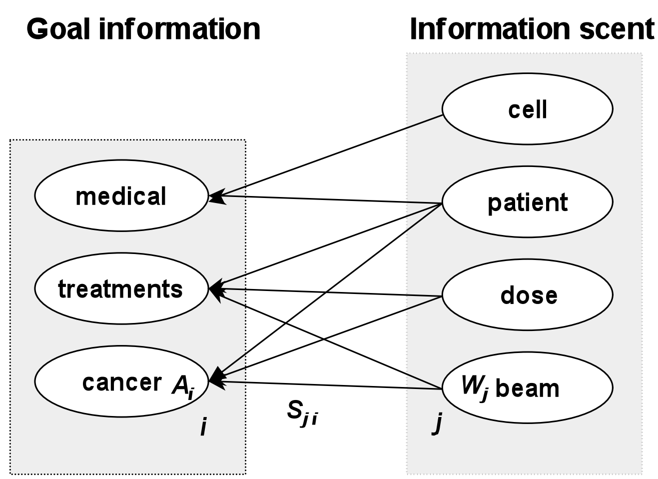


Figure 2. A cognitive structure in which cognitive chunks representing an information goal are associated with chunks representing information scent cues from a Web link.

The details of how users judge the utility of Web links (e.g, V_L) can be modelled at the finer-grained psychological band (for text links) by a *spreading activation model of information scent*. This model assumes that the user's cognitive system represents information scent cues and information goals in cognitive structures called *chunks* (Figure 2). These chunks can be thought of as representations of mental concepts in human memory. Figure 2 assumes that a user has the goal of finding information about "medical treatments for cancer," and encounters a Web link labelled with the text that includes "cell", "patient", "dose", and "beam". It is assumed that when a user focuses attention on a Web link their attention to information scent cues *activates* corresponding cognitive chunks into conscious awareness. *Activation* spreads from those attended chunks along associations to related chunks. The amount of activation accumulating on the representation of a user's information goal provides an indicator of the utility (e.g., V_L) of that link. When incorporated into a computational simulation of users, the rational band and psychological band models can predict up to 90% of the variance in link choice behavior of users [6].

These models of Web link choice can be mapped up to a social band model that simulates the flow of users through a Web site. This provides the core algorithm used by the Bloodhound Web usability systems, [4]. A flow network can model the aggregate flow of a group of users for a particular task where each node in the network represents a Web page, and each edge in the network represents the flow of users from one page to another. A simple version employs discrete steps (corresponding to users clicking to move from one page to another). The number of users arriving at Web page node *i* at step $t + 1$ from the *k* nodes from which it is reachable is modelled as,

$$N_{i,t+1} = f_t \sum_k S_{i,k} N_{k,t} \qquad (2)$$

where f_t is the proportion of users who continue to surf after *t* time steps, and S_{ji} is the proportion of users who decide to choose to go to page *j* from page *i* (which can be determined by Equation 1 above). Tests of this algorithm when incorporated into the Bloodhound usability evaluator showed that the predicted pattern of visits showed moderate to strong correlations with observed patterns in 29 out of 32 tasks conducted over four different Web sites.

Three Theses About the Modeling of ISSSs

This simple illustration about modeling the Web also demonstrates three theses [1] that promise a future science of ISSSs,:

- The *decomposition thesis*, which states that complex IS behavior occurring at large time scales from minutes to hours to days to months can be decomposed into smaller elements and events. For instance, the complex behavior of an individual interacting with a Web browser can be decomposed into individual elements of perception, judgment, decision, and action selection. Decomposition of the task of Web information seeking (of which link choice is just one subtask) is required to develop the full user simulations that make the prediction in Figure 1 Specific features of ISSSs can improve or degrade those elements in a way that has an impact on full shape of IS behavior in complex situations.

- The *relevance thesis* is that the microstructure of behavior is relevant to macrostructure phenomena. For instance, small perturbations in the quality of information scent can cause qualitative shifts in the search cost of Web navigation. There is also evidence [12] that changes in the time cost of fine-grained user interaction in information rating systems and social bookmarking systems has reliable effects at the social band on contribution rates in user populations. Tweaking the fine-grained structure of a user interface can have effects on the phenomena that emerge at the level of large social groups.
- The *modeling thesis*, which is the claim that predictive models can be developed to specify precisely how changes at the microstructure of individuals interacting with and through ISSS can percolate upwards to have effects on longer-term complex information seeking. In some cases, a single unified model can go from the lower bands to higher. For instance, the SNIF-ACT model [6] does this. More likely, however, there will be layers of models at different bands that capture essential features of models at the lower bands, just as statistical mechanics models particles at a level that only captures crucial features of the behavior of individual particles. The graph flow models in Bloodhound, for instance, only capture average or asymptotic behavior of many individual users interacting with the Web.

Enriching the Class of IS Tasks

As the Internet and communication technologies become ever more pervasive we see an astounding number of new ISSSs and behaviors. As a result, we continually need to expand the kinds of phenomena to be addressed by conceptual frameworks, theories, and models. I will just touch on a few interrelated conceptual frameworks as illustration of the expanding territory.

Sensemaking

Many information search tasks are part of a broader class of tasks called *sensemaking*. Such tasks involve finding and collecting information from large information collections, organizing and understanding that information, and producing some product, such as a briefing or actionable decision. Examples of such tasks include understanding a health problem in order to make a medical decision, forecasting the weather, or deciding which laptop to buy. In general, these tasks include subtasks that involve information search, but they also involve structuring content into a form that can be used effectively and efficiently in some task.

As an extreme example, intelligence analysts, do this sort of sensemaking as a profession, working to gather and make sense of vast amounts of incoming information in order to write briefings that shape decisions that affect national security. Cognitive task analyses of intelligence analysis [10] suggest that the overall process is organized into two major loops of activity: (1) an *information foraging loop* that involves processes aimed at seeking information, searching and filtering it, and reading and extracting information, and (2) a *sensemaking loop* that involves iterative development of a mental model (a conceptualization) that best fits the evidence. Information processing is both driven by *bottom-up* processes (from data to theory) and *top-down* (from theory to data). The foraging loop is essentially a tradeoff among three kinds of processes: information exploration, information enrichment, and information exploitation (e.g., reading). Typically, analysts cannot explore all of the space, and must forego coverage in order to actually enrich and exploit the information. The sensemaking loop involves substantial problem structuring (the generation, exploration, and management of hypotheses), evidentiary reasoning (marshalling evidence to support or disconfirm hypotheses), and decision making (choosing a prediction or course of action from the set of alternatives). These processes are affected by many well-known cognitive limitations and biases.

Exploratory Search

Another rich domain is *exploratory search* [7]. Exploratory search includes activities involving information lookup, leaning, and investigation that may overlap in time. For instance, looking for health-related information is one of the most prevalent information seeking activities on the Web. Typically this involves a prolonged engagement in which individuals iteratively look up and learn new concepts and facts and dynamically change and refine their information goals and come to ask better questions (it can be viewed as a subcomponent of sensemaking, above). Three major kinds of activities are involved in exploratory search: lookup, learn, and investigate. Whereas lookup activities have been the traditional focus of ISSSs (e.g., Figure 1), exploratory search emphases the learning and investigation activities. Searching to learn includes the sort of activities involved in making decisions (e.g., purchases) up through professional and life-long learning. It also includes social search to find communities of interest (for instance via social network systems). Investigative activities include analysis, synthesis, and evaluation activities.

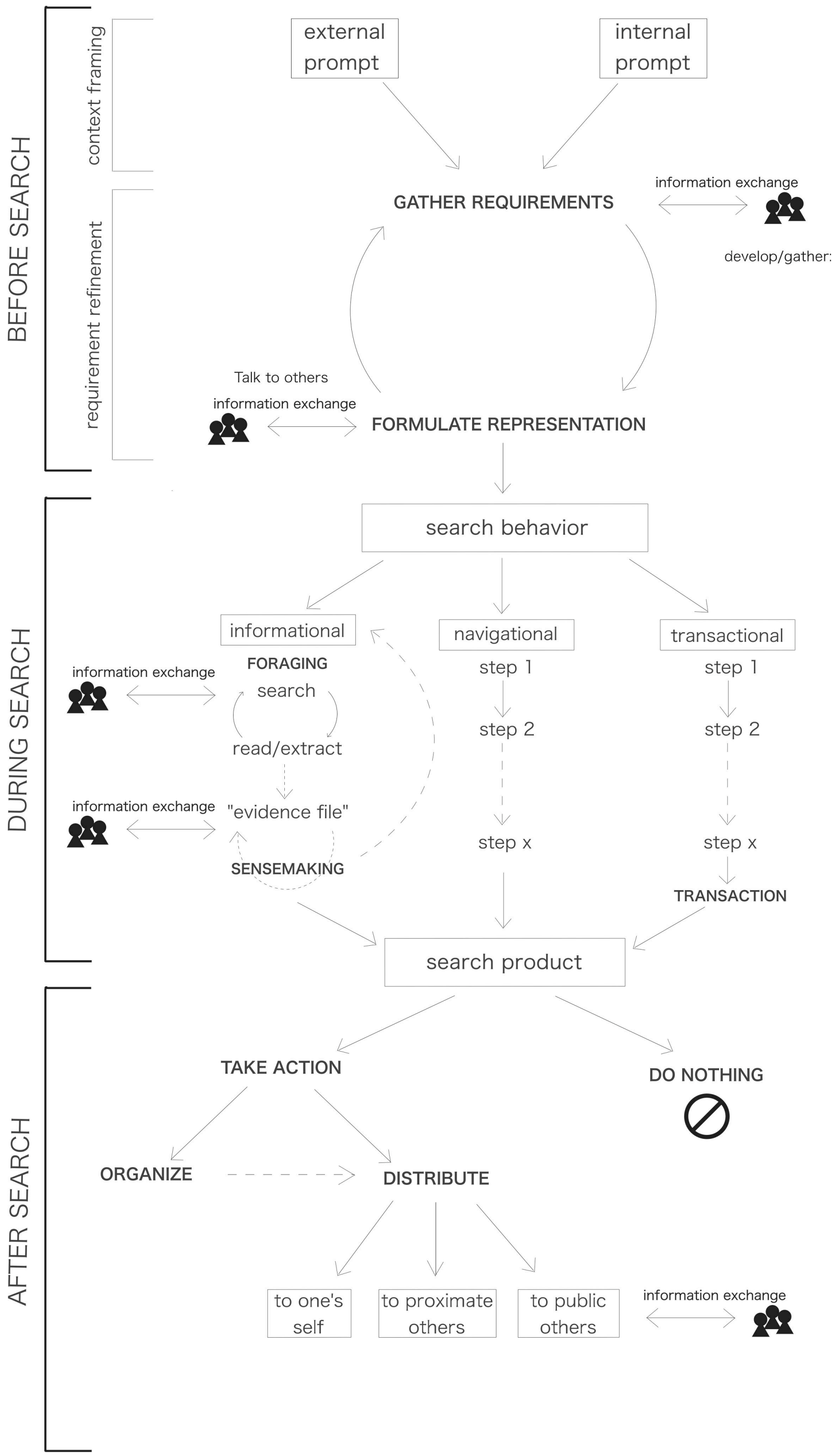


Figure 6. A conceptual model of social search (Adapted from [5]).

Social search

People frequently turn to their social networks to find information (see the Chi article this issue). Social search has become a recent focus of study [5] . Figure 6 shows some of the complexity that arises in analysis of social search. Social exchanges of information can occur before, during, and after an actual search for information. Social transactions can influence the gathering of requirements and formulation of a problem representation (before a search), the gathering and extraction of information (during search) and the distribution of results (after search).

Even more generally, there are new forms of socio-technical information systems, such as social bookmarking or rating sites that allow users to participate at low effort and contribute their unique nuggets of knowledge (e.g., about navigation, organization, recommendations) in a highly independent modular way, and these contributions improve the performance of the system as a whole. The emergence of such Web 2,.0 technologies shows how systems that can efficiently allow users to make contributions, and that have architectures that harness those contributions in a highly efficient way, are ones that can win big in the public an commercial world.

Discussion

Our conception of ISSSs has expanded from simple information search engines, and similarly the field is moving from conceptual frameworks to scientific models, and expanding the range of systems and phenomena that are studied. The field is moving away from a focus on simple precision/recall metrics (see Tunkelang sidebar) to a more comprehensive understanding of the utility of information to the range of human goals in modern day society, the need to better understand user experience, and user evaluations of credibility, trust, and reputation The field is realizing that we need to understand why people search, explore, annotate, and decide to participate and share the efforts of their knowledge work. There is a realization that people no longer work with a single ISSS, but even for simple tasks will effortlessly move among an array of tools. This is a complex territory to map out with scientific models. It spans 10 powers of ten of time scale, from tens of milliseconds to years, with enormous complexity at multiple bands of phenomena from the psychological to the social. One of the great opportunities is that these phenomena are increasingly happening in large living laboratories online (see Kelly et al. article this paper), and this a great attractor for scientific minds in diverse areas ranging from behavioral economics, incentive mechanisms, network theory, cognitive science, and human computer interaction, just to name a few.

References

1. Anderson, J.R. Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science*, *26* (1) 2002, 85-112.

2. Anderson, J.R. *The adaptive character of thought*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1990.
3. Bates, M.J. The design of browsing and berrypicking techniques for the online search interface. *13* 1989, 407-424.
4. Chi, E.H., Rosien, A., Suppattanasiri, G., Williams, A., Royer, C., Chow, C., Robles, E., Dalal, B., Chen, J. and Cousins, S. The Bloodhound Project: Automating discovery of Web usability issues using the InfoScent simulator. *CHI 2003, ACM Conference on Human Factors in Computing Systems, CHI Letters*, *5* (1) 2003, 505-512.
5. Evans, B.M. and Chi, E.H. Towards a model of understanding social search *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, ACM, San Diego, CA, USA, 2008.
6. Fu, W. and Pirolli, P. SNIF-ACT: A model of user navigation on the World Wide Web. *Human Computer Interaction*, *22* (4) 2007, 355-412.
7. Marchionini, G. Exploratory search: from finding to understanding. *Communications of the ACM*, *49* (4) 2006, 41-46.
8. Marchionini, G. *Information seeking in electronic environments*. Cambridge University Press, Cambridge, England, 1995.
9. Newell, A. and Card, S.K. The prospects for a psychological science in human-computer interactions. *Human-Computer Interaction*, *2* 1985, 251-267.
10. Pirolli, P. *Information foraging: A theory of adaptive interaction with information*. Oxford University Press, New York, 2007.
11. Robertson, S.E. Theories and models in information retrieval. *Journal of Documentation*, *33* (2) 1977, 126-148.
12. Wilkinson, D.M. Strong regularities in online peer production *Proceedings of the 9th ACM conference on Electronic commerce*, ACM, Chicago, Il, USA, 2008.

Complex and Exploratory Web Search

Anne Aula
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
1-650-253-8975
anneaula@google.com

Daniel M. Russell
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
1-650-253-2682
drussell@google.com

ABSTRACT

We suggest that there are two undiscussed dimensions to information searches: *complexity of the information need* and *clarity of what is needed to satisfy the information need*. Among the factors that affect the complexity of the information need is the amount of information that is needed and the number of steps that are required to collect the needed information. In addition, storing found information becomes increasingly important as the complexity of the task increases (too much information to hold in memory). Clarity of the goal refers to the searcher's understanding on the pieces of information that are needed to satisfy the information need. We show how *exploratory search* may sometimes be complex, but is not necessarily so, and is characterized more accurately by the degree of clarity the searcher has about the goal. *Complex search* tasks often include exploring the topic, but do not necessarily require exploration or may require exploration only in certain phases of the search process. We suggest that complex search tasks – especially those where the goal-state is initially unclear - are the ones where the current search tools do not offer enough support for the users.

Keywords

web search, complex search, exploratory search

1. INTRODUCTION

Over the years, a number of different taxonomies for classifying information seeking tasks have been proposed. For example, Rose and Levinson [6] proposed a taxonomy with three different search goals (navigational, informational, resource). For informational tasks, they included five sub-categories: directed (closed vs. open), undirected, advice, locate, and list. Similarly, Kellar et al. [2] proposed a web information classification with six different information tasks (fact finding, information gathering, browsing, transactions, communication, maintenance). In Rose and Levinson's taxonomy, undirected informational tasks and in Kellar's classification, information gathering are close to what other researchers sometimes call exploratory search tasks. [4][1]

Marchioni [4] grouped search activities into lookup, learn and investigate activities. According to Marchioni, exploratory search is especially pertinent to the learn and investigate activities. White et al. [1] propose that exploratory search tasks often require the users to submit a tentative query and use the results to find cues about the next steps. "Defining what constitutes an exploratory search is challenging…in exploratory search, users generally combine querying and browsing strategies to foster learning and investigation" [p.38].

The term "exploratory search" is generally used to refer to the challenging search tasks where the user's goal is to learn something by collecting information from multiple sources. We propose a novel classification whereby we differentiate between "exploratory search" and "complex search" by their relative levels of search goal complexity and search goal clarity. In our classification, exploration occurs when the searcher is unsure about what to do next, or has to learn more about the domain of interest. Exploratory search is often an intrinsic part of complex search tasks—but exploration is usually not the end goal—it is a property of the state of knowledge on the part of the searcher.

2. WHAT MAKES A SEARCH TASK COMPLEX?

We suggest *complex search task* as a term to refer to cases

- where the searcher needs to conduct multiple searches to locate the information sources needed,
- where completing the search task may span over multiple sessions (task is interrupted by other things),
- where the searcher needs to consult multiple sources of information (all the information is not available from one source, e.g., a book, a webpage, a friend),
- that often require a combination of exploration and more directed information finding activities,
- which often require note-taking (cannot hold all the information that is needed to satisfy the final goal in memory), and
- where the specificity of the goal tends to vary during the search process (often starts with exploration).

Earlier research has suggested that the current search tools are optimized for simple lookup or directed search tasks [3] and thus, these tasks are relatively easy to accomplish with the current tools. However, when we look at these simple tasks as parts of a complex search task, using the easy-to-find information to fulfill the overall search goal (that might require tens of little fact-finding tasks) may be very challenging. We have to think of search *tasks*, not individual searches or even search sessions. Individual searches might be relatively more-or-less sophisticated, but we need to analyze search behavior at the task level, not the individual query level. And we know that tasks might be broken up across multiple sessions spanning hours or days. [3]

3. EXPLORATION vs. COMPLEXITY

We propose two measures of search tasks that describe dimensions of user behavior during search.

1. *Goal abstraction level (GAL)*—this is a measure of how close the goal is to actionable behavior. That is, GAL is the distance

from the current level of understanding about the goal to a point where the goal can be executed (in terms of the fundamental domain operators). For example: an information goal of "find the root causes of the US Civil War" requires a decomposition to many discrete actions that can be taken by the searcher. Ultimately, a search is deemed exploratory when the searcher has a very abstract search goal—*i.e.*, when the GAL level is high. When observing the users when they are searching, an abstract (unspecified) goal can often be identified by the users' statements like "I'm interested in learning more about…"

2. *Search moves (SM)*—is an estimate measure of the number of web search steps it will take achieve the goal (i.e., how complex the search task is). That is, SM is a heuristic that can be used to estimate the number of search "moves" it takes to achieve the goal, given the searcher's current state and background knowledge. A move could be a search, or an information extraction subtask, but the idea is that SM is a factor that the searcher takes in to account when considering the next move to make in their search task.

For example, an overall task goal of "Find information that is needed to make the move to Helsinki successful" is too abstract to act on—thus, the searcher needs to decompose the goal into actionable sub-goals. For example, one such sub-goal could be: "find the best neighborhood to live in Helsinki". This sub-goal, in turn, has to be broken down into smaller goals (e.g., "find a listing of Helsinki neighborhoods and reviews/comments on each of them"). For the task of listing the neighborhoods, the goal is concrete and clear (list the neighborhood names and their reviews side-by-side); however the SM estimate (task complexity) could be quite high, including finding sites that list the neighborhoods (possibly on multiple sites), then collecting comments and reviews on each (again, potentially multiple sites, potentially requiring reconciliation between different opinions and reviews).

Clearly, both GAL and SM are heavily influenced by searcher expertise and domain knowledge. A search task such as "Find the root causes of the US Civil War" is a very high-level goal that will require a great deal of work to concretize—for a recent immigrant to the US. But a searcher with the background knowledge of a history undergraduate degree would see this as a much different kind of task.

4. MEMORY LOAD SUPPORT

Task goals that are both complex and exploratory cause large memory and cognitive demands on the searcher As the searcher discovers multiple exploration options and strategies, an effective tactic is to capture those options in some kind note. Similarly, in multi-session searches notetaking often bridges the gap of carrying information learned in one session to the next. This is especially true in search tasks that span multiple days.

In Figure 1 we've tried to capture this as the "Notetaking boundary," when exploration and complexity impose sufficiently large memory loads to require external capture.

We note that many systems have attempted to support searcher memory demands. Bookmarks are common, but flawed mechanism for storing information [1]. A number of notetaking systems (e.g., Google's Notebook product) have been attempted with varying degrees of success. Methods for capturing previously searched queries, context and history (and then making them available to the searcher) have been tried as well. [7]

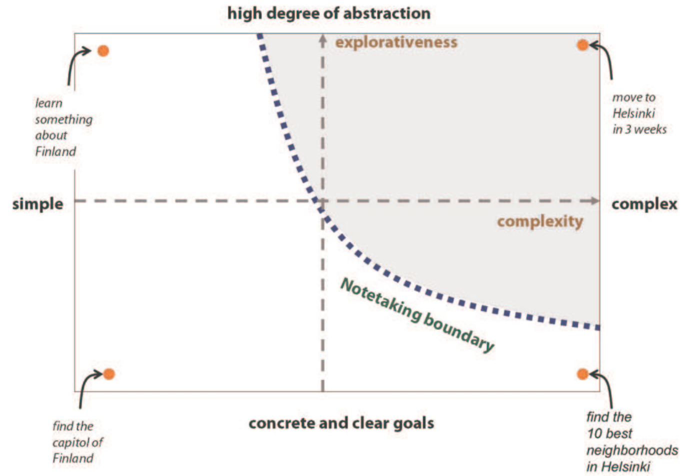


Figure 1: As a search task can vary from simple to complex, depending on the intrinsic complexity of the information being sought and the amount of time it takes to satisfy that information task. But as the searcher works, they may be more or less clear about what they need to do to satisfy their information needs. Exploration occurs when the searcher is unsure about what to do next, or has to learn more about the domain of interest. The Notetaking boundary marks the place where the memory loads of the task (due to both complexity or explorativeness) become so large as to require external memory aids to allow the searcher to succeed.

It seems clear to us that we have yet to identify the effective and powerful methods to support the memory demands that are especially pertinent to searches that are high both in complexity and in the degree of abstraction.

5. REFERENCES

[1] Aula, A., Jhaveri, N., Käki, M. 2005. Information search and re-access strategies of experienced web users. In Proceedings of the 14th international conference on World Wide Web, 583-592.

[2] Kellar, M., Watters, C., & Shepherd, M. (2006). A Goal-based Classification of Web Information Tasks. In the Proceedings of the Annual Meeting of the American Society for Information Science and Technology (ASIS&T 2006), Austin, TX. (CD-ROM Proceedings)

[3] MacKay, B. & Watters, B. (2008) Exploring multi-session web tasks. Proceedings of CHI'2008, 1187-1196.

[4] Marchioni, G. (2006) Exploratory search: from finding to understanding. Communications of the ACM, 49, 4, 41-46.

[5] Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.

[6] Rose, D.E. & Levinson, D. (2004) Understanding user goals in Web search. Proceedings of the WWW 2004, 13-19.

[7] Rose, D. E. and Raju, S. (2007). Encouraging Exploration with Elroy: A New User Experience for Web Search. Proc. of ACM SIGCHI 2007 Workshop on Exploratory Search and HCI (ESI 2007), 52-55.

[8] White, R.W., Kules, B., Drucker, S.M., and Schraefel, M.C. (2006) Supporting exploratory search. Communications of the ACM, 49, 4, 37-39.

Really Supporting Information Seeking: A Position Paper

Nicholas J. Belkin
SCILS, Rutgers University
4 Huntington Street
New Brunswick, NJ 08901, USA
+1 732 932 7500 x8271
nick@belkin.rutgers.edu

1. Introduction

It has been quite clear for quite some time that people engage in a wide variety of interactions with information when engaging in information seeking, whether in multiple information seeking episodes over time, or during the course of a single information seeking episode. Yet systems, both experimental and operational, intended to support information seeking, have been limited, with few exceptions, to the support of only one kind of information seeking strategy (ISS), and only one kind of interaction with information, non-contextualized specified search for particular information objects. In this paper, I argue that truly effective systems for supporting information seeking must be able to support the variety of interactions with information that people engage in, with knowledge of the context and situation which led them to engage in information seeking behavior, and, in the best case, without having to leave their task environment at all. I do this by indicating what I feel are significant challenges to which research in this area should respond.

2. Information-related goals, tasks and intentions

There is substantial and convincing evidence that the goals that lead people to engage in information behavior, the tasks associated with those goals, and with their behaviors, and the intentions underlying the behaviors, substantially affect their judgments of usefulness of information objects, and the ways in which they interact (or would wish to interact) with information objects. The challenges here lie in three spheres.

First, in the ability to characterize and differentiate among information-related goals, tasks and intentions in some principled manners that go beyond straightforward listings, that will apply across a wide variety of contexts, and from which design principles for IR systems can be inferred.

Second, we need to develop methods for inferring information-related goals, tasks and intentions from implicit sources of evidence, such as previous or concurrent behaviors. Without such techniques, any characterization of these factors, or even specification of how to respond to them in system design, is fruitless.

Third, going from characterization and identification of goals, tasks and intentions, to IR techniques which actually respond effectively to them, is a challenge that has been barely noted, much less addressed, to date.

3. Understanding and supporting information behaviors other than specified search

People engage in a wide variety of types of interactions with information, of which specified searching, as represented by, e.g., normal Web search engines, and standard IR models, is only one among many, and perhaps not the most important. For instance, since we know, from a good number of different theoretical and empirical studies, that people have substantial difficulties in specifying what it is (that is, what information objects) that would help them to realize their goals, only considering specified search as the basis for IR models and techniques, is clearly inadequate, and inappropriate. Furthermore, a number of studies have demonstrated that people do engage in a variety of different information behaviors within even a single information seeking episode. However, there is still little known about the nature, variety and relationships of different information behaviors, and the situations that lead people to engage in any one behavior, or sequence of behaviors. Without this basic knowledge, there is little hope for the development of systems which can proactively support multiple behaviors.

4. Characterizing context

The challenges here are obvious, and multiple. We need to have theories and data which lead not only to identification of at least classes of contextual features, but which also tell us in what ways these features do in fact affect information behaviors. We need to have ways to identify appropriate contextual features without explicitly eliciting them. We need to develop IR system techniques which can actually take account of knowledge of these features, individually and in combination, and we need to develop experimental methods and evaluation measures which will allow comparison of the effect of contextually-based interventions.

5. Personalization

The challenge with respect to personalization is first to consider the dimensions along which personalization could and should take place; then to identify the factors or values in each dimension that would inform personalization on that dimension; then to investigate how the different factors or types of evidence interact with one another; and finally, to devise techniques which take account of these results in order to lead to a really personalized experience.

6. Integration into the task environment

Engaging in information interaction in an IR system is almost always a consequence of a person's wanting to achieve some other goal, or accomplish some other task. In this sense, such engagement is inevitably a distraction from that other task. So, 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 in interaction with a separate IR system at all (although I'm quite willing to agree that there are certainly circumstances in which such engagement might indeed be desirable).

7. Other important challenges (in brief)

Taking account of affect;

New evaluation paradigms for interactive information retrieval;

(In)Formal models of interactive information retrieval.

Speaking the Same Language About Exploratory Information Seeking

Bill Kules
School of Library and Information Science, The Catholic University of America
Washington, DC USA
kules@cua.edu

1. INTRODUCTION

"Exploratory search" – little did I realize when I started my dissertation six plus years ago that a term that I naïvely thought my advisor and I had coined (we hadn't) would become a central element of discussion for an NSF-funded workshop. As a computer science Ph.D. student just beginning to learn about the field of library and information science, the notion of exploratory search seemed to capture the kinds of tasks that I was investigating. Of course, this was hardly the only term in use – it wasn't even the most common. As I read more, I realized that there was a broad and deep literature on the topic, but one which contained myriad terms, definitions, and models. The lack of common language presented a challenge as I tried to understand and apply that rich body of knowledge to my own work in faceted search interfaces, and it remains a challenge today.

This paper advocates the adoption of a common term (I suggest "exploratory information seeking"), as well as a common definition and reference model. I make this proposal at the risk of sounding naïve – perhaps it reflects too little knowledge of the rich history of the field; maybe it suggests oversimplifying or "dumbing down" the field to fit a least-common denominator approach; or it may simply echo a consensus already evidenced by the existence of this workshop. But I believe that the field is at a point where these three things are necessary in order to make the research and knowledge accessible to a larger audience.

We need to communicate with multiple audiences: researchers working directly in this field (many of whom are at this workshop); researchers outside of the LIS field (in areas such as computer science, human-computer interaction, psychology, management, and the humanities) who can benefit from and contribute to this work; students who are learning about the field; and the funding agencies, journalists, and general public that need to understand how our individual efforts fit into and advance a larger research agenda. A common framework will also support meaningful dialog about systems and evaluation. This need has been acknowledged at the workshops I have previously attended. In at least two instances, promising discussions were slowed by the lack of a common framework and the need to (re)establish a common understanding among participants.

2. TERMS AND DEFINITIONS

To start, we need to select a term and define it with care. The term should provide an immediate indication of the topic. In addition, we need a short definition that captures the essence (an "elevator speech"), and a long definition that more fully delineates the contexts, situations, tasks and techniques.

The term "exploratory search" is problematic, in part, because of its use of the word "search." Search has a variety of meanings and connotations, only some of which are relevant to the notion of exploratory search as I understand it. Past discussions at workshops have often spent a fair bit of start-up time trying to establish a working definition. They often stumble because there are many behaviors included under exploratory search that are not strictly search – browsing being a common example.

As a starting point for conversation, the term "exploratory information seeking" begins to capture the kinds of work that we are trying to support.

3. TASK MODEL

The second challenge is to develop a common reference model for tasks. This is important for the same reasons that a common definition is needed – we need to communicate effectively within and beyond our community. Of course, no single model will effectively capture the breadth of elements, concepts, variables, and relationships that have been identified. Rather, it should provide a rough "map" of the field that enables researchers to situate and elaborate on their work. As with the short definition, the model should be simple enough to be quickly comprehended.

The TREC conferences provide historical motivation for a common framework. Their system-oriented and quantitative focus has been criticized, but the common framework enabled dialog among researchers, development of tasks for evaluations, and comparison of systems. Over the years, they have carefully defined, refined, and examined specific information retrieval tasks, contributing to the robust development of the IR community. Our challenge is to adapt that idea to exploratory information seeking research without foreclosing on the correspondingly broader set of research methods that are required.

A particular challenge is to develop a taxonomy of exploratory information seeking tasks that span multiple levels of context. Because context and high-level goals are important, the tasks will span a range from individual tactics and actions within a specific interface to high-level scenarios such as writing a research paper or conducting competitive intelligence analysis. Inherent in this challenge is the need to accommodate both quantitative and qualitative methods. It may not be possible to directly compare data between a controlled laboratory study and an ethnographic investigation in the field, but a common understanding of the tasks will help us to triangulate between a varying study designs and compare evaluations in a systematic manner. This in turn will enable meta-evaluations that integrate the results of individual research projects into a larger understanding.

4. A SAMPLE REFERENCE MODEL

This section presents a three-level model of the exploratory information seeking process as a partial illustration of the proposals in this paper (Figure 1). This specific model does not capture the diversity of research interests represented by workshop attendees, but it illustrates a simple framework that can be used for situating discussions and comparing research.

This model combines Marchionini's electronic browsing model (Marchionini, 1995) with the three-level Byström & Hansen model. This model defines five activities: recognize an information need (to satisfy a work task), define an information-seeking problem (to satisfy the information need), formulate query, examine results, and view documents. It places activities in the context of the three levels of information-seeking and work tasks. It shows how search activities are sequenced within the iterative search process. Each higher-level activity can involve multiple subsidiary activities.

Search is a necessary step within a larger information seeking process, the objective of which is to satisfy a perceived information need or problem (Marchionini, 1995). In turn, the perceived information need is motivated and initiated by a higher-level work task (Byström and Hansen, 2002; Ingwersen & Järvelin, 2005) or personally motivated goal (Kari, 2006).

Work tasks are situated in the work organization and reflect organizational culture and social norms, as well as organizational resources and constraints. The work task is similar to Marchionini's recognition and acceptance of an information problem, but the work task specifically situates these in an organizational context. In the context of the work task, a second level of context is defined, in which information-seeking tasks are identified. These tasks vary as the work task progresses. The third level of context is the information retrieval context, wherein searchers identify sources, issue queries, and examine results. Reflection is inherent in each activity, and each activity can return to a previous or higher level activity.

The model is limited in a number of ways. For example, it is focused on work tasks; it considers a limited set of search behavior; and it does not account for non-IR tasks that are part of information seeking. Our challenge is to integrate the many models and theories already available (including Bates, 1972, 1989, 1990; Belkin, 1980; Choo, Detlor & Turnbull, 2000; Dervin & Nilan, 1986; Ellis, 1989; Fidel, 1985; Furnas & Rauch, 1998; Pirolli & Card, 1999, Saracevic, 1996) and summarize them in a manner that is meaningful to multiple audiences.

5. CONCLUSION

This paper proposes the adoption of a common term, definition and reference model for exploratory information seeking. It provides an example of a simple reference model as a straw man for discussion. One inherent challenge is to capture the breadth and depth of the field without oversimplifying it. This will benefit the field by providing a common frame of reference for discussion of systems and evaluation. It will also make the research and findings more accessible to researchers, practitioners, students, and funders – anyone who can benefit from and contribute to the field.

6. REFERENCES

[1] Bates, M. (1990). Where should the person stop and the information search interface start. *Information Processing and Management, 26*(5), 575-591.

[2] Bates, M. J. (1979). Information search tactics. *Journal of the American Society for Information Science, 30*, 205-214.

[3] Bates, M. J. (1989). The design of browsing and berrypicking techniques for the on-line search interface. *Online Review, 13*(5), 407-431.

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

[5] Byström, K., & Hansen, P. (2002). Work tasks as units for analysis in information seeking and retrieval studies. In H. Bruce, R. Fidel, P. Ingwersen & P. Vakkari (Eds.), *Emerging Frameworks and Methods* (pp. 239-251). Greenwood Village, CO: Libraries Unlimited.

[6] Choo, C. W., Detlor, B., & Turnbull, D. (2000). *Web Work: Information Seeking and Knowledge Work on the World Wide Web*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

[7] Dervin, B., & Nilan, M. (1986). Information needs and uses. In M. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 21, pp. 3-33). White Plains, New York: Knowledge Industries.

[8] Ellis, D. (1989). A behavioral model for information retrieval system design. *Journal of Information Science, 15*(4-5), 237-247.

[9] Fidel, R. (1985). Moves in online searching. *Online Review, 9*(1), 61-74.

[10] Furnas, G. W., & Rauch, S. J. (1998). Considerations for information environments and the NaviQue workspace. In *Proceedings of the Third ACM Conference on Digital libraries, Pittsburgh, PA* (pp. 79-88). New York: ACM Press.

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

[12] Kari, J. (2005). Evolutionary information seeking: A case study of personal development and Internet searching. *First Monday, 11*(1). http://www.firstmonday.org/issues/issue11_1/kari/index.html

[13] Kules, B., Wilson, M., schraefel, m.c., Shneiderman, B. (in process) From Keyword Search to Exploration: How Result Visualization Aids Discovery on the Web.

[14] Marchionini, G. (1995). *Information Seeking in Electronic Environments*: Cambridge University Press.

[15] Pirolli, P. L., & Card, S. K. (1999). Information foraging. *Psychological Review*, 106(4), 643-675.

[16] Saracevic, T. (1996). Relevance reconsidered. In *Information science: Integration in perspectives. Proceedings of the Second Conference on Conceptions of Library and Information Science (CoLIS 2).* Copenhagen (Denmark), 14-17 Oct.1996. pp. 201-218.

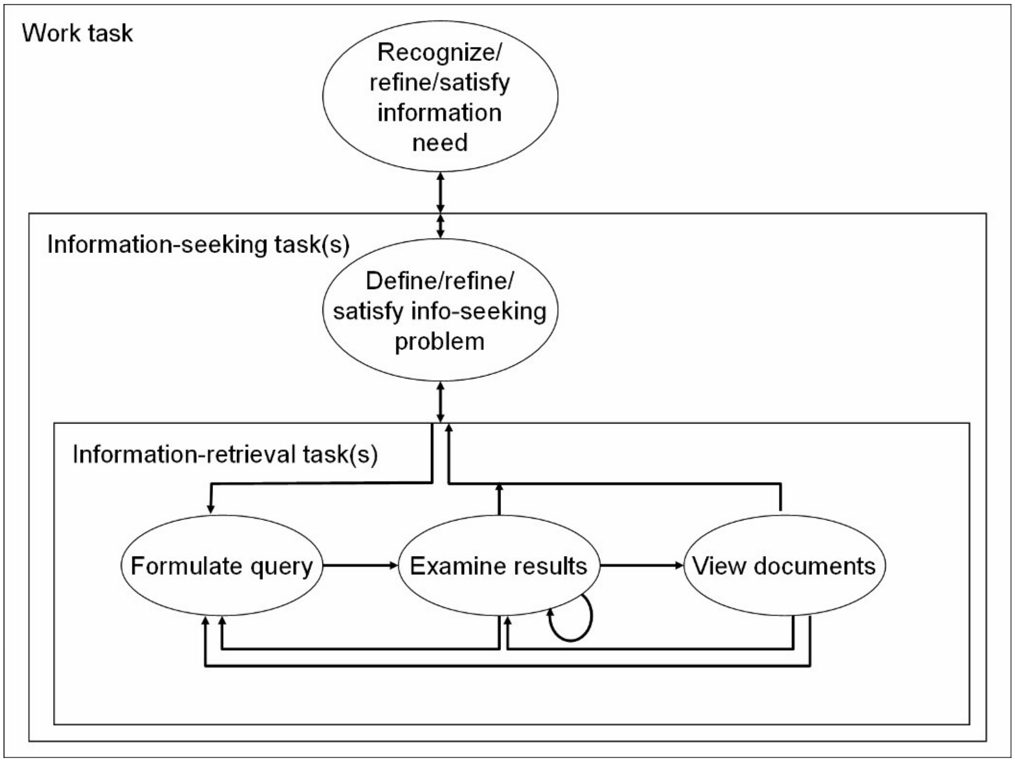


Figure 1. A three-layer model of exploratory information seeking, showing the context of work and information-seeking tasks.

Characterizing, Supporting and Evaluating Exploratory Search

Edie Rasmussen
University of British Columbia
Vancouver, BC, Canada V6T 1Z1
+1 604.827.5486
edie.rasmussen@ubc.ca

1. Characterizing Exploratory Search

In the past 15 years, the Web has completely changed our relationship with information and the ways we organize, access, and use it [1]. In the current environment, information to support any query, musing, idle thought or urgent need, is as near as the nearest keyboard, and the ubiquitous Google toolbar may be used a dozen or more times a day. For many people, search has replaced memory or retention of personal information. We have become so used to this information extravaganza that it is hard to remember a time when this kind of information came from a print encyclopaedia or dictionary, commercial database, or perhaps the reference collection of a public or academic library. In this new environment we have the ability to instantly explore our information world based on any stimulus, major or minor. The barrier to searching for information has effectively disappeared. Where search was once a serious endeavour, undertaken with a significant investment of time, energy and often financial resources, or not at all, it is now casual, periodic and opportunistic. While it is often directed toward a specific outcome, the volume and range of information which is now readily available facilitates, and in fact encourages, search which is exploratory in nature.

We are aware that this change in information behaviour has taken place in people's daily lives and work environments, but attempts to model it are in their infancy. What motivates individuals to search, what collections do they access, what strategies do they employ, how much time do they invest, how successful are they, what use do they make of the information, what roles do collaboration and social interaction play, and what synergies and syntheses occur? Studies are needed which examine real tasks integrated into people's daily lives at work and at home, based on longitudinal studies which capture the complexity of information behaviour [2]. Choo et al., [3] for instance, examined knowledge workers over a two week period and related their motivation in information seeking to the moves that they made online. While most such studies include only a small number of subjects and a limited time frame, at the other end of the scale is the Beauvisage study (described by Van Couvering [4]) which included 3,372 Internet users over 10 months and 597 users over 34 months. In this study 11.9% of sessions were "discovery" sessions, i.e. sessions in which the user seeks unknown information and expands his/her territory, a category which seems representative of exploratory search. We need more such studies which show how search is integrated into our lives.

2. Supporting Exploratory Search

Since exploratory search operates in areas which are unknown or poorly defined for the user, there are potentially many techniques to provide interactive system support for this activity, through tools for query development and improvement, navigation, organization and visualization [5]. Many such tools have been developed and evaluated in the context of defined search tasks, using metrics such as task time, error rate, success rate, and user satisfaction. However the challenge in providing tools to support exploratory search is in accommodating its external characteristics as an information-seeking activity. Modern search, particularly in the context of the Web, is ubiquitous, periodic, opportunistic, contextual, social, and progressive. What is needed is not simply a set of tools which make an information seeking task easier or more successful, but ones which are so well embedded in the overall information seeking environment and context that users turn to them automatically and integrate them in their work or everyday information seeking activities. This is a much greater challenge than simply developing a support system for exploratory search that performs well in a laboratory test or even on a set of realistic tasks.

3. Evaluating Exploratory Search

There has been much debate over the role of laboratory experiment in information retrieval, on which precision-recall based metrics are appropriate indicators of success in the real world, and on the degree to which metrics based on averages can predict individual query performance. Despite its limitations, the Cranfield model comprising test collections, queries and relevance judgements serves a very useful purpose in evaluating retrieval system performance, and certainly the model as implemented at TREC has led to significant improvements in term weighting schemes and rapid transfer of the technology to competing systems. It is important to know, at the system level, that best achievable performance is being delivered.

Evaluation at the operational level is far less standardized, more difficult to interpret. This is particularly true where the problem being searched is not well defined, or where the objective is to explore a new area rather than to arrive at a specific outcome. If the task is exploratory, then many paths through the information are possible, and the outcomes are fungible. Quality of support, nature of the interaction, learning outcomes, instances of synthesis or knowledge creation, and the quality of the experience are all difficult to measure but potentially important characteristics of a search support system.

The third stage of evaluation should move out of the laboratory or the staged experiment and into the real world of user behaviour. If the algorithms are as effective as possible, if the system supports exploratory search on controlled or user-generated tasks, then it will still not be successful unless users make it an automatic part of their information seeking activities. A successful system need not serve all areas of exploratory search or even a broad domain; Baby Names Wizard's NameVoyager[1] comes to mind as a tool which supports exploration and discovery in a very narrow domain, but does it very well. However, whether the system is designed for a specific workplace or search activity, or is intended to play a more general role in Web search, unless we design and evaluate systems to meet real-world, ubiquitous information needs they will not be successful in making the transition from the laboratory to the desktop.

[1] http://www.babynamewizard.com/voyager

REFERENCES

[1] L. Estabrook, E. Witt, and L. Rainie (2007). *Information Searches That Solve Problems: How People Use the Internet, Libraries and Government Agencies When They Need Help.* Pew/Internet. http://www.pewinternet.org/pdfs/Pew_UI_LibrariesReport.pdf

[2] K. Martzoukou. (2004). A review of Web information seeking research: considerations of method and foci of interest. *Information Research*, 10(2) paper 215. http://InformationR.net/ir/10-2/paper215.html]

[3] C.W. Choo, B. Detlor and D. Turnbull. (2000). Information seeking on the Web: an integrated model of browsing and searching. *First Monday*, 5 (2) http://firstmonday.org/issues/issue5_2/choo/index.html

[4] E. Van Couvering (2006). *Web Behaviour: Search Engines in Context.* . http://personal.lse.ac.uk/vancouve/thesis/navigation.pdf

[5] G. Marchionini (2006). Exploratory search: From finding to understanding. *Communications of the ACM* 49(4): 41-46.

The Information-Seeking Funnel

Daniel E. Rose
A9.com
130 Lytton Ave., Suite 300
Palo Alto, CA 94301 USA
+1 650-331-2670
danrose@a9.com

ABSTRACT

We introduce a new conceptual framework for information seeking called the information-seeking funnel, inspired by a common model of how people are attracted to products they end up buying. By using this model, we are able to better understand why current search tools are inadequate for several stages of the information-seeking process. At the same time, we explore what characteristics tools that are designed for these neglected stages might have, and how we might evaluate them.

1. INTRODUCTION

Information-seeking is a complex activity that can have varying degrees of directedness. Some parts of this activity involve focused search; others are more exploratory. But what does it mean to be exploratory? Is exploratory search an attitude or a process? What is the relationship between different kinds of information-seeking behaviors? We investigate these questions by using an analogy drawn on another type of familiar activity: shopping. While the task domains are quite different, they share the property that successive stages have different characteristics, and that different tools are appropriate at each stage.

2. SHOPPING FOR IDEAS

In the commercial world, there is a model called the *buying funnel* or the *sales funnel*. The buying funnel depicts the changing sets of people at different stages of the buying process, from all those who might possibly be interested in a product or service to those who actually purchase it. Sometimes the sets are described from the perspective of a sales representative, as “suspects,” “prospects,” “qualified leads,” and “customers.” Other times, the funnel is characterized by the mindset of the potential customer, where the four stages are “Awareness,” “Desire,” “Interest,” and “Action,” or (similarly) “Awareness,” “Research,” “Decision,” and “Purchase.” The “funnel” is so called because the graphical depiction illustrates the successively smaller sets of people that occupy each stage (see Figure 1). At each stage, the goal of the seller changes to correspond to the mindset of the buyer. For example, at the first stage, the seller’s goal is to make the potential buyer aware of the product.

Prepared as a position paper for National Science Foundation workshop on Information-Seeking Support Systems (ISSS), June 26-27, 2008, Chapel Hill, NC. Workshop organizers: Gary Marchionini and Ryen White.

The buying funnel model seems to be widely used in business. It’s mentioned everywhere from books on how to be a successful salesperson to blogs for search engine optimizers discussing how to get more clicks on paid search advertising. But what’s striking is that the funnel embraces all phases of the user’s mindset starting *before* she is even aware of the product or service.

2.1 The Information-Seeking Funnel

Now imagine that instead of potential customers buying products, we have users seeking information. By analogy, we can model the situation using an *information-seeking funnel*. And as with the buying funnel, we’ll start before users are aware of an information need, and end when they know it has been satisfied. Just as we can think of a product’s manufacturer hoping to draw a person deeper into the funnel, we can think of the supplier of some information (an author, perhaps) wanting that information to be consumed. Alternatively, we can think of the information nuggets themselves as memes [4] that “want” to be disseminated.

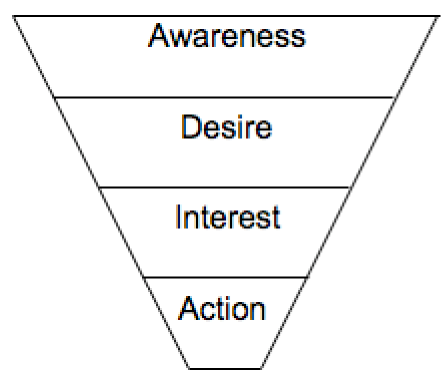


Figure 1: The buying funnel.

Of course, there have been many excellent models of information-seeking (e.g. ASK [3]) developed over the years as a result of studying how users search. But these models tend to focus on the situation after the user knows what information she is seeking. Some, like the Berrypicking model [1], acknowledge that the user’s information need starts out poorly formed and evolves. But in general, there is an assumption of a certain amount of directedness toward an information goal.

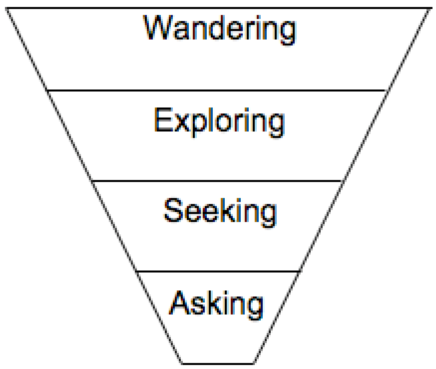


Figure 2: The information-seeking funnel.

The information-seeking funnel, shown in Figure 2, consists of four stages:

- *Wandering*. At this stage, the user does not have an information seeking-goal in mind. However, she may have a meta-goal (e.g. "find a topic for my final paper.")
- *Exploring*. In the Exploring stage, the user has a general goal (e.g. "learn about the history of communication technology") but not a plan for how to achieve it.
- *Seeking*. Here, the user has started to identify information needs that must be satisfied (e.g. "find out about the role of the telegraph in communication."), but the needs are open-ended.
- *Asking*. At the final stage, the user has a very specific information need that corresponds to a closed-class question ("when was the telegraph invented?").

The funnel shape now represents the narrowed scope of relevant information. At the first stage, all information in the world is potentially relevant; at the last stage, only a small amount of information is relevant.

3. INTERACTION AT EVERY STAGE

It is fair to say that most existing work on technology that supports information-seeking has focused on the last two stages, where the information need is more narrowly focused and the information-seeking task more directed.

If that's true, then opportunities exist to make a difference in neglected stages of the process. What might tools that are aimed at these other parts of the funnel look like? To investigate this, it is instructive to return to the shopping experience and think about the many types of interaction that are supported throughout the buying funnel.

3.1 Variety of Buying Funnel Interactions

In the early stages, users are interacting with the entire world. In order to gain their awareness, marketers rely on brand advertising in a variety of media. When a new class of products is introduced, the seller often has to convey some information about why someone would find such a product useful. Essentially, part of the task is education – initially, people may not even know what a phonograph / clothes dryer / cell phone is or why they would want one.

In the middle stages, users – now potential customers – are starting to seek out information about products. They may go to a store and look at the product or try it out; they may read reviews in online or offline media; they may rely on recommendations, either from friends or strangers, as in collaborative filtering; they may look at ratings from others who have bought the product.

Importantly, users at these stages employ a variety of modalities. For example, online purchases often involve complex combinations of searching, navigating, filtering through a faceted metadata space, and simply following suggestions that attract their attention.

As users narrow in on specific products, merchants provide specialized tools to facilitate that process. For example, articles of clothing sold online may be viewed from several different angles or with the option to zoom in on details. The task changes from learning about these kinds of products to learning about this specific product. What's important is not only that there are such a wide variety of tools, but also that the tools are appropriate to the funnel stage.

3.2 Stateful Search

Returning to the information-seeking funnel, what might some tools to support the upper stages look like? It is hard to predict which of the many new approaches being explored in the research community are likely to be successful, but it seems clear that at least one aspect of these systems will be what we call *stateful search*.

In recent years, most general-purpose search engines tend to behave in a stateless way. That is, each interaction with a user is performed in isolation, without taking into account any previous interactions with users.[1]

In contrast, a stateful search system treats the process, not just the result, as a first-class object of interest. A stateful search system will not only keep a history of each step in the information-seeking process (searching, browsing, viewing results, etc.), it will also allow users to interact with that state. Furthermore, such a system will allow users to gather, analyze, annotate, and organize information, creating new explicit representations of state alongside those kept internally by the system.

But doing all this is going to require different kinds of user interactions than search engines expect today – interactions that in many cases might be characterized as "more work" than they expend when entering a query today. Is it reasonable to expect that users can change their behavior?

3.3 The Myth of the Two-Word Query

In 1996, the ATG Information Access Group at Apple packaged the V-Twin search software we had developed[2] into a free web site search engine called "Apple e.g.," which powered an early

[1] In a sense, "personalized" search systems, which analyze the past actions of individual users to learn their general preferences, may be considered stateful, but not at the level of granularity we mean here.

[2] Doug Cutting was the architect and lead developer of V-Twin.

version of Apple's own web site as well as several other university sites. Observing that user queries on these sites were quite short, we came up with a method for improving performance of the engine on short queries. We presented our results at TREC [8], where we simulated the behavior of short queries on the TREC ad hoc task by truncating the topic titles to the two words with the highest term weight. At the time, we had trouble convincing people that worrying about two-word queries was worthwhile. I overheard another participant at the conference questioning the value of our work during one of the breaks. "If we're only given two words to work with," our critic said, "then there's pretty much nothing we [IR researchers] can do."

A few years later, the pendulum had swung absurdly far in the opposite direction, where it remains today. A series of web use studies starting in 1998 [10] confirmed that short queries are the norm on the web. Not only are all major search engines optimized for two-word queries, there is now conventional wisdom that says that users will never type more than this. This is sometimes characterized as "users are lazy," and other times as "users are efficient." In some ways, the latter sentiment is worse, implying that search engines are so good that no more than two words are necessary for describing a user's information need.

As we have discussed above and elsewhere [7, 9], the existing search engine interaction paradigm of typing two words in a text box and getting a linear list of ten results is ill-suited to many information-seeking tasks. In particular, supporting all stages of the information-seeking funnel will require a variety of interaction modalities, just as is the case with the buying funnel. But are users so strongly trained in the existing paradigm, and so reluctant to enter more than two words, that tools with alternate interaction styles are doomed to failure?

Several pieces of evidence suggest that they are not. We know from multiple studies [2, 5] that the user interface – even the shape of the search box – can influence the length of user queries. Perhaps more importantly, there are thriving communities where users routinely expend much more effort and enter far more information. At Yahoo! Answers [11], a service where users ask questions and others in the community answer them, the "short" version of the question is typically 10 or 12 words – and this is usually followed by a more detailed version that can be as long as a paragraph. A few recent short questions are:

- *Does Barack Obama play video games?*
- *Where can I find a good mariachi in LA county or OC for my wedding?*
- *What documents are required to bring the child of my fiancée to the USA on a K1 visa?*

Just part of one typical detailed question starts:

> *I am 12 years old and i need a dress for my friends bar mitzvah. I want it to be either red, gray, or black and not too expensive. I got some really cute black wedges but i dont have a dress but i could exchange the shoes. I need it by the 21st so it has to be a store i can go buy it at…*

What these examples suggest is that users are not only adaptable but also quite willing to switch to a different interaction style if they believe that style will provide more value in that system. If the application is compelling enough, users will happily take the time to provide the information that will make them successful.

4. UPPER-FUNNEL EVALUATION

Once we have these new tools for supporting the upper states of the information-seeking funnel, how will we evaluate them? For specific information-seeking tasks at the lower half of the funnel, we can evaluate a system based on how well it helps users satisfy their information need. Metrics such as recall, precision, and discounted cumulated gain (DCG) can answer questions like "did the system find all the material on the topic" and "were the results presented in the optimal order." But those metrics are not useful when the information need is not yet clearly defined.

Some systems are also judged by the *speed* with which users can "get answers to their questions." I have heard many product managers in the search industry, and even some HCI professionals, claim that a good search engine is one in which users "get in and get out as quickly as possible." Not only does this perspective treat the *quality* of the information obtained as irrelevant, it is predicated on the assumption that there is no value to the user in exploration [6].

In short, metrics like these can't tell us whether a system did a good job helping users explore different ideas or prompting them to make connections.

4.1 Collateral Knowledge

At the top half of the funnel, then, what we need are ways to measure the ability of the system to convey unsolicited information – essentially, to put the user in an information-seeking frame of mind.

We began exploring this concept during research at Apple in the 1990s on the roles of browsing and searching in information-seeking systems.[3] We hypothesize that while a user is conducting an information-seeking task, he acquires insights and information along the way that he was not explicitly looking for, but that may prove to be valuable. This serendipitously-acquired information we call *collateral knowledge*.

To measure the ability to convey collateral knowledge, we designed the following methodology: First, we gave users a pre-test to measure their general knowledge of a certain subject area. Next, we gave them a set of directed information-seeking tasks using a particular system. Finally, we gave them a post-test on the subject, including topics they had not been asked to learn about, but which they might have been exposed to during the process. By comparing the post-test answers to the pre-test answers, we were able to measure the collateral knowledge obtained during the information seeking process.[4]

Since users who are at the early stages of the information funnel don't exactly know what they're looking for yet,

---

[3] This work was done with Paul V. Biron.

[4] The specific study we conducted at Apple was intended to compare the abilities of search-oriented vs. browse-oriented systems to convey collateral knowledge, but for a variety of logistical reasons, the study was never completed.

collateral knowledge may be an interesting measure of how well the system is supporting exploratory search.

5. CONCLUSIONS

The information-seeking funnel is a model of the information-seeking process encompassing the period from before the user knows that an information need exists, to the point at which an information need is satisfied. By means of its analogy to the buying funnel, it serves as a way to frame the problem of information-seeking tools and their evaluation. It helps explain why mainstream search engines aren't doing a good job of supporting exploratory search. It makes clear that – just as with shopping – different tools with different interaction styles will be needed at each stage. It illustrates why metrics for the latter stages make little sense for the earlier ones. In summary, the information-seeking funnel provides a conceptual framework for thinking about the problem space.

6. REFERENCES

[1] Bates, M. J. 1989. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Review*, 13(5):407-24.

[2] Belkin, N. J., et al. 2001. Rutgers' TREC 2001 Interactive Track Experience. *Proceedings of the Tenth Text Retrieval Conference* (TREC-10).

[3] Belkin, N. J., Oddy, R. and Brooks, H. 1982. ASK for Information Retrieval. *Journal of Documentation*, 38, 61-71 (part 1) & 145-164 (part 2).

[4] Dawkins, R. 1976. Memes: The New Replicators, in *The Selfish Gene*. Oxford University Press.

[5] Karlgren, J. and Franzen, K. 2000. Verbosity and Interface Design. *SICS Technical Report T00:04*, Swedish Institute for Computer Science, Stockholm.

[6] Rose, D. E. 2004. Why Is Web Search So Hard... to Evaluate? *J. Web Eng.* 3(3-4): 171-181.

[7] Rose, D. E. 2006. Reconciling Information-Seeking Behavior with Search User Interfaces for the Web. *JASIST* 57(6):797-799.

[8] Rose, D. E. and Stevens, C. 1997. V-Twin: A Lightweight Engine for Interactive Use, *Proceedings of the Fifth Text Retrieval Conference* (TREC-5).

[9] Rose, D. E. and Raju, S. 2007. Encouraging Exploration with Elroy: A New User Experience for Web Search. *Proceedings of the ACM SIGCHI 2007 Workshop on Exploratory Search and HCI* (ESI 2007), 52-55.

[10] Spink, A., Bateman, J., and Jansen, B. J. 1998. Users' Searching Behavior On The Excite Web Search Engine. *WebNet 1998.*

[11] Yahoo! Answers, http://answers.yahoo.com, accessed 12 June 2008.

Five Challenges for Research to Support IS^3

A Position Statement

Elaine G. Toms
Centre for Management Informatics, Faculty of Management
Dalhousie University
Halifax, Nova Scotia, Canada B3H 4J5
+1 902-494-8374
etoms@dal.ca

1. INTRODUCTION

While many problems remain to develop effective IS^{3}s (Information Seeking Support Systems), I have identified five issues that focus primarily but not totally on the information consumer, reflecting a user rather than system centric focus. These include: a better understanding of needs, tasks and intentions; an examination of applicability of collaboration to IS^3, the development of search as a "real" application; the identification of what constitutes success for IS$^{3;}$ and a call for more holistic exploration of the IS^3 ecology.

2. THE CHALLENGES

2.1 Needs, Tasks and Intentions

Why does a person search for information (either implicitly or explicitly)? To date this has been explained in multiple ways from Taylor's initial conceptualization of need [3], to that of task. Confusing is the interchangeably use of expressions of needs, intentions of the consumer, articulation of task, and query.

In recent years, task has emerged as a significant mediator of the search process, and as a result, many taxonomies of task have been developed (see [4] for example). Immersed in task is the consumer's intention for the search. A single task may result in multiple, but inter-related intentions that are often expressed as information tasks, each of which may result in one or more queries or other mechanisms for acquiring information. Thus what is often described as a need or as a task is in fact a dynamic complex process that starts from a realization and works toward resolution, some of which is internalized to the individual and some of which is manifested in system interaction. Research to date has focused primarily either on the query, or on the high level task.

Queries are, falsely, often considered synonymous with need and task. Analogically, a three to four word query can be equated with a doctor-patient exchange in which the patient expresses the problem as "pain arm" and the doctor is expected to distinguish among a host of problems from a potential heart attack to a bug bite. The predominant two to five keyword queries suffer from the same problem. How can consumer intensions, information goals, and/or tasks be predicted from such a limited amount of information?

At the other end of the spectrum, significant qualitative work have explored task, but not at a level that enables operationalization of task. Confusing also is the multiple roles that task takes in an experimental human study. Tasks are the vehicle by which the system use is observed, and at the same time may be the experimental variable that is being tested. As a result, tasks when developed for experimental purposes have so many confounds that it is difficult to assess which aspect of task (if any) makes a difference. We tend not to create exemplars that are representative of particular categories of task.

Untangling needs, from intentions and tasks is a first step that is needed before designing how those intentions can be converted into concrete interaction-able processes.

2.2 Collaboration

Collaboration has become a new element in the search research agenda fueled by the massive effort in mining and analyzing community use relationships, e.g., links/hubs, popularity, and recommendations. But a community is distinctly different from a team or group whose activities occur via collaboration. Collaboration, the process by which a team or group works together with a shared purpose, may involve a dyad (i.e., two person groups) or a team of three to ten people (more and the group will become a community) who may share a common physical space within a common timeframe.

The challenge for collaboration in IS^3 is sifting through the information space to separate distinctly individual activities or processes from those that can be performed by a group. Collaboration may be performed in multiple configurations: work is handed off from one person to the next so that the work is done in an asynchronous like way, or work is done at the same time and in the same or different place. Now rather than a single information consumer, multiple people are sharing in the operation. Whatever the process, the needs for IS^3 will differ. But first we need to know how much of information seeking/search is truly individual and which aspects are performed more effectively by a group. Within an IS^3, is the "handed off" type of collaboration more likely than the joint collaboration, or does it matter?

2.3 The Search Dashboard

The search interface has been anchored in the query-box structure – a dumb waiter style of interaction – since that format was imposed on the search environment without any requirements assessment. We have essentially replaced the Unix command line prompt with a box. Sadly, now a global information consumer community has been "trained" to think only in those terms, and to think of search as the two-mouse-click/three-web-page solution that has been promoted by Google's minimalist design. From a design perspective, search has been treated only at the physical level of interaction: push a button, get a response. Yet, much like any other application there is immense potential for tools and widgets to aid the search process, and for re-thinking search as an application, and not just as a support tool.

Part of the challenge is that search is often thought of in the generic one-size-fits-all sense, yet search is integral to practically all other applications, from simple help functions in typical office applications (e.g., word processors, accounting information systems), to standard content-rich systems (e.g., newspapers and the global information network – the web), and intense information use environments such as educational programs and knowledge work in business and government. What tools (other than a query box) are required to aid the search for and use of information in those settings? Will they vary from application to application (e.g., news reading versus help systems)? In particular what tools are needed to support the original "work" task (including intention and use) that triggered the information quest? This will require an examination of search as an application (e.g., a place to collect useful pages, methods for visualizing results and content, support for query construction and results assessment). Notably these tools are tightly interwoven with aggregation of the content that are created using association and relationships among the various information objects.

Part of this is also re-assessing the two-mouse-click/three-web-page solution that currently exists: is there a one-page, no mouse-click solution? An element that bears assessing is the concept of awareness which emerged as a core element within CSCW (computer supported cooperative work) research and application development. Awareness is often referred to as "knowing what is going on;" rarely does a search interface provide the type of fluid information that is required to maintain control over action with useful feedback from the system. It is not just about the results.

2.4 Measurement and Evaluation

Together, topical relevance and usability have driven the measurement of search systems for the past few decades. Yet the research community continues to be challenged by what to measure, and in turn what constitutes appropriate metrics for each. We tend to measure the obvious (e.g., time on a (sub)process, mouseclicks, number of words) and report what is easily countable or quantifiable in some way.

Most recently biometrics have emerged as another technique for measuring – something! And that leads to the crux of the problem: we still have not clearly defined "success" in the context of IS^3 [6]. When is a IS^3 successful? Only when a relevant item is located? Only when no relevant items are located? Only when something that is immediately useful is located? When the process has been engaging? When the system is intuitive? Likely success is a complex element with multiple mediating factors.

Well known in the management information systems area is the Delone and McLean model of information systems success in which information, system and service quality predict use and intention to use which predicts benefits [1]. The intent is a holistic perspective that is not focused only on one aspect of the problem, e.g., topical relevance or user satisfaction. The Delone and McLean model provides a useful example of what is needed in IS^3.

In addition to evaluating outcomes, we have no way of evaluating process. Since the process of information seeking/searching is a dynamic one with all variables in a constantly changing mode, we need a novel method to do that assessment.

2.5 Complexity of the Search Problem

The information seeking/searching environment is a complex world of multiple competing variables that impact how information is sought, what information is examined and how information is subsequently used, and thus on the design of systems to support information seeking/searching. There are multiple competing factors, each of which has an impact on how an IS^3 is subsequently used. A view of the complexity is illustrated in Figure 1 which is derived from [5].

We know for example a little about how individual differences (e.g., experience, spatial ability, knowledge) among consumers affect their search capabilities, but we know little about how the design of IS^3 s can account for those differences. We have focused primarily on analysis of the content of information objects, but have rarely taken into account the classic metadata that has existed and used primarily in library science for decades (e.g., genre, date, author, specialized indexing) to examine their contribution to meaning or to meaningful results (see [2] for one example). The formal "work" task has been all but ignored and the relationship between the work task and the resulting information task is tenuous. Except in discussions of information behaviour which tend to provide only a high level and non-operational view of information seeking, we rarely find the situation or work environment considered as a factor in the information seeking.

Overall we know something about some of these variables but we know little about the inter-relationships among the variables. Which ones are the most influential? What is the most parsimonious set that will influence the use of an IS^3 ? Once we understand more about the information ecology of the process, then we will be in a better position to determine the requirements for an IS^3 .

3. REFERENCES

[1] DeLone, W.H., & McLean, E.R. 2003. The DeLone and McLean Model of Information Systems Success: A Ten-Year Update, Journal of Management Information Systems 19(4 2003), 9-30.

[2] Freund, L, Toms, E.G., & Clarke, C. Modeling task-genre relationships for IR in the workplace. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval: SIGIR '05, Salvador, Brazil, August 15-18, 2005*, pp. 441-448.

[3] Marchionini, G. Exploratory search: from finding to understanding. *CACM* 49(4 2006,) 41-46

[4] Taylor, R.S. Information use environments. *Progress in Communication Sciences*, 10 (1991), 217-255.

[5] Toms, E.G., Bartlett, J., Freund, L., Dufour, C., & Szigeti, S. Identifying the significant contextual factors of search. In: *Proceedings of the SIGIR 2004 Workshop on Information Retrieval in Context, ,Sheffield (UK), 29 July2004.*

[6] Toms, E.G., O'Brien, H., Kopak, R. & Freund, L (2005). Searching for relevance in the relevance of search. In *Context: nature, impact, and role: COLIS5: Proceedings of the Fifth International Conference on Conceptions of Library and Information Sciences (Glasgow, UK, June 5-8*, 2005), pp. 59-78. Lecture Notes in Computer Science 3507

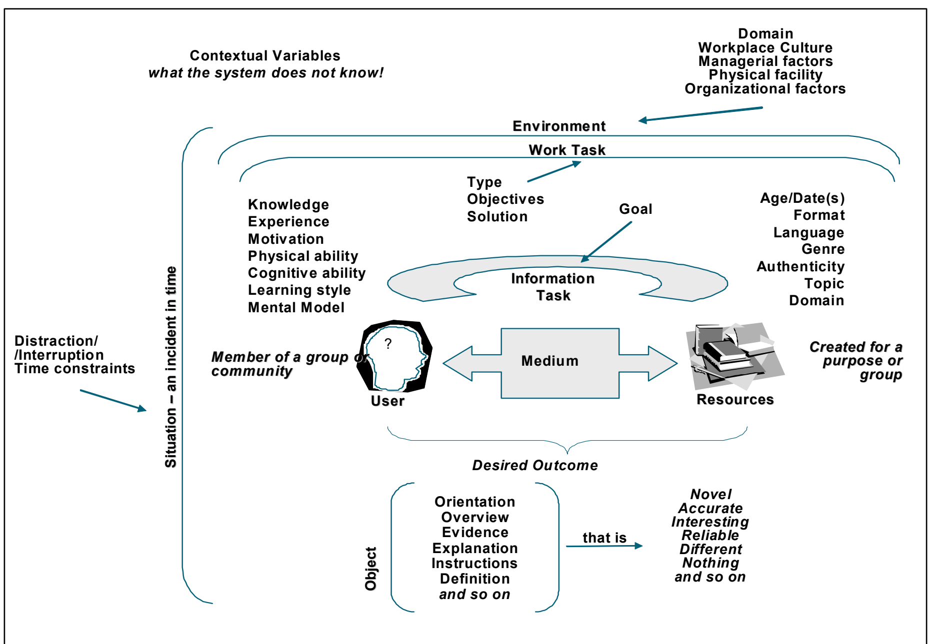


Figure 1. The IS^3 Environment (adapted from [5])

Information Seeking can be Social

Ed H. Chi

Augmented Social Cognition Area
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304 USA
echi@parc.com

Submitted to special issue on 'Exploratory Search'

Nov. 23, 2008

As part of information seeking, exploratory search involves ill-structured problems and more open-ended goals, with persistent, opportunistic, iterative, multi-faceted processes aimed more at learning than answering a specific query [Marchionini 2006]. Whereas for the fact-retrieval searches, an optimal path to the document(s) containing the required information is crucial, learning and investigation activities lead to a more continuous and exploratory process with the knowledge acquired during this "journey" being essential as well [White et al. 2007]. Therefore, information seeking systems should focus on providing cues that might make these explorations more efficient.

One possible solution is building social information seeking systems, in which social search systems utilizes social cues provided by a large number of other people. What is social search? How might be build social search systems? Is there a need for such solutions?

The need for Social Search

Information retrieval researchers typically depict information seeking as solitary activities of a single person in front of a web browser. This view is slowly changing. For example, other pieces in this theme issue point to recent interest in collaborative, co-located search [Golovchinsky sidebar] as well as social bookmarking [Millen sidebar]. These recent trends point to the social nature of information seeking. Indeed, recent research done with 150 participants by the Augmented Social Cognition group at Palo Alto Research Center suggest that many information seeking activities are interwoven in-between social interactions [Evans and Chi, 2008]. Our research suggests analyzing the search process by looking at three stages of before, during, and after the search:

(1) **Before**: We saw users engaged in social interactions 43% of the time before exploring on the web. These social interactions supported information gathering by providing opinions and advice such as websites or keywords to try. For example, a programmer might have engaged in a series of emails with coworkers asking about the existence of various monitoring software packages for a web server, and the merits of each package. An analysis of only search engine logs might have simply recorded several refinements of queries in a single 30-minute session rather than detecting the considerable amount of social preparation done before searches.

(2) **During**: Social interactions are also common during the search act itself. For example, people sometimes conducted searches with others who are co-located, in which they might take turns suggesting and trying out search keywords. In these cases, users are likely to interact with others during informational exploratory searches. Around 40% of the users engaged with others both before and during the information search.

(3) **After**: Users often either organize the search results or distribute them to others in their social network. For example, after a barista found a particular recipe, she printed it out and shared it with all of her coworkers. In fact, we observed users distribute search results to others quite frequently at around 60%.

We have integrated our findings with models from previous work on sensemaking and information-seeking behaviors [Evans and Card, 2008] to present a canonical model of social search. Figure 1 below depicts this descriptive model. We see that, when viewed holistically, information seeking is more than just a database query. Instead, information seeking is often embedded within social relationships. The social networks are both sources of requests as well as suggestions. They are also sinks in which refined results are distributed.

Our results and analysis demonstrated that users have a strong social inclination throughout the search process, interacting with others for reasons ranging from obligation to curiosity. Self-motivated searchers and users conducting informational searches provided the most compelling cases for social support during search.

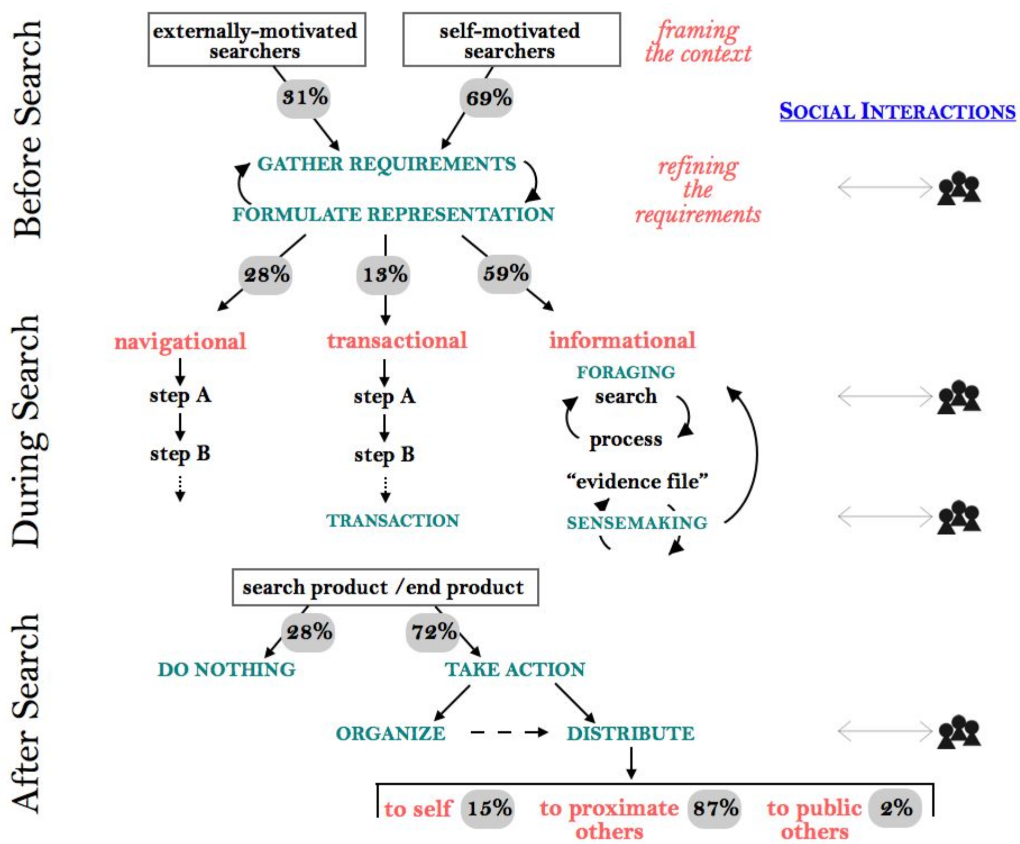


Figure 1. Combining with previous models of information seeking behavior, a canonical model of social search shows three stages in a search act weaved in between social interactions.

Different kinds of Social Search Systems

Researchers and practitioners now use the term "social search" to describe search systems in which social interactions or information from social sources is engaged in some way [Evans and Chi 2008]. Current social search systems can be categorized into two general classes:

(a) **Social answering systems** utilize people with expertise or opinions to answer particular questions in a domain. Answerers could come from various levels of social proximity, including close friends and coworkers as well as the greater public. Yahoo! Answers (answers.yahoo.com) is one example of such systems. Early academic research includes Ackerman's Answer Garden [Ackerman, 1996], and recent startups include Mechanical Zoo's Aardvark (vark.com) and ChaCha's mobile search (chacha.com).

Some systems utilize social networks to find friends or friends of friends to provide answers. Web users also use discussion forums, IM chat systems, or their favorite social networking systems like Facebook and Friendfeed to ask their social network for answers that are hard to find using traditional keyword-based systems. These systems differ in terms of their immediacy, size of the network, as well as support for expert finding.

Importantly, the effectiveness of these systems depends on the efficiency in which they utilize search and recommendation algorithms to return the most relevant past answers, allowing for better constructions of the knowledge base.

(b) **Social feedback systems** utilize social attention data to rank search results or information items. Feedback from users could be obtained either implicitly or explicitly. For example, social attention data could come from usage logs implicitly, or systems could explicitly ask users for votes, tags, and bookmarks. Direct Hit[1] was one early example from early 2001 that used click data on search results to inform search ranking. The click data was gathered implicitly through the usage log. Others like Wikia Search (search.wikia.com), and most recently Google, are allowing users to explicitly vote for search results to directly influence the search rankings.

Vote-based systems are becoming more and more popular recently. Google's original ranking algorithm PageRank could also be classified as an implicit voting system by essentially treating a hyperlink as a vote for the linked content. Social bookmarking systems such as del.icio.us allow users to search their entire database for websites that match particular popular tags.

One problem with social cues is that the feedback given by people is inherently noisy. Finding patterns within such data becomes more and more difficult as the data size grows [Chi and Mytkowicz, 2008]

In both classes, there remains opportunity to apply more sophisticated statistical and structure-based analytics to improve search experience for social searchers. For example, expertise-finding algorithms could be applied to help find answerers who can provide higher-quality answers in social answering systems. Common patterns between question-and-answer pairs could be exploited to construct semantic relationships that could be used to construct inferences in question answering systems. Data mining algorithms could construct ontologies that are useful for browsing through the tags and bookmarked documents.

---

[1] http://www.searchengineshowdown.com/features/directhit/review.html

MrTaggy as an example of Social Search Systems

At PARC, Rowan Nairn and I have been constructing a social search system based on using statistical machine learning analytics. Our system, called MrTaggy (mrtaggy.com), relies on 150 million bookmarks crawled from the web to construct a similarity graph between tag keywords. MrTaggy's tag-search browser uses this similarity graph to recommend and search through other tags and documents.

Figure 2. MrTaggy's user interface with related tags list on the left and search results lists presented on the right.

The Figure above shows a typical view of the tag search browser. MrTaggy provides typical search capabilities (query input textbox and search results list) combined with explicit relevance feedback for query refinements. Users have the opportunity to give relevance feedback to the system in two different ways, at the fine-grained item level and at a coarse descriptor (tag) level:

Related Page Feedback: Clicking on the upward or downward arrow on a search result includes or excludes it from the result list. This feedback also results in emphasis of other similar or de-emphasis of other dissimilar web pages.

Related Tag Feedback: On the left a *related tags list* is presented, which is an overview of other tags related to the current set of tag keywords. For each related tag, up and down arrows are displayed to enable the user to give relevance feedback by specifying relevant or irrelevant tags.

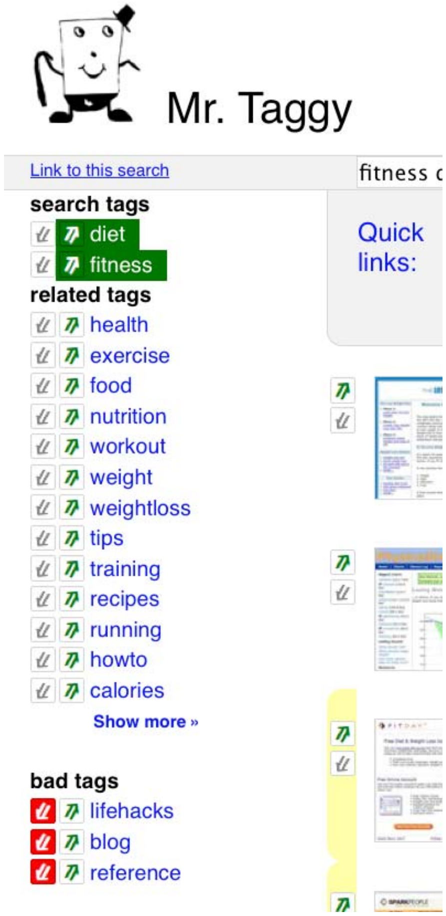


Figure 3. MrTaggy user interface for adding relevant and irrelevant tags.

For a search result, MrTaggy displays the most commonly used tags describes the content of the web page, in addition to the title and the URL of the corresponding web page. Other people applied these tags to label the corresponding Web page. When hovering over tags presented in the snippet, up and down arrows are displayed to enable relevance feedback on these tags as well.

Having just described the interaction of the relevance feedback part of the system, we now describe how it operates in concert with the backend. Figure 4 below shows an architecture diagram of the overall system.

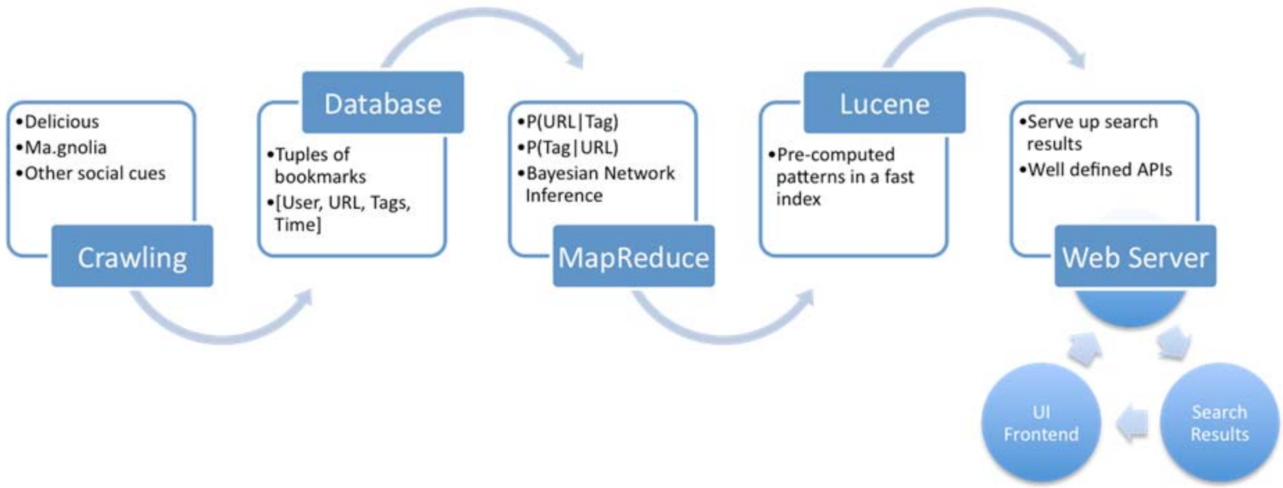


Figure 4. Overall architectural diagram of the MrTaggy tag-based search browser.

First, a crawling module goes out to the web and crawls social tagging sites, looking for tuples of the form <*User, URL, Tag, Time*>. Tuples are stored in a MySQL database. In our current system, we have roughly 150 million tuples. A MapReduce system based on Bayesian inference and spreading activation then computes the probability of each URL or tag being relevant given a particular combination of other tags and URLs. Here we first construct a bigraph between URLs and tags based on the tuples and then precompute spreading activation patterns across the graph. To do this backend computation in massively parallel way, we used the MapReduce framework provided by Hadoop (hadoop.apache org). The results are stored in a Lucene index (lucene.apache.org) so that we can make the retrieval of spreading activation patterns as fast as possible.

Finally, a web server serves up the search results through an interactive frontend. The frontend responds to user interaction with relevance feedback arrows by communicating with the web server using AJAX techniques and animating the interface to an updated state.

An Example Evaluation of Social Search System Effectiveness

We recently completed a 30-subject study of MrTaggy [Kammerer et al. 2008]. In this study, we analyzed the interaction and UI design. The main aim was to understand whether and how MrTaggy is beneficial for domain learning.

We compared the full exploratory MrTaggy interface to a baseline version of MrTaggy that only supported traditional query-based search. We tested participants' performance in three different topic domains and three different task types. The results show:

(1) Subjects using the MrTaggy full exploratory interface took advantage of the additional features provided by relevance feedback, without giving up their usual manual query typing behavior. They also spent more time on task and appear to be more engaged in exploration than the participants using the baseline system.

(2) For learning outcomes, subjects using the full exploratory system generally wrote summaries of higher quality compared to baseline system users.

(3) To also gauge learning outcomes, we asked subjects to generate keywords and input as many keywords as possible that were relevant to the topic domain in a certain time limit. Subjects

using the exploratory system were able to generate more reasonable keywords than the baseline system users for topic domains of medium and high ambiguity, but not for the low-ambiguity domain.

Our findings regarding the use of our exploratory tag search system are promising. The empirical results show that subjects can effectively use data generated by social tagging as "navigational advice". The tag-search browser has been shown to support users in their exploratory search process. Users' learning and investigation activities are fostered by both relevance feedback mechanisms as well as related tag ontologies that give scaffolding support to domain understanding. The experimental results suggest that users' explorations in unfamiliar topic areas are supported by the domain keyword recommendations presented in the related tags list and the opportunity for relevance feedback.

Conclusion

Since social search engines that depend on social cues rely on data quality and increasing coverage of the explorable web space, we expect that the constantly increasing popularity of social bookmarking services will improve social search browsers like MrTaggy. The results of this project point to the promise of social search to fulfill a need in providing navigational signposts to the best contents.

References

1. Ackerman, M. S.;McDonald, D. W. 1996. Answer Garden 2: merging organizational memory with collaborative help. In Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (Boston, Massachusetts, United States, November 16 - 20, 1996). M. S. Ackerman, Ed. CSCW '96. ACM, New York, NY, 97-105. DOI= http://doi.acm.org/10.1145/240080.240203
2. Chi, E. H.; Mytkowicz, T. Understanding the efficiency of social tagging systems using information theory. Proceedings of the 19th ACM Conference on Hypertext and Hypermedia; 2008 June 19-21; Pittsburgh, PA. NY: ACM; 2008; 81-88.
3. Evans, B.; Chi, E. H. Towards a Model of Understanding Social Search. In Proc. of Computer-Supported Cooperative Work (CSCW). ACM Press, 2008. San Diego, CA.
4. B. M. Evans and S. K. Card. Augmented information assimilation: Social and algorithmic web aids for the information long tail. In Proc. CHI2008, ACM Press, pages 989–998, 2008.
5. Kammerer, Y.; Nairn, R.; Pirolli, P. L.; Chi, E. H. Signpost from the Masses: Learning Effects in an Exploratory Social Tag Search Browser. (to appear) In Proc. CHI2009. Boston, MA.
6. Marchionini, G. Exploratory search: From finding to understanding. *Communications of the ACM*, *49*, 4 (2006), 41-46.
7. White, R.W., Drucker, S.M., Marchionini, M., Hearst, M., schraefel, m.c. Exploratory search and HCI: designing and evaluating interfaces to support exploratory search interaction. *Extended Abstracts CHI 2007*, ACM Press (2007), 2877-2880.

Collaborative Information Seeking

Gene Golovchinsky, Pernilla Qvarfordt, Jeremy Pickens

FX Palo Alto Laboratory, Inc.

{gene, pernilla, jeremy}@fxpal.com

Introduction

People often work together when searching for information, but traditional information seeking support systems do not support such behavior, causing people to adopt workarounds such as sending links by e-mail or over-the-shoulder-surfing to compensate for a lack of adequate tools [6].This observation has led several research groups to explore various aspects of collaboration in support of information seeking [10] [7] [8]. The goal of that research is to create tools that support collaborative search based on the notion of explicitly-shared information needs.

Collaboration and information seeking can assume many forms. The needs and search activities of John and Gail described in the following paragraphs illustrate some of the challenges for collaborative information seeking.

John is a middle-aged accountant who recently underwent successful surgery for colon cancer. He knows that he will have a long life as a cancer survivor as long as he maintains a healthy life style. He continues to learn about the disease, joins online support groups, and desires tools that allow him to find and organize information about colon cancer. As part of his support group he offers encouragement and advice to others.

John joined an online community to identify a group of peers – fellow cancer survivors, researchers, doctors, and other motivated and knowledgeable people – with whom he can co-ordinate his ongoing investigations. Searching independently, they form a trusted search community [1] that will make automatically-generated recommendations more coherent since the community has similar information needs. They can also form synchronous collaborations in small groups to explore and share information about new developments and to track ongoing research in the field.

John has been lobbying his insurance company and hospital to expand their post-treatment counseling services to obtain online search help from registered nurses (RNs) who specialize in particular diseases. Working through a dedicated portal, John should be able to get feedback from his RN counselor on the results he is finding on this own and with his peers. The RN would have access to John’s medical record (the details of which may not be available to John in a useful form) that she could use to identify information relevant to John’s condition.

Finally, John would like a tool to manage the information set related to his cancer both to maintain a record of what he has learned, and to publish a subset of that, along with his commentary, to help his family and friends understand his condition, and perhaps to draw them into his circle of collaborators.

Gail is a neuroscientist who leads a multi-campus team working on new techniques for using viruses as vectors for neuronal therapies. Gail also consults for a large biotech company. She scans a dozen journals regularly, participates in online forums, and has a post-doctoral fellow who regularly conducts systematic searches of the biomedical literature. Her research group maintains a database of papers and much of her time is spent re-finding and integrating new ideas she gains from preprints, discussions with colleagues, and results from her lab work. She needs tools that integrate search activities across diverse collections of public and personal information streams and allow flexible organization and analysis.

Gail often relies on her clinical colleagues to identify and interpret newly-published research results. She often wishes that they could share particular search sessions, as well as the summaries she currently receives. Gail can often direct less experienced colleagues to search on topics they did not initially consider. Sometimes in meetings, Gail and her colleagues discuss particular searches and run additional searches during the discussion.

Gail can access the biotech company's library of technical reports and pre-prints through a secure web portal. While this is a useful resource, she is frequently confused by the organization of the sub-collections and by the company-specific terminology and metadata. Having online access to a reference librarian during search sessions would make her more productive.

We can use these scenarios to consider how collaboration can be enhanced by information-seeking support tools.

Forms of collaboration

The term "collaboration" has been used in the information seeking literature to refer to a variety of forms of mediation, communication, and coordination [9] that may not have as much in common as the name suggests. In an attempt to distinguish among the various forms of computer-supported collaboration for information seeking, we propose that such systems can be classified along four dimensions: intent, depth of mediation, concurrency, and location [3]. Collaboration also implies that different people can play different roles in the overall human-computer system. In the following sections, we first explore the dimensions of collaborative search, and then focus on roles.

Dimensions of collaboration

Intent: explicit vs. implicit. Explicit information seeking occurs when two or more people set out to find information on a topic based on a *declared* understanding of the information need (that may evolve over time). In our scenarios, examples of explicit collaboration are when Gail collaborates with clinical colleagues or works with the librarian, and when John receives help from the RN. In contrast, implicit intent characterizes collaborative filtering systems that *infer* similar information needs based on users' actions or opinions. Recommendations such as "People who bought this product also bought…" [4] and "What sort of things are my co-workers finding and reading, that I also need to be aware of" that use aggregate statistics to make recommendations are examples of implicit collaboration, Such recommendations are only useful if previous people were motivated by similar information needs, as in the case of John and his online community. See the "Social Search" [2] and "Better Exploratory Search Through Social Bookmarking" [5] sidebars in this issue for a discussion of implicit search.

Depth of mediation. Depth of mediation is the level at which collaboration is represented in the system (User Interface (UI) vs. search engine back end). The level of mediation affects how aware a system is of the contributions of different people, and how it uses those contributions to affect searches. Collaborative filtering systems, for example, keep track of each user's data separately, before aggregating them to make specific recommendations. Cerchiamo [8] uses relevance judgments of the collaborators to influence the ranking of search results. Systems such as SearchTogether [7], on the other hand, distinguish among users only in the user interface; the search engine component is unaware that multiple people have contributed queries, saved documents, etc.

Concurrency. People can collaborate synchronously or asynchronously. Synchronous collaboration implies the ability of people to influence each other in real time as Gail does when she meets with her colleagues to search and discuss the results. Asynchronous collaboration describes situations where influence is created from previous searches, either personal or aggregated from a community, as when John receives recommendations based on results found by others in his peer group. Systems can support both synchronous and asynchronous collaboration: in SearchTogether, for example, users can save search results for others who may not be online, or they can run a query and examine search results together.

Location. Finally, collaborators may work in the same place at the same time, which allows them to communicate in a variety of ways not necessarily mediated by the computer, or they may be distributed, increasing opportunities for collaboration but decreasing the fidelity of possible communications.

Roles in collaboration

When working with others, Gail and John may divide the task in different ways depending on the roles they and their collaborators assume.. How tasks are divided may depend on the nature of the task, peoples' expertise, and on the capabilities of the system that mediates information seeking. Roles may be implicit in the functionality of the interface (specifying queries, making relevance judgments), or they can be more explicit (where people use different interfaces for different subtasks).

Peer. The most obvious role set is the peer. All collaborators use the same interfaces to control the system and to coordinate their activities. This is the most common situation with existing (non-mediated) tools. Each collaborator uses an independent system, and they combine their results manually, as Gail and her colleagues need to do. Experimental collaborative search systems such as Físchlár-DiamondTouch [10] and SearchTogether [7] implement peer search roles. In SearchTogether, each person is able to specify queries, to examine results, etc. Físchlár-DiamondTouch uses a multi-touch surface to allow groups of people to manipulate search results in parallel.

Domain A expert-Domain B expert. A variation on the *Peer* role is collaboration between people with symmetrical user interfaces but with different domain knowledge, (e.g., when Gail collaborates with physicians). Mediation can increase the chances that documents relevant to both sets of expertise will be recognized by the users.

Search expert-Search novice or Domain-expert-Domain-novice. One common asymmetry in user skill and experience is the degree of expertise or familiarity with a domain and with search tools. One way to support different levels of familiarity with tools is to allow expert searchers to customize their interfaces with more sophisticated functionality. For example, one person may use an “advanced search” interface, while another uses a simple text box. Asymmetry in domain expertise may require a more nuanced approach. In SearchTogether, for example, experts’ and novices’ relevance judgments can be treated differently to produce more reliable results.

Search expert-domain expert. This configuration introduces true asymmetries in contributions of team members. This configuration can facilitate collaboration between a skilled searcher and a person with a complex information need, as when Gail works with the reference librarian. The expert searcher knows how to select collections, and how to formulate queries against those collections, but can only make rudimentary relevance judgments based on a description of the information need provided by the domain expert. The domain expert, on the other hand, has a better grasp of the evolving information need, and is able to evaluate retrieved documents.

Another configuration of this role combination pairs a lay person with an information need with an expert domain coach, as when John collaborates with the RN. The lay person acts as a search expert, performing most of the querying and iterative refinement, while the domain expert makes occasional query suggestions and (positive and negative) relevance judgments. Other examples of this collaboration include student (search expert) – adviser (domain expert), and law clerk (search expert) – attorney (domain expert). Suitably mediated, these roles allow people with different skill sets to produce results not easily obtainable independently.

Prospector-Miner. The roles above focused on various combinations of the searchers’ *expertise*, while Prospector-Miner roles [8] focus on searchers’ *activities* during the search. These roles allow one user to search broadly and the other to search deep. The Prospector generates many queries to explore the collection and makes a few relevance judgments for each result set, before moving onto the next query. The Miner makes detailed relevance judgments on documents identified by the Prospector’s queries. One example of these roles is when Gail suggests new search topics for her colleagues. In some ways these roles parallel the search expert-domain expert roles described above, but the specialization into roles is not driven by users’ knowledge but rather by decomposing the information seeking task into subtasks, and developing specialized user interfaces for each subtask.

Other role combinations are also possible, and roles are not limited to pairs. Multiple *Peers* can collaborate trivially, a *Search expert* can mediate information seeking of multiple (collaborating) *Domain experts*, or any number of *Prospectors* and *Miners* can work together.

Looking ahead

Most complex tasks are improved through collaboration. *Ad hoc* collaboration has been observed in a variety of information-seeking situations, but is poorly supported by existing tools. Workarounds such as e-mailing URLs and over-the-shoulder-surfing suggest that a real gap exists between users’ desire to collaborate and the capabilities of the tools they use. While social search has flourished over the past several years, true collaboration is still in its infancy. The dimensions of collaboration and roles described in this article can serve as a design framework for collaborative systems that support explicit collaboration. We expect that tools for explicit collaboration will slowly move from the research domain to specific vertical markets such as legal or medical research and then to subsequent mass-market acceptance.

References

1. Briggs, P. and Smyth, B. 2007. Trusted search communities. In *Proceedings of the 12th international Conference on intelligent User interfaces* (Honolulu, Hawaii, USA, January 28 - 31, 2007). IUI '07. ACM, New York, NY, 337-340.
2. Chi, Ed. Social Search, this issue
3. Golovchinsky, G., Pickens, J., and Back, M. A Taxonomy of Collaboration in Online Information Seeking. In *Proceedings of the 1st International Workshop on Collaborative Information Retrieval*. June 20, 2008. Available online at http://www.fxpal.com/?p=abstract&abstractID=454
4. Linden, G., Smith, B. and York, J. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing,* 7 (1). 76-80.
5. Millen, D. "Better Exploratory Search Through Social Bookmarking" this issue
6. Morris, M. R. 2008. A survey of collaborative web search practices. In *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy, April 05 - 10, 2008). CHI '08. ACM, New York, NY, 1657-1660.
7. Morris, M. R. and Horvitz, E. 2007. SearchTogether: an interface for collaborative web search. In *Proceedings of UIST '07*. ACM, New York, NY, 3-12.
8. Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., and Back, M. Algorithmic Mediation for Collaborative Exploratory Search. In *Proceedings of SIGIR 2008,* July 22-25, 2008, ACM Press.
9. Shah, C. Toward Collaborative Information Seeking (CIS). In *Proceedings of the 1st International Workshop on Collaborative Information Retrieval*. June 20, 2008. Available online at http://workshops.fxpal.com/jcdl2008/submissions/tmpE1.pdf
10. Smeaton, A.F., Lee, H., Foley, C., McGivney, S. and Gurrin, C., Físchlár-DiamondTouch: Collaborative Video Searching on a Table. In *Proceedings of the SPIE Electronic Imaging-Multimedia Content Analysis, Management, and Retrieval* (2006).

Better Exploratory Search through Social Bookmarking.

David Millen

An ongoing challenge facing workers in large organizations is to find valuable information and then to re-find that information when it is needed again. Equally challenging is finding effective ways to share the information with colleagues who have similar needs and interests. Increasingly, large enterprises are experimenting with social bookmarking systems to support workers in these challenging search and sharing tasks.

One early system developed for enterprise use was the *dogear social bookmarking system* [1]. Developed and initially deployed at IBM, workers could bookmark important information resources and easily annotate each bookmark with *keywords* or *social tags*, creating a personal folksonomy over time. Field studies showed that these social tags were often used to browse or rediscover information in a worker's bookmark collection, significantly supporting information reuse.

Colleagues also benefit from these bookmarking applications due to the social nature of their design and use. Design details encourage public (visible) sharing of bookmarks, which can then be searched and browsed by coworkers. Social networks made up of workers with common interests are discernable from both shared tags and bookmarks, and these social networks can then be used to browse, filter or track information resources of mutual interest. Furthermore, early dogear users expressed conscious tagging strategies that benefit coworkers by promoting community or supporting common tag use among teams.

Perhaps even more important than the benefits of bookmark discovery provided *within* the social bookmarking application is the significant improvement in enterprise search enabled by integrating search and the bookmark data. A Firefox plug-in that integrated dogear bookmarks and an enterprise search application was well received by dogear users and offered significantly better search results for enterprise content. In many respects, when an employee bookmarks an intranet resource, they are performing a rating task, signaling explicitly that the resource is interesting and valuable. Follow-on studies within IBM show bookmarked resources are more relevant and more often clicked than other intranet resources. Other approaches to collaborative/social search are highlighted in this issue (see articles by Ed Chi and Gene Golicvhinsky).

Current research is underway to more fully understand the use of social bookmarking applications in exploratory search activities. Social bookmarking applications are a natural way to collect, store, and manage information resources over the longer periods of time typically associated with exploratory search. Better tools are needed, however, to help workers identify important relationships across resources and to detect emerging trends. Integration with other enterprise tools such as new authoring environments (e.g., wikis and blogs) may be desirable. The bookmark collection for a community or organization is an important shared asset and it will be important to find ways to better leverage this asset.

[1] Millen, D., Feinberg, J., and Kerr, B. Social bookmarking in the enterprise. *Queue* 3, 9, November 2005, 28-35.

Author BIO:
David Millen is a research manager at IBM T J Watson Research Center, where his group explores new social software for large organizations. David holds a PhD in Cognitive Psychology from Rutgers University. He is a member of ACM & SIGCHI. Email: david_r_millen@us.ibm.com.

Social bookmarking and information seeking

David R Millen
IBM T.J Watson Research
One Rogers Street
Cambridge, MA 02142
david_r_millen@us.ibm.com

1. Social Bookmarking Systems

In recent years, there has been tremendous growth in shared bookmarking applications. Introduced in 2003, the del.icio.us social bookmark website was one of the first of this kind of application, and has enjoyed an early and large base of committed users. A flurry of similar offerings has since been unveiled [see (Hammond, et al., 2005) for a recent review].

These internet oriented social bookmarking services have been adapted for use in large organizations. Examples include the *dogear* (Millen, et al., 2006) and *onomi* social bookmarking services (Damianos, et al., 2007). Both of these enterprise-ready bookmarking services support bookmarking of both internet and intranet information sources, and provide user authentication via corporate directories.

There are two distinguishing characteristics of social bookmarking systems. The first is the use of keywords, or tags, that a user enters to describe the links he or she saves. These tags allow users to organize and display their collection with labels that are meaningful to them. Furthermore, multiple tags allow bookmarks to belong to more than one category, a limitation of the traditional hierarchically organized folders found in most Web browsers. The second significant characteristic of these social bookmark applications is the social nature of their use. While bookmark collections are personally created and maintained, they are also typically visible to others. As a result, users benefit by getting pointers to new information from others while at the same time getting a general sense of other people's interests.

These new social bookmarking applications are a natural and powerful extension of existing social navigation tools and practices (see, for example, (Dieberger, 2003; Munro, 1999). They provide a mix of both direct (intentional) navigational advice as well as indirect (inferred) advice based on collective public behavior. By definition – these social bookmarking systems provide "social filtering" on resources from the web and intranet. The act of bookmarking indicates to others that one is interested in a given resource. At the same time, tags provide semantic information about the way the resource can be viewed.

Social bookmarking systems arguably provide support for search activities that range from simple fact-finding to more exploratory or social forms of search. Fact-finding or what is called "known-item" retrieval is supported by traditional application *explicit search* capabilities. Users generate query terms and sift through lists of search results to find the appropriate bookmark (and associated web site). These known-item search tasks are usually characterized by a well understood search problem and reasonable understanding of the search domain.

Known-item retrieval is also supported in social bookmarking applications by *browsing* through collections of one's own (personal) bookmarks, which have been explicitly created, tagged and annotated by end-users. Social bookmarking applications typically allow personal bookmark browsing in one of two ways. The first is by sifting through scrollable pages of bookmarks, and the second is by performing a *tag query* of the collection by clicking on a tag.

Social bookmarking tools also support *exploratory* search activities. In exploratory search, the problem definition is less well structured and the emphasis may be on learning or analysis (Marchionini, 2006). One form of this less goal-oriented browsing found in social bookmarking applications is to browse bookmarks by *time*, enabling end-users to serendipitously follow recent bookmark that they find interesting. A second exploratory browsing strategy supported by social bookmarking applications is to explore *popular* bookmarks, where frequency of bookmarking a specific URL is a simple measure of popularity.

Another particularly interesting form of exploratory search supported by social bookmarking services is where end-users engage in person-centric browsing by clicking on a clickable name and the bookmarks for that person appear. This kind of browsing enables the reputation of the original bookmarker to be considered when viewing the bookmark result. At the same time, this person-centric search allows a form of people sensemaking to occur.

These uses of social bookmarking in exploratory search have been reported in a large-scale study of use in a large global corporation (Millen, Yang, Whittaker & Feinberg, 2007).

2. Research Challenges

There are a number of important research challenges for social bookmarking systems. Several are described here:

2.1 Adoption and sustained use of social software

If social applications like social bookmarking are to be useful in general information seeking tasks, then it is a necessary prerequisite that the applications are inherently useful and have a strong user base. The extended valueof the application lies in the vibrancy of the community. As with any social web application, there are questions about how these kinds of tools are adopted and how communities of use develop. Models of use that inform system designers about different task or role patterns are needed and would enable better application design and point to desirable areas of innovation and new functionality.

2.2 Social tag management

The use of social tags are arguably the most novel aspect of social bookmarking applications. Tag recommenders could help increase tag production and help with more efficient bookmark search and exploration. While some work has been done to understand tag and folksonomy development, there remain a number of interesting questions about how to optimize these vocabularies for various search tasks.

2.3 Integration of social tagging with traditional search applications

Social bookmarking applications have been integrated with enterprise search with considerable success (Millen et al, 2007). However, the use of social bookmarking to improve both intranet and internet search engines is an area of research that is important and poses significant challenges. For example, research into novel search algorithms that fully harness the value of social tags to improve search results is needed.

3. REFERENCES

[1] Damianos, L. E., Cuomo, D., Griffith, J., Hirst, D. M. and Smallwood, J. (2007): 'Exploring the Adoption, Utility, and Social Influences of Social Bookmarking in a Corporate Environment'. In 40th *Hawaii International Conference on System Sciences* Waikoloa, HI, January, 2007, pp. 86-86.

[2] Dieberger, A. a. G., M. (2003): 'CoWeb - experiences with collaborative Web spaces.' In C. a. F. Lueg, D. : From Usenet to CoWebs: Interacting with Social Information Spaces. Springer-Verlag, New York, 2003, 155--166.

[3] Hammond, T., Hannay, T., Lund, B. and Scott, J. (2005): 'Social bookmarking tools (I) A General Review', D-Lib Magazine, 2005, ISSN: 1082-9873.

[4] Marchionini, G. (2006): 'Exploratory search: from finding to understanding', Communications of the ACM, 49, 4, 41-46.

[5] Millen, D. R., Yang, M., Whittaker, S. and Feinberg, J. (2007) Social bookmarking and exploratory search. ECSCW 2007, Limerick, Ireland, Sept 26-28, 2007.

[6] Millen, D. R., Feinberg, J. and Kerr, B. (2006): 'Dogear: Social bookmarking in the enterprise'. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Montreal, Quebec, Canada, April, 2006, pp. 111-120.

[7] Munro, A. J., Hook, K., and Benyon (1999): Social Navigation of Information Space. Springer, London, 1999.

Designing Information-Seeking Support Systems

Ryen W. White, Microsoft Research

Tools have been leveraged by mankind for millennia; from the rudimentary stone implements used to hunt and harvest, to the navigational aids that facilitated world exploration, the sophisticated machinery that drove the industrial revolution, and the electronic hardware and software that catalyze the ongoing scientific and information revolutions. The development of tools has affected fundamental change on the nature of human existence. In today's information age, a significant portion of the world economy centers on the manipulation of information rather than the harvesting of crops or the production of material goods. Information may be a valuable resource, but to realize its value it must be used effectively by individuals, workgroups, and organizations in support of discovery, innovation, and decision making. To this end, citizens require sophisticated tools that use state-of-the-art technology from a range of computer and information science disciplines to help them find the right information, but also to help them to understand how to more effectively use that information to affect transformational change on personal, organizational, and societal levels.

Existing information retrieval (IR) systems are effective at finding potentially relevant information fragments in a corpus given indications of user intent. Commercial search engines created by companies such as Google and Microsoft index and search billions of Web pages in a fraction of a second in response to a user's request. It seems that when the target of the search is well defined the challenge of finding relevant information is close to resolution. However, many information seekers have ambitious search quests, especially as they move into emerging domains of knowledge such as sustainable energy, healthcare delivery, or disaster response. To truly help people we must journey beyond basic information finding and toward the attainment of high-level learning objectives such as analysis, synthesis, and evaluation. Well-designed information-seeking support systems (ISSS) empower people to address challenges in areas such as education, workforce productivity, disaster management, terror prevention, and energy conservation. While not every user will experience the thrill of making a historically-important discovery, effective information seeking tools will enable more people to be creative and collaborative in their work and help develop an enlightened society.

The design of ISSS was discussed in detail at the workshop during targeted breakout sessions and workshop-wide discussion. Some of the key challenges to be addressed in designing ISSS that emerged from these activities are summarized below.

Finding

Today's search systems provide adequate support for information finding through keyword-based access to document collections. However, searcher-defined queries are based on searcher's existing knowledge and create limited opportunity for exploration or discovery. Techniques such as relevance feedback (RF) (c.f. Salton and Buckley, 1990) can help searchers choose additional query terms by first eliciting examples of relevant items. ISSS must offer query (re)formulation assistance and where appropriate support personalization (including "groupization" where the actions of similar users are used as a proxy for the individual if personalization is not possible), in support for information behaviors, based on personal and environmental contexts, and including support for different information behaviors, and the ability to recognize and respond to intent.

Facets

ISSS must offer the ability for searchers to filter result sets by specifying one or more desired attributes of the search results. Faceted search interfaces provide a mechanism for people to explore a domain via its attributes. Faceted search interfaces seamlessly combine keyword search and browsing, allowing people to quickly and flexibly find information based on what they remember about the information they seek. They allow flexible navigation, provide previews of next steps, organize results in a meaningful way, and support both the expansion and refinement of the search. Facets can be enriched to incorporate coordinated and dynamic updates useful in understanding relationships between items in a collection or for general exploration tasks. An alternative way to eliminate unwanted information from search results is through negative RF (Cool et al., 1996).

Visualization and Exploration

When exploring an area of interest searchers require a task-appropriate form of data display. ISSS must provide overviews of the searched collection and large-scale result sets to allow information to be visualized and manipulated in a variety of ways. Information visualization and the use of graphical techniques help people understand and analyze data. Information visualization tools provide users with the ability to explore a range of data dimensions seamlessly. These capabilities combined with computational data analysis, can be applied to analytic reasoning to support the sense-making process and exploratory search. Tunable visualizations make it possible to ask questions easily-expressed in a keyword search, and also facilitate rapid refinement of queries with real time direct manipulation. Dynamic query interfaces (Ahlberg et al., 1992) use mouse actions such as slider adjustments and brushing techniques to pose queries and client-side processing, and immediately update displays to engage information seekers in the search process. The tight coupling between queries and results is especially valuable for exploration, where high-level overviews of collections and rapid previews of objects help people understand data structures and infer relationships among concepts, as well as hypothesis generation and information need clarification.

Learning

ISSS must help searchers learn more about the subject area in which they are searching and comprehend the information they encounter. This may be accomplished through the incorporation of expert behavior / knowledge into knowledge base of non-experts, but care must be taken to tailor the search experience to suit the user's expertise level. As well as supporting novices in becoming more expert, ISSS can help those in specialized communities (e.g., health, education, e-science, drug discovery, law) further their knowledge by offering tailored solutions. Historical data comprising many users' search behavior and trails previously followed by expert users (and shared by them) can be mined and leveraged as a way to educate novices about effective search strategies and useful resources. Lessons can be learned from the e-learning and intelligent tutoring communities (Corbett et al., 1997), and users purposely engaged in sustained reasoning activities during browsing. ISSS must also offer topic coverage and controllable result diversity to allow users to learn more about an entire subject area topic or focus their search on a particular subtopic.

Collaboration

ISSS must effectively support interpersonal collaboration. Collaboration is an important aspect of non-digital knowledge building in the workplace and beyond. Techniques such as brainstorming, resource sharing, and division of labor are used to make knowledge workflows more efficient and productive. Pooling cognitive resources may also yield benefits in terms of coverage of the solution space; as more people bring with them ideas on complex problem solving.

Tagging and Data Sharing

Collaboration was previously defined as the simultaneous completion of a work task by multiple individuals, each potentially attempting a different aspect of the task in parallel. In contrast, users can collaborate by annotating shared resources to assist future information seekers in finding and understanding information. As well as creating value added information based on human editing, ISSS must allow the sharing of the results of intellectual work, including the process itself. To support data sharing, ISSS should implement common data exchange formats and provide tools to their users for creating new data in the required format or transforming old data. We must decide where shared information will be stored, address intellectual property issues, and preserve privacy.

History

Information seeking episodes typically involve the examination of multiple information sources and transcend many search sessions. Searchers require tools that allow them easily re-visit previously encountered items. Knowing how much of the information space has been explored on a topic and what remains to be seen, is useful for information seekers. ISSS must: (i) offer a smart and structured history, records of paths users followed to get to findings, and easy re-visitation of results, and (ii) keep track of user progress (and alert them if they are straying), remember dead ends, and record what has already been seen.

Notetaking and Collection Building

Many information seeking episodes involve gathering multiple information fragments from a variety of disparate sources. There may be no single document that contains all of the required information. ISSS must contain "workspaces" to support a spectrum of activities, from unstructured note-taking to integrated authoring environments. A well known way of supporting intelligence enhancement is to be able to annotate information for a specific context. Google's Notebook application allows users to collect snippets of content from several Web pages and combine them in a single document. Analysis of the notes taken can help one find gaps in ones notes or research (e.g., find the missing references in my paper), generate summaries or topic overviews, and potentially through a facility to search notes in common, incidentally find other people working on this topic. Notetaking templates for specific tasks can provide structure to notes and help elicit more relevant information than may have been recorded had freeform notes were used.

Novelty and Comparison

ISSS must provide users with new information on request, including answers to statements such as "tell me something that I don't know that I need to know" or "tell me what I should have asked". To do this, ISSS may offer a comparison facility that monitors the content being read and the content available from a number of input streams. Novel relevant information would be automatically detected based on the current context and outliers that may be of interest to the user would be flagged. Users may be interested in what has changed over a time interval, whether new information on the current topic is available, or whether the current document contains any new information or could be immediately ignored (or even hidden from the user). ISSS must have the capability to show conflicting evidence or perform negative search (Garfield, 1970) where the goal is to show that information is not present in the collection.

Integration and Coordination

Information workers typically use of many different software applications in parallel (e.g., office suites, Web browsers, email clients). Support for the coordinated use of multiple tools and the agile transitions between tools and activities are essential elements of ISSS design. Also important is the seamless integration of tools so that the seeking is not a conscious activity and users do not need to leave the current task. This is vital in minimizing disruption to user concentration and workflow.

Engagement and Joy

The information seeking process is inextricably linked to user emotion regarding the outcomes of the search and information encountered during it. Existing search systems provide a satisfactory user experience. ISSS must shape the search experience to make search delightful, joyful, inspiring, and engaging. Systems must consider these positive emotions as well as negative emotions such as user frustration and dissatisfaction as important indicators of user happiness.

Summary

Information is a critically important resource that drives economic and scientific growth. Information seeking is about more than just finding information; how found information is used to better peoples' lives is an important consideration that is often overlooked by system designers. There is an outstanding opportunity for America to assume a global leadership role and gain economic and scientific advantage by researching and developing effective ISSS. In this section we have considered some of the nascent efforts that have been developed around non-keyword search paradigms and challenges that need to be addressed in the next generation of ISSS to shift the focus beyond search and retrieval, and toward grander challenges such as intelligence amplification and scientific discovery. If the challenges laid out were addressed and the systems that address them shared with elite researchers and laypeople alike, we can readily imagine that the process of discovery and innovation would accelerate not only within research laboratories, but also in the wider world.

References

Ahlberg, C., Williamson, C., and Shneiderman, B. (1992). Dynamic queries for information exploration: An implementation and evaluation. In *Proc. ACM SIGCHI*, pp. 619-626.

Cool, C., Belkin, N.J., and Koenemann, J. (1996). On the potential utility of negative relevance feedback in interactive information retrieval. In *Proc. ACM SIGIR*, p. 341.

Corbett, A.T., Koedinger, K.R., and Anderson, J.R. (1997). Intelligent tutoring systems. In Helander, M.G., Landauer, T.K., and Prabhu, P.V. (Eds.) *Handbook of Human-Computer Interaction* (pp. 849-874). Amsterdam, The Netherlands: Elsevier Science.

Foltz, P. and Landauer, T. (2007). Helping people find and learn from documents: Exploiting synergies between human and computer retrieval with SuperManual. In Landauer, T.K., McNamara, D.S., Dennis, S., and Kintsch, W. (Eds.), *The Handbook of Latent Semantic Analysis*, pp. 323-345. Mahwah, NJ: Erlbaum.

Garfield, E. (1970). When is a negative search result positive? *Essays of an Information Scientist*, 1, 117-118.

Salton, G. and Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41(4), 288-297.

Building Knowledge: What's Beyond Keyword Search?

m.c. schraefel

The success of the Web as the main provider of information is indisputable A key to the Web's phenomenal success is intriguingly less in the information than in our ability to find the information it contains. Indeed, the main way we access the Web is via the white box that from a few words seems to read our mind and return a list of links to resources we want. So successful has this approach to finding information become that on the one hand it is difficult to remember how we managed to find any information at all prior to web-based keyword search, and on the other, it's difficult to envision needing or wanting any other tool for information discovery. Successful paradigms can sometimes constrain our ability to imagine other ways to ask questions that may open up new and more powerful possibilities. The Newtonian model of the universe-as-clockworks, for instance, still explains a great deal of physical phenomena. Indeed, one may say it was only some niggling phenomena that were not well described by that model already that begged the question might there be a better model, a different paradigm? Relativity and quantum mechanics, very different way to imagine the behaviours in the manifest world, opened up new ways to make sense of more of our universe.

The success of the search engine may be our Newtonian paradigm for the Web. It enables us to do so much information discovery that it is difficult to imagine what we *cannot* do with it. But how does this paradigm help a busy mom find a better job quickly, effectively, that is a match for her passion and skills? And if that mom could use some extra training to support that skill to get that better job, how would the paradigm bring in that highly-relevant information that is outside the constraints of the keyword search?

In the Information Retrieval and Information Seeking literature, these kinds of more complex, rich information discovery and knowledge building tasks have been modeled in terms of Search strategies and tactics (Belkin, 87). In the relatively recent work classed as Exploratory Search (White, 06), the emphasis has been on harmonizing human-computer interaction design approaches with models of information seeking to develop new tools that will support these alternative kinds of search and knowledge building.

Examples of such approaches include:

- *knowledge building by association*: creating new knowledge through building associations between one domain/concept with another (Golivchinsky, 97), rather than by seeing "an answer" in any one item.
- *wanting to explore a domain without sufficient knowledge of the domain*. Someone who is not an expert may look for one piece of information without realizing that another component, not matched by a keyword search, is highly relevant.
- *annotations and notes.* A well known way of supporting knowledge building is to be able to annotate information for a specific context. For instance, "The socket described worked well for this project but was miserable for this other - despite what the authors claim here" Similarly being able to create notes ABOUT something and add references easily from related sources is another powerful knowledge building technique

- *Collections.* Pulling together information resources as they are discovered for future knowledge building, as part of information triage (Marshall, 97) is another approach for developing knowledge
- *History Review.* Interrogating both previously looked for information as well working back through the paths taken to that information.
- *Collaborative knowledge building.* brain storming, shared search, shared component development are examples of such knowledge building

Each of these approaches to knowledge building involves exploration of information that yes, pull together a wide array of information resources, but that have less to do with specific iterative searches for a particular pre-existing answers, than support for the development of a New Answer through the interrogation and association of these sources. To support these different kinds of knowledge-building goals, we need to develop the tools that will support these kinds of approaches to exploration. This article looks at some nascent efforts that have been developed around these alternative search paradigms.

Exploratory Search Tools to Date

The pre-history of Exploratory Search can be seen in the raison d'etre of hypertext: to support human-made associations through knowledge spaces. Nelson (Nelson, 81), who coined the term "hypertext" in 1965 was inspired by Vanevar Bush's close-of-WWII vision of the Memex (Bush, 45). The goal of the Memex was to support better knowledge management of a post war Science Explosion by helping scientists build, maintain *and share* their own paths through the document space. Bush called these paths Trails. He postulated that these human-made Trails of associations would be more meaningful for scientific discovery than having to track up and down through library taxonomies of texts. Nelson took Trails and imagined what was to become the key component of the Web: the Link, the ability to "transclude" or connect by reference into a new document both one's own thoughts with others' work to develop a perpetual exchange of ideas. A key attribute of the hypertext link was to support non-linear exploration of information for free form association building. 15 years later, prior to the networked web, Trigg's Notecards system of 1984, put Elnglebart’s NLS on steroids via somewhat richer visualizations of the types of linking functions already described in NLS (Halaz, 87). While most hypertext researchers point to Trigg’s formalization of link types as his key contribution, from an HCI perspective that he chose the note card as the metaphor for his system is for our purposes significant. The card paradigm would later be developed into spatial hypertext (Marshall, 97) to support not just a temporal model of seeing one card at a time (a limit of 1984 display systems) but of being able to support the cognitive model of presenting information akin to the layout and re-organization of cards in a physical world in order to build new knowledge through the association of this information.

Some take-aways from these pre-web representations of knowledge building across automated resources is that Search as keyword search has been largely absent from the main visions of these systems. Perhaps it was simply assumed as a rudimentary tool/strategy such as rooting through the various categorizations of a card catalogue, but it seems important to realize that strategies such as recovering the path through a document space from start to goal (Trails) were seen as critical. Likewise visualizations that privileged non-linear, non-temporally restricted representations of information (such as operations that can be carried out with notecards - stacking, sorting, selectively displaying, sharing, tagging) were also seen as key parts of information building and communication of that information. And then the Web happened.

This pre-history of current Web-based exploratory search approaches is likewise important because it motivates a kind of *recherche du temps perdu* - we have been here before, asking how to best enable knowledge discovery - not as fact retrieval but in terms of how to support and enhance that retrieval for building new knowledge. With the astounding success of the web, however, we occasionally demonstrate a kind of amnesia about what we once sought to achieve. Part of this amnesia may be driven by a similar kind of Newtonian Model success: we've gotten so much out of this approach so far, why not keep digging away at it, push *its* limits?

Early Web Serendipity and Serendipity Redux

One of the celebrated features in the early days of the web was "surfing"- to come upon information serendipitously. The lack of a powerful search engine made this navigational hit and miss approach to information finding on the early web a feature rather than a bug. It accelerated exponentially any kind of equivalent effort in the physical world of grabbing a book, checking references, finding those references physically, grabbing and scanning them, and perhaps discovering something tangentially, serendipitously of interest along the way.   What has happened to web surfing? The scale of the web has grown so profoundly that surfing has been largely replaced by text search interspersed with reference to select sources of mediation, such as blogs, RSS feeds and social networks: we leverage each other's serendipity.

We *serendip*, to coin a term, within a smaller set of known resources and search with intent for particular answers. We might be said to surf less broadly as a result. Part of this loss may be attributed to what happened to another exploration tactic of the early web in the rise of text search. Category engines fell into disuse. While requiring human intervention to construct domain categories, their advantage is, like looking at library shelves, to give a person a sense of the scope of an information space. An opportunity for knowledge building in foregrounding such attributes of a space is the ability to foreground associations among these attributes. Exploratory Search paradigms over the past five years have largely focused on how to faciliate this kind of associative view. These models have become known as Facetted Search.

Facetted Search: the Metadata is the Message

Whereas a keyword search brings together a list of ranked documents that match those search terms, the goal of a facetted search is to enable a person to explore a domain via its attributes. One of the most well known examples of such a browser is Apple's iTunes application (Figure 1) which is an interface to access and play back tracks or sets of tracks from a collection of music files.

Figure 1. Apple iTunes Browser

The browser to the collection presents three columns, representing three facets of the Music domain: genre, artist, album. Attributes matching these facets are populated into the columns. A selection in any column acts as a filter on the column to its right. Once a selection is made, and the right column(s) filtered, a list of individual tracks matching those selected is presented in the lower most browser pane. Keyword search is integrated into iTunes such that the list of data matching the search terms populates the facets in the columns as well as returns a list of individual track results. This layout means that even after the keyword search results are returned, the facets can be operated upon to further explore the collection. If results returned cover multiple genres it is easy to highlight those instances that are associated with a given artist, genre or album.

Exploration by facet enables one to make new connections about a domain or its attributes within a domain. One might, for instance discover that someone perceived to be a Jazz artist has also recorded Country music, which may lead one to explore Country music - something previously thought to be of no interest. This same ability to reconsider a domain via attributes also supports creating new knowledge about the domain: a person may not *know* that these attributes are a way of interpreting a domain. Exposing these facets may implicitly help build domain knowledge.

Enriched Facets. Another attribute of note in most store examples is that quantity is also represented. The facets not only provide the categories of sweater possible, but how many of each there are. In a sense this is reminiscent of seeing the number of books on a shelf for a particular topic: we immediately get a greater sense of the domain from this simple cue.

A facetted browser that has made particular use of representing quantity is the Relation Browser, (detailed in the sidebar). The compelling attribute of the Relation Browser is the integration of multiple types of information against a single facet. Histograms reveal totals of single facets, but selections of other facets immediately show another histogram within the histogram to show how that query will effect the space. These light weight information markers like numbers and spatial cues from histograms provide glanceable, additional attributes about an information space that are not available from keyword search alone.

Backwards Highlighting (Wilson, 08) in the mSpace browser (http://mspace.fm) is a similar to support multi-directional links across facets (Figure 2) in what otherwise appears to be a directional browser like iTunes. In iTunes, a selection in the middle or left column only filters to the right; it does not populate back to the columns to the left of that selection. Picking the artist "radiohead" in other words does not show with what Genres that band is associated. Backwards highlighting shows both the filter to the right as well as the possible paths that could be associated with that selection from the left. In the example of a newsfilm space below, where the facets are decade, year, theme, subject and story, a person has picked the 1940's in the leftmost column. The columns to the right are all filtered by that choice. They next choose a Theme in the third column. The effect of this selection is both to filter the remaining columns to the right, but also to highlight two items in the Year column to the left from which the selected third column item is related. The intensity of the highlights also shows a person which attributes were deliberately selected (the bright highlight) and which were calculated (the duller highlight). These simple information guides have been shown to assist both recall and descriptions of information in a domain.

Figure 2. mSpace slice demonstrating backward highlighting

Making Sense of the Facets themselves. Another sense making attribute that can be associated with an individual item in a facet is a Preview Cue (Figure 3). One may not know a sonata from a symphony, but associating appropriate music with each will at least give the person an example of each as well as a way to decide if it's of interest to them – do they like what they hear? . Once the samples are triggered the person can either step through those samples, or based on the first one played decide if they wish to explore that area of the domain further, or move on.

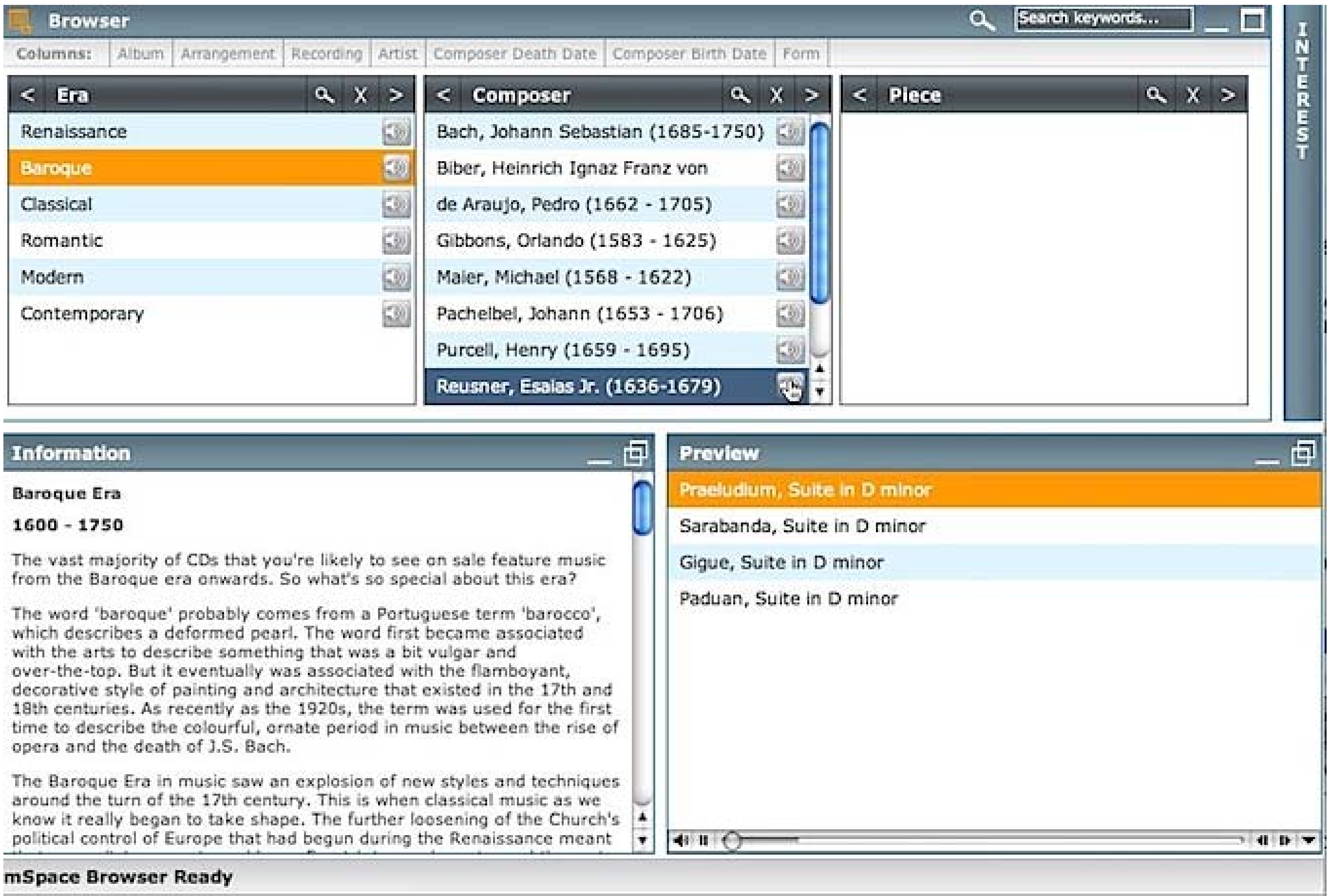


Figure 3. mSpace preview cues with in-context information about selected term.

In the image above, hovering over the Speaker icon has triggered a preview cue for the Baroque Composer Reneau. Three selections by the artist are also cued up in the preview cue. Note also that where Baroque in Period has been selected, a description of the selected facet is presented. Likewise, to help develop an understanding of the domain, when an item associate with a facet is selected, information about that facet is presented.

mSpace refers to the presentation of facets as a "slice" through a domain space, and enables the facets in the slice to be reordered, as well as enabling other facets to be added or removed to a slice.This ability to reorganize a slice according to a person's interests was motivated by the desire to enable a person to explore a domain by what is relevant or known to them: to enable them to have more facility to make sense of a domain in ways that are meaningful to them. In the newsfilm world for instance, one may be more interested to organize a space around the work of a particular reporter than around a particular topic.

Visualizations to Enhance Representations for Knowledge Building

While the above discussion has highlighted the simple ways in which information facets can be decorated to enable rich exploration of a domain, mash ups have also shown us the value of re-presenting those attributes across a variety of visualizations. Exhibit (Huynh, 07) is an example of a tool that provides facetted exploration of data along with visualizing that data against maps and timelines. The value of these representations is in the questions they foreground for interrogation. The Presidents facets makes it easy to see at a glance that most Presidents were born on the eastern side of the US. That Cleveland was the last president to hold office completely inside the 19th Century (MacKinley bridges 19th and 20th C).

Projects like LifeLinesII (Wang 08) have taken larger sets of data such as patient's health records and medical test results, mashed them up, in order to enable medical professionals to align, rank and sort them according to the attributes available on the data (Figure 4). This visualized and parameterized data readily facilitates seeing where there might be correlations across populations in responses to a combination of condition and drug. While IBM's manyEyes (http://manyeyes.alphaworks.ibm.com/manyeyes/) shows the value of being able to share visualizations of data quickly for powerful analysis, by adding manipulable facets onto the visualization. For instance, one can align all patients by heart attack in order to surface significant patterns that may precede or follow such an incident.

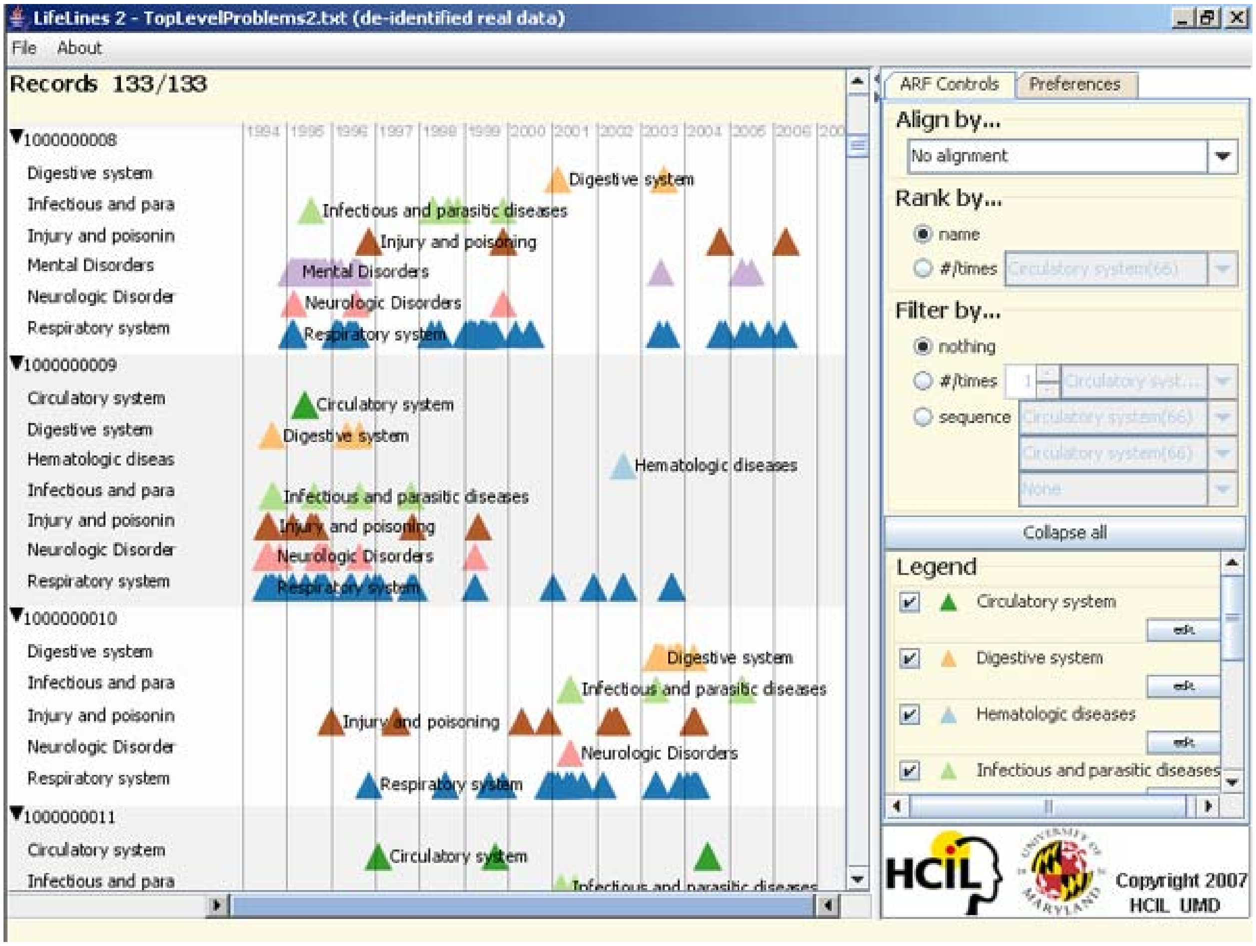


Figure 4. LifelinesII interface, showing attributes to be aligned.

Facetted browsers and tunable visualizations make it possible to ask questions not easily-expressed in a keyword search. They also facilitate rapid refinement of queries with real time direct manipulation. Spatial layout of the data's attributes for manipulation allows relationships within the data to remain available for rapid comparison. All these types of interactions are components of enhancing our information seeking and knowledge building practice.

Moving from Data Manipulations to Tracking New Ideas

We have seen the use of tags-as-annotation as one strategy to enhance the personal or social network value of found things: a tag helps gather that artefact into many potentially relevant contexts (see the Chi paper in this issue). Indeed, the popularity of online photo tagging has shown that people *will* add metadata to their data. Indeed the social sharing value that tags enable,has high value: someone on the team found this thing relevant to our work; now all of us can get it quickly by hitting that tag. Projects like Folksonomies are considering how more structured taxonomies may emerge from these flat spaces in order to add the value of semantics to these annotations.

The authors of SpartTag.us (Hong, 08) make the case that much of the Web's available content, from news articles to blog posts, is frequently reprinted verbatim. SparTag.us enables not only notes to be associated with a Web page and shared, but these notes can automatically show up anywhere online the document may be cloned. To reduce interruption while exploring and make it possible to capture quickly components of a web, Hunter Gatherer (schraefel, 02) enabled components of Web pages to be captured into Collections. Drawing on earlier hypertext ideas and modern graphics processing, work by Donetcheva and Drucker takes the collection notion and enables each component captured to be laid out as an individual card (Donetcheva 07). LiveLabs recent version of this project, Thumbtack (http://thumbtack.livelabs.com), adds machine learning processes so that extracted addresses from a collection can be automatically mapped; books can be explored via extracted author or genre information, and cars by price, engine size, model and so on. New techniques in machine learning are making such semantic extraction increasingly tractable across an increasing area of data.

Whither the Note Book, History and what i don't know i need to know?

At a recent NSF workshop on Information Seeking, two components that the discussants kept resurfacing as critical for exploratory search were History and Note Keeping. An expressed desire was for tools that would help surface things we should know about if and when we're looking at a given topic.

For history currently, we have the History list of our browsers. In mSpace, when someone shares an article with another person, they also share the state of the facets to get to that artefact so a larger context of discovery is available. Going outside the context of a single application, the Jourknow project (http://projects.csail.mit.edu/jourknow/) proposes being able to use local computer context to associate and recover information across personal facets like location, date, and applications to support questions like "what pages was i looking at when i was in the cafe last sunday?" The philosophy beyond Journknow is that any process might inform any other process of interrogation and discovery: how can we make them available to each other for exploration?

These questions lead us back to issues around how we capture and reflect upon the knowledge building we are doing? Right now, the main paradigm for exploration is to "go to the web" - via a browser - to trawl for information. Is this the optimal interaction? It seems there are at least two challenges for knowledge building via information seeking while we are working on our own thoughts, taking notes. We may wish to take notes about something while we're reading it - hence being able to select and annotate web documents, as imagined by Nelson decades ago, is as yet uncommon. But likewise we write notes on our own thoughts. Blogging is a popular demonstration of how well writing notes, thoughts or articles is supported - where we can effortlessly add in links to other information. Indeed, with trackbacks, we can also inform those to whom we've linked that a conversation involving their work is underway. Comments on blogs set up meta conversations around the initial seed of a discussion. But blogging is still largely text based. Sure we can link in photos and YouTube videos, but there are many other kinds of data that we might want to reflect upon and share with others (including active code). A scientist may want to gather up scientific data generated from an experiment, add some notes, tie in some data about the apparatus, along with several quotations about the informing theory, all to give as a blog to a colleague to ask "why aren't my results what your theory predicted? On a more casual note, someone has used VIKI thoughtfully to gather considerable data about various digital cameras. In the mix is the camera they've selected to purchase. How would that annotation be captured to be shared? As the data rapidly goes out of date, how might the person share the attributes of their choice to act as a template for a friend's future choice? Backstory (Venolia 08) is a search tool that has been developed to look at some of these issues within a software developer support group. Gathering up web based sources with local resources and notes on contexts of use, Backstory makes it possible to share local knowledge within a team across data object types,

If these kinds of data gathering and sharing tasks for enhanced knowledge building were better supported, we can readily imagine that the process of discovery and innovation would accelerate. As we have seen with search engines, when a process accelerates, such as finding a phone number or a paper or the answer to a "what is it" question, the activities supported by those processes change. If we can do something quickly, trivially now that used to take days or hours, we can move on more rapidly from information seeking to knowledge building.

Enhancing Search with Awareness. A repeated demand at the NSF workshop was, "tell me what i don't know i need to know." Such a challenge goes beyond related recommendations of people who read this also bought that. How might we assess search patterns of “successful” persons in a domain to associate related content on a topic to a search? For instance, if someone is looking for nutrition information. Some strategies involve collaborative search(see paper on collaborative search, this issue); others may be based on mining of search patterns across the web. The design challenges here are significant: how can we surface this kind of valuable associated knowledge that would not show up in a keyword search? How do we reflect back why information of this type was being surfaced? Are there ethical issues around how information is selected to be associated? e.g., people who are interested in explosives might also want to know about offshore suppliers of hydrogen peroxide?

These kinds of challenges are exciting to contemplate. They suggest that there are many more ways in which we already want to be able to find, manipulate, ponder, share and reflect upon information - all with the facility of keyword search, but none of which keyword search addresses. All which are part of the larger space of "information seeking" beyond simple "search."

Belkin, N.J., Croft,W.B. Retrieval techniques, Annual review of information science and technology, vol. 22, Elsevier Science Inc., New York, NY, 1987.

Bush, V. As We May Think. *Atlantic Monthly* July 1945. http://www.theatlantic.com/doc/194507/bush. context histories for shared science.

Dontcheva, M., Drucker, S.M., Salesin, D., Cohen, M.F., Relations, cards, and search templates: user-guided web data integration and layout. Proc. UIST '07. 2007: 61-70.

Golovchinsky, Gene Queries? Links? Is there a difference? Proc. Of CHI, 1997, 407-414.

Halasz, F.G., Moran, T.P., Trigg, R.H., Notecards in a nutshell, Proc of SIGCHI/GI, 1987: 45-52.

Hong,L., Chi, E.C., Budiu, R., Pirolli, P., Nelson, L. SparTag.us: a low cost tagging system for foraging of web content. Proc AVI ’08. 65-72.

Huynh, D.F., Karger, D.R., Miller,R.C. Exhibit: lightweight structured data publishing. Proc WWW07: 737 -746.

Marshall, C., Shipman, F.M. Spatial hypertext and the practice of information triage, Proc. Hypertext, 1997: 124-133.

Nelson, T. Literary Machines, Mindful Press, Sausalito, California, 1981.

schraefel, m.c., Zhu,Y., Modjeska, D. Wigdor, D., Zhao, S. Hunter gatherer: interaction support for the creation and management of within-web-page collections. Proc. WWW '02: 172-181.

White, R. W., Kules, B., Drucker, S. M. and schraefel, m. c. Introduction. Special Issue on Exploratory Search. CACM, 49 4 (2006). 36--39.

Venolia, G. Backstory: A Search Tool for Software Developers Supporting Scalable Sensemaking. Microsoft Research. MSR-TR-2008-13.

Wang, T.D., Plaisant, C., Quinn, A.J., Stanchak, R., Murphy, S. Shneiderman, B. Aligning temporal data by sentinel events: discovering patterns in electronic health records. Proc. Of CHI '08: 457-466.

Bio:
mc schraefel (lower case deliberate), PhD, Human Computer Interaction, Senior Research Fellow, Royal Academy of Engineering, Senior Lecturer, Electronics and Computer Science, University of Southampton, UK, mc+ieee@ecs.soton.ac.uk

Supporting interaction and familiarity

James Allan, University of Massachusetts Amherst
Information Seeking Support Systems Workshop, June 26-27, 2008
Position Paper – Draft for workshop

I am interested in building "information seeking" systems to support expert searchers (not the general populace) in the tasks of collecting, organizing, and monitoring information related to their issues of interest. Examples of expert searchers include political analysts tracking an important issue, researchers checking on the latest developments in some experiments, or information analysts watching for terrorist activities in Europe. Rather than document retrieval, I am focusing on "fact finding", more along the line of several Question Answer tasks, including "complex interactive QA" that was recently run at TREC.

The particular inspiration for this position paper comes from a series of studies carried out by Patterson, Woods, and Roth at the Ohio State University's Institute for Ergonomics, in the Cognitive Systems Laboratory. (I make no claim that this is the only study of this sort; it is merely the immediate inspiration of this paper.) They invested considerable effort in understanding how people, particularly information analysts, cope with problems of data overload. In exploring how humans extract useful information from data, their work points toward some important criteria that a support system needs to satisfy.[1] Based on their observations, I propose two ideas as important for information seeking support systems: (1) assisted information finding and (2) moving from familiar to unknown sources.

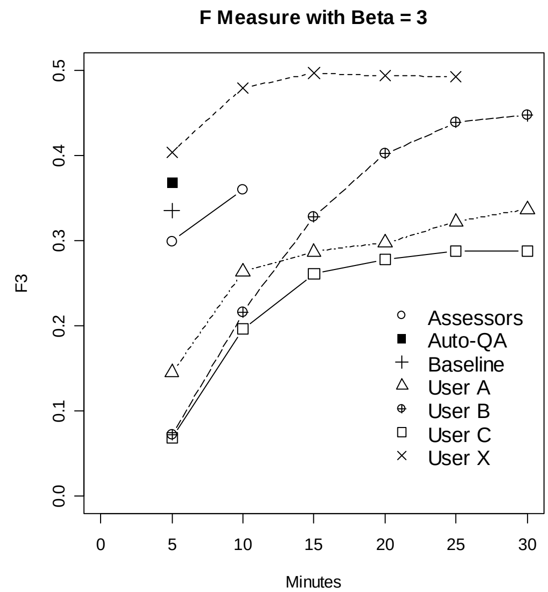
Assisted information finding

The expert searcher I aim to support wants to find a collection of facts and relationships that answers a question. The question is complex, perhaps even difficult to articulate succinctly—certainly not in a 2½-inch field on a Web page. The searcher may know in advance whether there are likely to be a few or hundreds of useful facts, whether some or all are needed, and whether the facts are static or constantly being expanded upon. In such an environment, the classic model of information finding no longer makes sense: an effective system cannot accept a small query, list documents in response, and then wait to see if any are selected. A system designed for this situation must engage with the searcher, working with him or her to develop an expressive query, to extract useful facts, to identify their relationships, and then build incrementally on the results by repeating the process if necessary. (This is probably *Motherhood and Apple Pie* to the attendees of the workshop.)

Existing search and organization systems tend toward one of two extremes. The most common approach is the simple search engine, exemplified by all major Web search engines. They accept a query and present a list of documents ranked by the likelihood of relevance. They might remember and be able to leverage past queries, and might be able to recommend documents that were interesting to

[1] D.D. Woods, E.S. Patterson, and E.M. Roth (2002), "Can we ever escape from Data Overload? A cognitive systems diagnosis." *Cognition, Technology, and Work*, Springer London, 4(1):22-36.

other people. They do not work in conjunction with the searcher. At the other extreme are powerful visions of systems that do everything for the expert: given a query, they are intended to return precisely the desired list of facts. These systems are the goals of government funded initiatives such as DARPA TIDES, DARPA GALE, and the ARDA/DTO AQUAINT programs. The most successful work in those areas can answer simplistic questions well and complex questions with widely varying but limited success.



We have recently developed and evaluated (as part of TREC 2007's ciQA track) an interactive search system designed for finding answers. The system allows the searcher to describe a complex information need (though participants only used free text descriptions), to see documents that were on the right topic, and then to identify answers to the need. The graph on the right shows search effectiveness on the vertical axis and time on the horizontal. The best performance (across the entire evaluation), regardless of time on task, was achieved by an expert searcher ("User X") familiar with the system. Using a system designed to help a trained searcher find answers to questions can result in substantial gains over purely automatic techniques. However, major challenges remain.

Two key challenges are helping the searcher develop a framework for organizing his or her thoughts and using that framework both to help organize results and to help formulate queries. When the expert searcher knows what sort of result is likely to be correct in advance, the search approach can be tuned to expand or restrict the scope of the search, to look for supporting evidence or counter-examples, and to develop queries based on positive or negative examples. Our prototype system provides a rudimentary version of the needed framework, but does not use it to assist with the actual search process. I suspect that relevance feedback may be able to leverage negative information more usefully in such an environment.

From the familiar to the unknown

The 1999 study by Patterson et al. shows a common practice of analysts faced with daunting amounts of raw data. When trying to find their answers, they tend to "narrow returned sets based on the number of hits almost indiscriminately…." [2] We and others in the research community have similarly found that searchers prefer a sense of control, whether it is of the data being explored or of the language used to describe requests. Most implemented search systems address this issue in limited ways, if they do at all. They might provide very simple search capabilities to avoid overwhelming the searcher—e.g., Web search engines allow keywords with a handful of rarely used operators—or they might construct

[2] E.S. Patterson, E.M. Roth, and D.D. Woods (1999), "Aiding the Intelligence Analyst in situations of data overload: a simulation study of computer-supported inferential analysis under data overload." Interim report to the Air Force Research Laboratory, AFRL-HE-WP-TR-1999-0241, p. 77.

elaborate visualizations intended to provide better understanding, hence control—e.g., color bars used to identify which query terms appear in a document and where. Neither approach actually recognizes that searchers are more comfortable with smaller amounts of data.

I believe that a key challenge for information seeking support systems is developing techniques and interfaces that allow a searcher to work in three stages: (1) facilitate rapid and appropriate winnowing of material to a small and comfortable set of items, (2) support exploration and organization on the small set, and (3) enable expansion back to the original large data set to find information that supports, refutes, or builds on what was found in the small set. Each of those stages requires investigation of techniques that work for searchers and algorithms that support the data processing efficiently and effectively.

From Web Search to Exploratory Search: Can we get there from here?

Peter Anick
Yahoo! Inc.
panick@yahoo-inc.com

1. INTRODUCTION

A common complaint of users of search engines is that they deliver too many results. In a recent study of biology teachers, for example, 34% of respondents indicated "too many results" as a primary reason for not using search engines (Perrault 2007). At first blush, this seems like an odd criticism. After all, the results are ranked by relevance and one can always narrow the results by refining the query. It is likely, therefore, that this complaint really reflects the frustration of users when they are engaging in exploratory searches. When a user knows the domain, it's easy enough to assess the relevance of documents via titles and snippets or refine a search that is too broad in scope. But when the user is less familiar with the subject matter, such assessments and refinements can be considerably more difficult. As Marchionini (2006) has noted, exploratory search is more about learning and investigating than lookup.

For this reason, developers of exploratory search interfaces have looked to devices that help users organize and scan documents. Kules and Shneiderman (2006), for example, advocate organizing web search results by "meaningful and stable categories" while clustering interfaces (e.g., clusty.com) partition result sets into categories constructed on the fly. In this position paper, we consider how some of the tools already developed for general web search, such as terminological search assistance and snippet generation, can support the learning and investigative functions of exploratory search, given that a user engage with them from the proper perspective. Terminological search assistance, such as that offered in Yahoo's Search Assist tool (described in the next section), does not provide stable categories, nor are the terms necessarily "meaningful" to a user unfamiliar with the domain. However, term suggestions can provide a convenient way to navigate through a dense and unfamiliar result set when used in conjunction with document snippets, the short fragments of web pages generated to provide a succinct summary of each of the top matched results. Snippets are finely tuned to best express how each document is relevant to the query. Ideally, they contain fragments of text that capture the query terms in close proximity and span just enough lexical context to allow the user to glean the sense of the terms.[1] These aspects of snippets make them quite useful in the "learning" phase of an exploratory search.

[1] Many users have noted that queries posed as explicit questions often produce some snippet containing the answer.

The paper is organized as follows. First, we'll introduce the features of Yahoo's Search Assist tool that make it particularly amenable to exploratory search. Then we'll review two example scenarios. Finally, we'll consider some of the issues and next steps if we were to evolve these web search features into a tool more tuned to supporting exploratory search.

2. SEARCH ASSIST

Yahoo's Search Assist interface (version 1.0) combines two refinement tools within a single display tray. We will refer to these tools as *Gossip* and *Viewpoint*. *Gossip* offers suggest-as-you-type expansions for a user's query string on a character by character basis (as displayed in the left column in figure 1). Its suggested query completions are derived from frequently occurring queries mined offline from search logs. Clicking on a *gossip* suggestion has the effect of replacing the query in the search box and executing the search.

Once a search has been run, *Viewpoint* offers up related terms which are derived from an analysis of the top search results. Clicking on a *viewpoint* suggestion effectively AND's the clicked term with the current query and generates a new result set. However, in this case, the interface keeps the query and refinement terms separate. Rather than merge the refinement into the query box, the interface instead places it into the left hand side of the tray (under the caption "Showing results containing" in figure 2) and re-populates the right-hand-side with new related terms generated from the revised query. Selecting a new suggested term replaces the previous selection in the composite query, i.e., it ANDs the new selection with the original query. The interface thus aims to help the user deal with the problem of "too many results" by allowing him to pick one item at a time to focus on within those results. Note that this behavior differs from most clustering interfaces, which create a single document partition based on the initial query. By refreshing the terms after each term selection, the viewpoint model effectively enables the user to move about within the domain of the initial query.

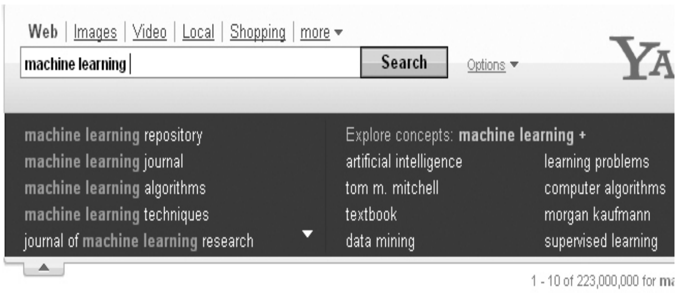


Figure 1. Search Assist tray showing *gossip* suggestions on left and *viewpoint* concepts on right after running a query.

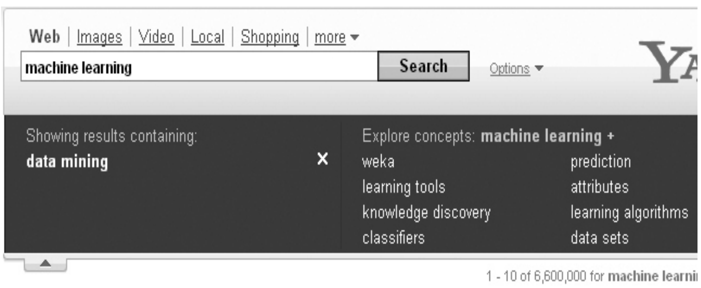


Figure 2. Search Assist tray after clicking on *viewpoint* suggestion "data mining" in right hand side of figure 1.

3. FOCUS AND SKIM

In this section, we present two short hypothetical scenarios to show how the tool may be used to support an exploratory search strategy of focusing on one term suggestion at a time and skimming snippets to quickly glean domain knowledge prior to (or even in lieu of) visiting any actual documents. For each example, we provide several *viewpoint* "focus" terms that might logically be selected and show samples of the snippets returned as a result of each search.

Example 1. Having heard a Moroccan *rebab* player at a recent conference, the user is curious to learn more about the instrument and the contexts in which it is played. He knows next to nothing about Moroccan music.

Initial query: **rebab morocco**

Focus term 1: "**Moroccan music**"

Snippets:

> **Moroccan music** is of many types; it includes Arab, Berber, classical, and **...** instruments - the **rebab**, played by the leader and considered **...**

> **...** master musicians performing on the **rebab**, oud, violin, viola, tar, and darbouka. **...** illustrates the Arabic, Judaic, and sub-Saharan roots of **Moroccan music**. **...**

> **Moroccan music** reflects the country's hybrid culture, blending Arabic, African, **...**

Focus term 2: "**Berber Music**"

Snippets:

> **...** instruments - the **rebab**, played by the leader and considered the most important; **... Berber music**, even more closely linked to poetry than Arab music, is usually **...**

> **Morocco's** Berber villages are rich with music. **... Berber Music** Today **...** Raiss instruments typically include the **rebab**, a one-stringed fiddle, not same **...**

Focus term 3: "**ahwash**" (selected because it is an unfamiliar term)

Snippets:

> The Berber **rebab** should not be confused with the **rebab** andalusi, a two **...**

> **... Ahwash** is exclusively village music, probably unchanged for centuries or longer. **...**

Example 2: This example is based on the query used in Michael Levi's position paper, "Musings on Information Seeking Support Systems". We assume the user has little background in economics

Initial query: "**Is the US in a recession?**"

Focus term 1: "**definition of recession"**

Snippets:

> A period of general economic decline; specifically, a decline in GDP for two or more consecutive quarters.

Focus term 2: "**GDP**"

Snippets:

> But what **is GDP**? Gross domestic product is the value of all final goods and services produced in **...**

> Imports, which are a subtraction in the calculation of **GDP**, decreased. **...**

Focus term 3: "**real GDP**"

Snippets:

> **...** the "two quarters" method indicates **the US is** not **in a recession** technically, **...**

> **...** A second point is that **real GDP is** not reported in a timely fashion. **...**

Focus term 4: "**NBER**" (unknown term for this user)

Snippets:

> **...** official judges from **the National Bureau of Economic Research** (**NBER**) pour through various **...**

> **The National Bureau of Economic Research**, an official panel of senior economists, has declared that **the US** entered **...**

Focus term 5: "**economists**"

> 60.8% of **economists** believe **a recession is** here or approaching

> **Economists** Disagree on Whether the Country Actually **Is in a Recession**

4. OBSERVATIONS

These two short scenarios illustrate how a web search user might go about learning domain vocabulary or identifying key facts through interaction solely with term suggestions and snippets. While not all suggested terms lead to relevant information, it is usually possible to judge the importance of unknown terms from skimming the snippets returned, some of which are likely to include either explicit definitions or contextual information that serves to define them.

What might a more sophisticated interface, specifically designed for exploratory search, do to better support a "focus and skim" interaction model?

- Provide longer snippets or a mechanism to expand snippets for more context. As proposed in White et al (2005), it may even be reasonable to organize the display as a collection of top ranked sentences rather than as per-document abstracts.
- Weight term suggestions so as to prefer those that are more domain specific. This is less useful for general query refinement where terms like "images" and "reviews" might make perfectly appropriate refiners, but for exploratory search purposes domain specific terminology should in principle offer the user more opportunities to learn.
- Provide tools for marking and remembering useful snippets. This would allow the user to organize snippets (and hence documents) by their focus terms, generating a dynamic "table of contents" through which to visit the documents themselves in an orderly fashion.
- Help the user choose a suitable initial query. The initial query should be broad enough to circumscribe the domain of interest while narrow enough to eliminate ambiguities or irrelevant aspects that might lead to poor term suggestions.
- Offer more term suggestions on demand, sorted into categories where appropriate.

5. CAN WE GET THERE FROM HERE?

We have argued that some key elements of exploratory search support may already exist in web search engines as a side effect of efforts to support query refinement and search result assessment. It remains to be seen if users themselves will recognize that such features can be exploited in the manner we have suggested. Log analysis reveals some cases of iteration over viewpoint terms (there is a 20% chance that a session containing one viewpoint term selection will contain a second selection) but real life examples such as the scenarios we presented here are difficult to validate from logs alone since it requires determining whether the lack of a click on a document reflects sufficient knowledge gleaned from snippets vs. dissatisfaction with the search results.

What is known is that users have such deeply ingrained habits for their conduct of web searches that it is difficult for them to integrate, or even notice, new features on a search results page. Naïve users are typically reluctant to interact with features whose consequences are uncertain while sophisticated users tend to construct erroneous interpretations of how a feature works and then use it sub-optimally. The greatest challenge in the development of systems to support exploratory search will be to bring the mental model of the user in line with the actual behavior of the tool.

6. REFERENCES

Levi, M. D. (2008). Musings on Information Seeking Support Systems, ISSS 2008 position paper.

Marchionini, G. (2006). Exploratory Search: From finding to understanding. Communication of ACM, 49(4): 41--46, 2006.

Perrault, A. M. (2007). An exploratory study of biology teachers' online information seeking practices. School Library Media Research, 10. www.ala.org/ala/aasl/aaslpubsandjournals/slmrb/slmrcontents/volume10/biology.cfm

Kules, B and Shneiderman, B. (2006). Improving the search experience by organizing Web search results with meaningful and stable categories. HCIL Symposium, University of Maryland, College Park, 2006

White, R. W., Jose, J. M. and Ruthven, I. (2005). Using top-ranking sentences to facilitate effective information access. *Journal of the American Society for Information Science and Technology*, 56 (10), pp. 1113-1125.

Transparent and User-Controllable Personalization For Information Exploration

Peter Brusilovsky
School of Information Sciences, University of Pittsburgh
135 N. Bellefield Ave.
Pittsburgh, PA 15256, USA
peterb@pitt.edu

ABSTRACT

Personalized Web search has emerged as one of the hottest topics for both the Web industry and academic researchers. However, the majority of studies on personalized search focused on a rather simple type of search, which leaves an important research topic – the personalization in exploratory searches – as an under-studied area. In this paper, we present a study of personalization in task-based information exploration using a system called TaskSieve. TaskSieve is a Web search system that utilizes a relevance feedback based profile, called a "task model", for personalization. Its innovations include flexible and user controlled integration of queries and task models and on-screen visualization of task models. Through an empirical study using human subjects conducting task-based exploration searches, we demonstrate that TaskSieve pushes significantly more relevant documents to the top of search result lists as compared to a traditional search system. TaskSieve helps users select significantly more accurate information for their tasks, allows the users to do so with higher productivity, and is viewed more favorably by subjects under several usability related characteristics.

1. INTRODUCTION

It is commonly accepted that lookup search is just one of several types of searches performed by Web users. Marchionini [6] calls searches "beyond lookup" as *exploratory searches*, which can be further distinguished as *search to learn* and *search to investigate*. Exploratory search assumes that the user has some broader *information need* that cannot be simply solved by a "relevant" Web page, but requires multiple searches interleaved with browsing and analyzing the retrieved information. The research on supporting exploratory search attracts more and more attention every year for two reasons. On one hand, the number of users engaged in exploratory search activities is growing. With the growth of information available on the Web, almost any Web user performs searches "beyond lookup" on such occasions as planning a vacation or choosing the most relevant product (i.e., digital camera). Moreover, some classes of users, such as intelligence analysts, perform multiple exploratory searches every day as a part of their job. On the other hand, traditional search systems and engines working in a mode "query – list of results" provide very poor support for exploratory search tasks [6]. Neither is it easy for users to formulate a query when it is not really clear what they are looking for, nor is the result presentation in the form of a linear list helpful to make sense of the retrieved information.

Our team investigated the issue of exploratory search in the context of DARPA's GALE project. Our goal was to develop a better information distillation interface for intelligence analysis. We focused on personalized search expecting that adaptation to an analyst's global task beyond a single query may help our system to bring better results to the analyst's attention.

Personalized search emerged as one of the hottest topics for both the Web industry and academic researchers [7]. Unlike traditional "one-size-fits-all" search engines, personalized search systems attempt to take into account interests, goals, and preferences of individual users in order to improve the relevance of search results and the overall retrieval experience. In the context of a tight competition between search engines and technologies, personalization is frequently considered as one of the technologies that can deliver a competitive advantage.

We expected that personalized search will be appreciated by users engaged in information explorations and will allow them to achive a sizeable performance increase. However, an evaluation of our personalized intelligence analysts discovered that traditional personalized search does not provide the proper level of support in an information exploration context. While appreciating the value of personalization, the analysts repeatedly asked for an interface that provides "more transparency" and "more control" over the search process. Unfortunately, neither transparency, no control are provided by the traditional personalized search systems provide. Personalization works as a black box, which starts a query produces a user-adapted list of results with no direct user involvement. Inside this black box, the personalization engine applies a user profile either to generate query expansion or to reorder search results [7].

In our recent work we explored an alternative approach to implementing personalized search, specifically geared to information exploration context. In our TaskSieve system [2], we attempted to implement personalization as an information exploration tool, which offers the user both: a reasonable control over the process and a better transparency of its mechanism. We consider transparency as an important component of user control: without clear understanding of the process, which is supported by transparency, an effective control is hardly possible. The remaining part of the paper briefly presents the components of TaskSieve interface, which demonstrates our vision of transparent and controllable personalization. The results of TaskSieve evaluation can be found elsewhere [2].

2. TaskSieve: A PLATFORM FOR TASK-BASED INFORMATION EXPLORATION

2.1 A Transparent Task Model

Unlike the majority of known personalized search systems, TaskSieve aims to support the task-based exploratory search process. In place of a traditional model of user interests,

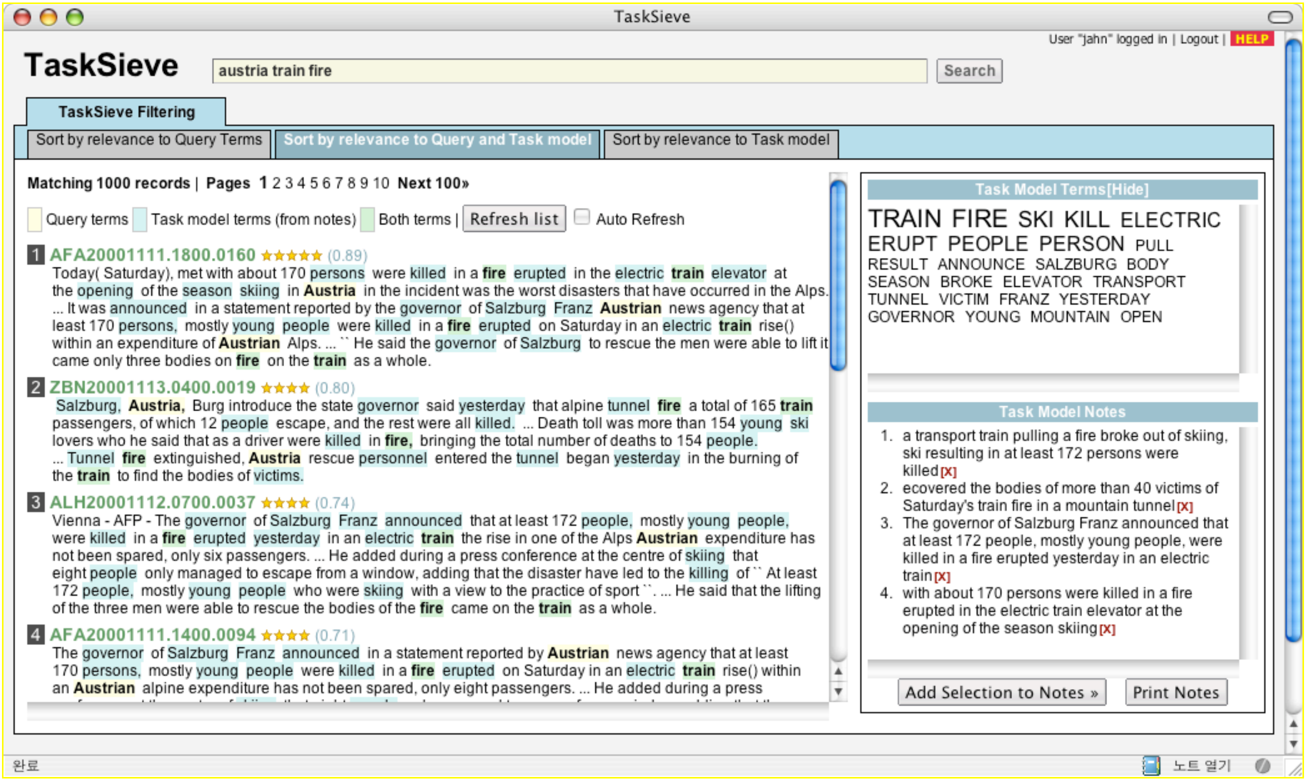


Figure 1 TaskSieve interface

TaskSieve applies a more focused *task model*, which attempts to accumulate information about the task explored by the user. A task model is a relatively short-term model in comparison with a long-term model of user interests, yet it can support the user over a lengthy sequence of queries (frequently spread over several sessions) as long as the user is focused on a specific task. The model is constructed unobtrusively while the users are interacting with the system. There is no task description to enter, as in AntWorld [5] or SERF [4]. The user simply starts working on a new task by entering the first query and processing the list of initial, but not yet adapted, search results. Standard stemming and stopword removal procedures are applied to these task model vectors. Among the hundreds of terms from the user notes, the top 300 important terms are selected according to their TF-IDF weights in the document corpus.

TaskSieve was designed to assist users who perform exploratory searches reasonably often, i.e., it focuses on relatively experienced searchers up to the level of professional information analysts. These users appreciate more powerful and sophisticated information access tools; but as we learned from our earlier work on adaptive filtering [1], they also want to be in control of the system's work and highly value the transparency of the system mechanisms. This requirement contradicts the traditional approach taken by personalized search systems, which tend to make personalization decisions without user consent and hide the underlying personalization mechanism. Unlike these systems, TaskSieve attempts to make the personalization transparent. It starts with using a relatively simple, but easy to understand task model form: weighted term vectors. In addition, it makes the task model visible to the user through the model viewer (upper right in Figure 1). The viewer shows terms, which form the task model, sorted by their importance (weight). A larger font size is used for more important terms. The model visualization is kept up-to-date according to the task model changes. This visible task model is expected to help users to understand the task-based engine of TaskSieve; however, users who consider the model less useful or need more space for other parts of the interface can hide the viewer at any time.

2.2 Controllable Personalized Ranking

As in many other personalized search systems, TaskSieve uses the post-filtering approach to personalized search results, using the task model to re-rank the plain search results retrieved by a search engine (**Error! Reference source not found.**). The idea of re-ranking is to promote documents, which are more relevant to the user task as measured by their similarity to the task model. For transparency reasons, TaskSieve uses the traditional linear approach to combine query relevance and task relevance:

(1) Retrieve documents along with their relevance scores by submitting the user query to a search engine.

(2) Calculate similarity scores between retrieved documents and the model.

(3) Calculate combined score of each document by equation (1).

$$alpha * Task\_Model\_Score + (1 - alpha) * Search\_Score \quad (1)$$

(4) Re-rank the initial list by the combined score from step 3.

TaskSieve uses Indri[1] as a search engine and normalizes its scores, dividing by the maximum score (score of the rank 1 item) of the

[1] http://www.lemurproject.org/indri

corresponding list (step 1). Task model scores are calculated by measuring the similarity between each document vector and the task model vector. We use BM25 **Error! Reference source not found.** for this task (step 2) and the scores are also normalized.

In equation (1), *alpha* controls the power of the task model. It can vary freely from 0 to 1. The traditional approach is to fix *alpha* either ad-hoc, or by learning the "optimal" value and using this value to fuse all search results. We believe this approach contradicts the desire of our target users to be "in control", and instead give the control over the fusion to users. TaskSieve allows the users to alternate among three preset ranking options: "Sort by relevance to Query Terms", "Sort by relevance to Query and Task model", and "Sort by Relevance to Task Model" (which correspond to *alpha* values 0, 0.5, and 1.0 respectively). If *alpha* is 0, the ranking is the same as plain search. If *alpha* is 1.0, then the search rank is completely ignored. If *alpha* is 0.5, which is the default, the system considers equally the importance of query and task.

Figure 1 shows an example of the task-based ranked list (lower left in the screen). A user enters a query "austria train fire". Important task model terms such as "TRAIN", "FIRE", "SKI", and "KILL" were extracted from the user notes in order to re-rank the original search result to the query "austria train fire" generated from the baseline search engine. Just above the ranked list, there are three tabs labeled with three ranking options explained above. Users can explore different query terms and control the task-based post-filtering engine in order to complete their tasks.

2.3 Using Notebook for Task Model Update

In addition to the innovative ways of using the task model, TaskSieve explores a new transparent approach to updating this model. This approach is based on the idea of a *notebook*. A notebook is a collection of document fragments (which we call *notes*) extracted and saved by the user. From one side, the notebook supports the user's need to collect the most important information for further processing. A note collection tool is frequently used in the process of information exploration (analysts call it a "shoebox"). From the other side, the content of the collected notes represents the task much better than the documents from which they are extracted. It allows TaskSieve to use the content of the saved notes to increase the quality of modeling in comparison with existing personalized search systems.

TaskSieve encourages the user to take notes and make this process very simple. The users can highlight any text from the search snippets or whole document and add it to the notebook by a single button click. When a new note is saved, it is displayed in the notebook (lower right in Figure 1). Each note can be removed by clicking on the "X" beside it if the user doesn't think she needs it anymore.

Every action in the notebook (adding and removing) instantly affects the task model – the weights of the task model terms found in the added or removed note are increased or decreased correspondingly. The important task model terms in the task model viewer are immediately updated to reflect the new set of weights. The ranking of the current search result list can also be updated immediately after each task model change if *Auto Refresh* is checked. However, this option is switched off by default, because our previous studies in a similar context of information filtering demonstrated that automatic update of ranking confuses users and causes performance decreases [3]. Therefore, TaskSieve offers a "Refresh list" button, allowing the user to re-rank the search results according to the current state of the task model whenever it is most convenient for her.

3. STUDY RESULTS AND FUTURE WORK

We conducted an empirical study with human subjects using TaskSieve for task-based exploration searches. The study demonstrates that TaskSieve – compared to a traditional search system – can utilize the information available in the task model to return significantly more relevant documents at the top of the ranked lists. The data also show that the average precision values of the baseline system's ranked lists at the last 10 minutes is still lower than that of the experimental system's first 10 minutes. This shows that the improvement obtained through task model is even higher than that through human users learning about the search topic and the retrieval system over the time.

The study also shows that TaskSieve can help user performance, too. TaskSieve's users were not only able to select notes that contained significantly more relevant information, they also can select more notes even during the first 10 minutes of the search session when they were still relatively unfamiliar with the search tasks. This demonstrates that TaskSieve significantly improved the productivity of the users' searches.

The flexibility in controlling the integration mode between queries and the task model also demonstrates its usefulness. First, we observed subjects switching among the different modes in their searches. Second, the searches with the half-half mode produced the best results. Third, the searches in query-only mode produced better results than the baseline, which indicates that the users really mastered the preset manipulations and used the appropriate mode for different searches. Finally, it is clear that none of the modes significantly dominates all the searches. All of these indicate that it really makes sense for TaskSieve to let users decide the best mode for their searches.

4. ACKNOWLEDGMENTS

This paper is partially supported by DARPA GALE project and by the National Science Foundation under Grant No. 0447083.

5. REFERENCES

[1] AHN, J.-W., BRUSILOVSKY, P., GRADY, J., HE, D., SYN, S.Y. Open user profiles for adaptive news systems: help or harm? in Proceedings of the 16th international conference on World Wide Web, WWW '07 (Banff, Canada, May 8-12, 2007), ACM 11-20.

[2] AHN, J.-W., BRUSILOVSKY, P., HE, D., GRADY, J., LI, Q. Personalized Web Exploration with Task Models in Proceedings of the 17th international conference on World Wide Web, WWW '08 (Beijing, China, April 21-25, 2008), ACM 1-10.

[3] HE, D., BRUSILOVSKY, P., GRADY, J., LI, Q., AHN, J.-W. How Up-to-date should it be? The Value of Instant Profiling and Adaptation in Information Filtering in Proceedings of the 2007 international conference on Web Intelligence, WI '07 (Silicon Valey, CA, USA, November 2-5, 2007), IEEE 699-705.

[4] JUNG, S., HARRIS, K., WEBSTER, J., HERLOCKER, J.L. SERF: Iintegrating human recommendations with search in Proceedings of ACM 13th Conference on Information and Knowledge Management, CIKM 2004 (Washington, DC, November 8-13, 2004 571-580.

[5] KANTOR, P.B., BOROS, E., MELAMED, B., MEÑKOV, V., SHAPIRA, B., NEU, D.J. Capturing human intelligence in the net, Communications of the ACM (2000) 43, 8, 112-116.

[6] MARCHIONINI, G. Exploratory search: From finding to understanding, Communications of the ACM (2006) 49, 4, 41-46.

[7] MICARELLI, A., GASPARETTI, F., SCIARRONE, F., GAUCH, S. Personalized search on the World Wide Web, In: Brusilovsky, P., Kobsa, A., and Neidl, W., Eds. The Adaptive Web: Methods and Strategies of Web Personalization. Springer-Verlag, Berlin Heidelberg New York, 2007, 195-230.

Faceted Exploratory Search Using the Relation Browser

Robert G. Capra and Gary Marchionini
School of Information and Library Science
University of North Carolina at Chapel Hill
rcapra3@unc.edu, march@ils.unc.edu

ABSTRACT
The Relation Browser (RB) is a tool developed by the Interaction Design Lab at the University of North Carolina at Chapel Hill for understanding relationships between items in a collection and for exploring an information space (e.g., a set of documents or web pages). It is implemented as a Java Applet that can be embedded in a web page. The Relation Browser provides a dynamic user interface that allow users to explore a data set through the use of faceted browsing and keyword search. The current Relation Browser, RB07, is available at: http://idl89.ils.unc.edu/rb07

1. Relation Browser 07
The Relation Browser has a long history and has been through a number of significant revisions [5][7]. At JCDL 2007 [2], we reported on two studies we conducted to compare three different interface styles (a handcrafted web site, a simple facet interface, and the previous RB++ version of the Relation Browser) for three different task types (simple lookup, complex lookup, and exploratory search) for the U.S. Bureau of Labor Statistics (BLS) web site data. Based on the results of the studies and on expert reviews of the interface, we developed and implemented a set of design changes for the next-generation RB [1][3], dubbed the RB07, while still maintaining a primary goal of providing a tool for exploring data spaces – especially for gaining a better understanding of documents and how they are related to each other.

The Relation Browser has two "hallmark" features. First, all the facet values are displayed "front and center" in the interface, with visual components (bars) that indicate the number of matching documents for each facet value (see Figure 2). Second, the interface supports dynamic queries [6], meaning that as you mouse-over facet values, the display is automatically updated to show a preview of including the moused-over facet value in the current query. This can cause the length of the bars to change as well as the results shown in the results area. These and other features of the RB are designed to support exploratory search [4].

Additional features of the RB07 are described below (numbers correspond to the numbered circles in Figure 2).

1. *Multiple facet views* – The RB07 supports multiple, pluggable facet views. Users can switch between views using tab controls. The "Facet List" view presents the facets and facet values as lists of TriBars (described below) that indicate the number of matching documents. In addition, there is a "Facet Cloud" view (Figure 4) that displays the facet values in the style of a tag cloud, using font size to indicate the number of matching documents. Both the Facet List and Facet Cloud dynamically update on mouse-overs.

2. *Static facet list* – An additional listing of the facets was added down the left side of the screen to provide a constant view of the facets that does not dynamically update (except to change the color of facets in the current query). This is especially useful when working with the facet cloud since the position of the facet values in the cloud dynamically update as the mouse is moved within the cloud.

3. *Multiple result views* – Results of our RB++ study revealed that many users were accustomed to search engine style interfaces that display results in linear list. However, the RB++ compactly displayed results in a grid. The new RB07 allows users to select either the grid view of the prior RB++, or a list view that is similar to search engine result displays.

4. *Current query display and control* – In the new RB07, the current query is displayed near the top of the screen and acts similarly to breadcrumb trails that many web sites use – it shows a history of the facets and search terms entered and provides buttons to remove individual facets or terms. This allows users to quickly remove items from an over-constrained query.

5. *and 6. Full-text search and search within the results* – Observations and results from our RB++ study indicated that many users wanted to employ a "keyword search first" strategy to information seeking tasks. The RB++ did not support full-text search and encouraged a "facets first" approach to exploratory search. In the new RB07, we have sought to support the use and mixing of both strategies by including full-text search and search within the results, while maintaining a prominent display of the facets and document space.

2. TriBar Displays
The list view of the Relation Browser uses a custom user interface element called a TriBar (see Figure 1) that visually displays three pieces of information about each facet value.

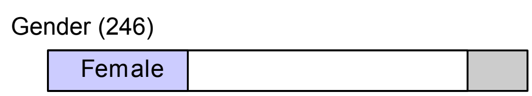


Figure 1. An example TriBar display

The total number of documents that match this facet is shown in parentheses next to the name of the facet (e.g. "Gender (246)"). The full width of the TriBar (the right edge of the grey area) represents the total number of items *in the collection* that have this *facet* (e.g. 246 items have the Gender facet). Second, the right edge of the white area indicates the total number of items *in the collection* that have this *facet value* (e.g. About 90% of the items that have a Gender facet have the value Female). Note that items may have more than one facet value for each facet, as is the case with the Gender facet shown in Figure 2 (many documents discuss both males and females in the same document). Third, the right edge of the purple area on the left side indicates the total number of items *in the current query* that have this *facet value*.

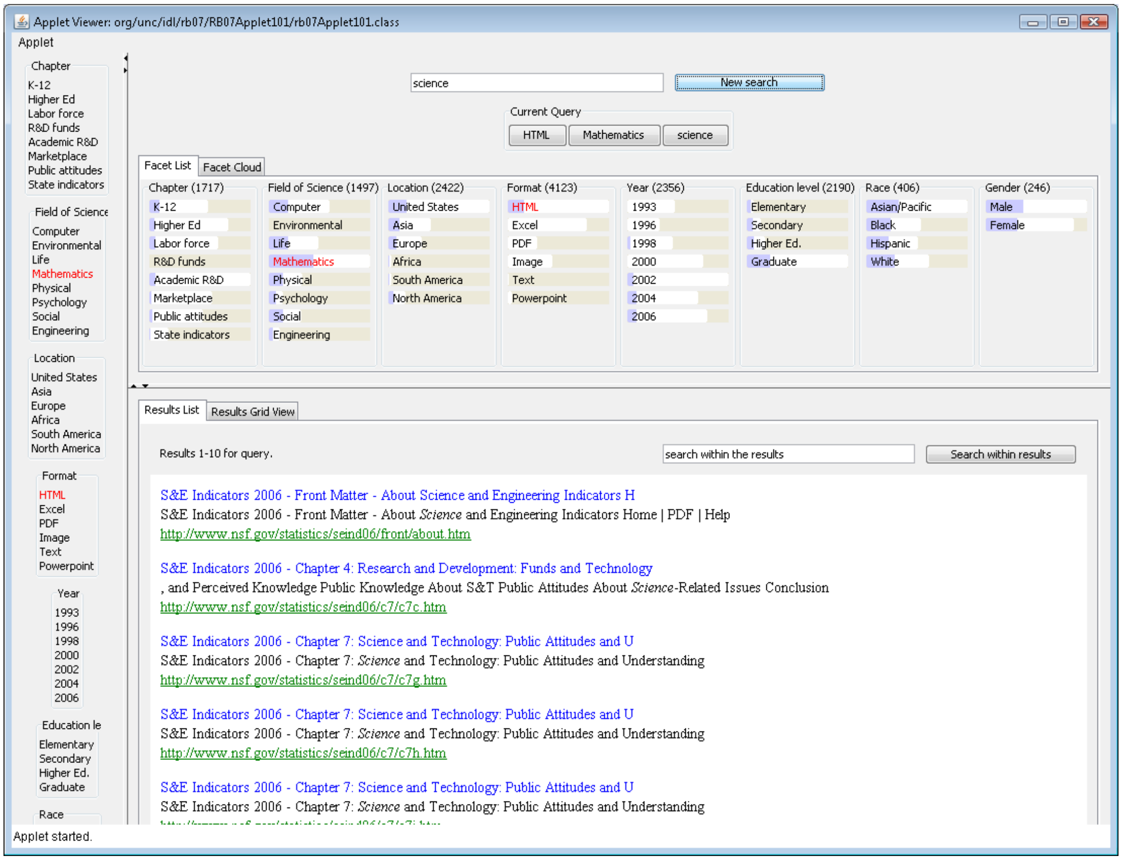


Figure 2. Screenshot of the Relation Browser 07

3. Architecture

The RB07 is implemented as a Java Applet that communicates with an Apache web server and Apache SOLR search engine as shown in Figure 3.

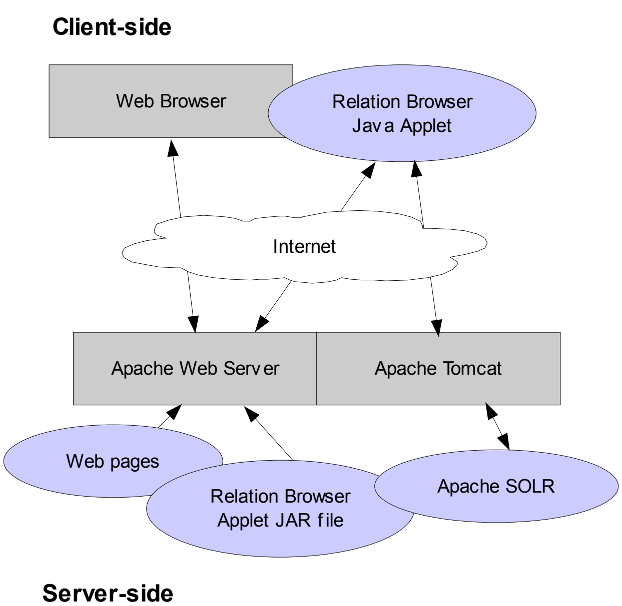


Figure 3. Relation Browser Architecture

When the applet is started in the client browser, it connects to the server and loads the facet data so that it can calculate and display the facet counts in the TriBars. Facet data is stored and transmitted as a compressed bit set to minimize the transfer time from the server to the client. All computation and updating of the facet counts is done on the client so that the dynamic display can be updated smoothly. Keyword search is handled using an Apache SOLR server (based on the Lucene search engine). When a keyword search is entered, the query is sent to the SOLR server and the results are merged (on the client side) with the existing facets in the current query. Keyword search results are cached so that subsequent query manipulations do not require a re-query of the SOLR server.

4. Configuration

Each Relation Browser instance is configured using two types of XML configuration files. The first type of configuration file is an instance file that describes the facets, facet values, and other properties of the Relation Browser. There is only one instance file. An example is shown below.

```
<rbinstance>
  <resultfields>
    <resultfield>docid</resultfield>
    <resultfield>Title</resultfield>
    <resultfield>URL</resultfield>
    <resultfield>Description</resultfield>
  </resultfields>
  <fulltextfields>
    <fulltextfield>fulltext</fulltextfield>
```

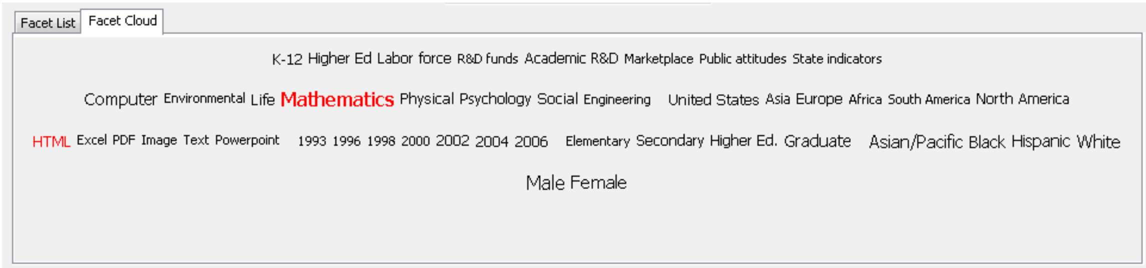


Figure 4. Screenshot of the Facet Cloud View of the Relation Browser 07

```
  </fulltextfields>
 <facets>
   <facet>
     <facetname>Chapter</facetname>
       <facetvalues>
        <facetvalue>K-12</facetvalue>
        <facetvalue>Higher Ed</facetvalue>
        <facetvalue>Labor force</facetvalue>
        <facetvalue>R&amp;D funds</facetvalue>
        <facetvalue>Academic R&amp;D</facetvalue>
        <facetvalue>Marketplace</facetvalue>
        <facetvalue>Public attitudes</facetvalue>
        <facetvalue>State indicators</facetvalue>
       </facetvalues>
   </facet>
   . . .
```

The second type of configuration file are the document description files. There is one document description file for each document in the collection. These files specify what facets, fields, and values apply to each document. An example is shown below.

```
<doc>
   <field name="docid">125</field>
   <field name="URL">
     http://www.nsf.gov/statistics/seind06/c1/c1s2.htm
   </field>
    <field  name="Description">  S&amp;E  Indicators  2006
- Chapter 1: Elementary and Secondary Education          -
Student C</field>
   <field name="Year">2006</field>
   <field name="Chapter">K-12</field>
   <field name="Field of Science">Mathematics</field>
   <field name="Location">United States</field>
   <field name="Format">HTML</field>
   <field name="Year">2006</field>
   <field name="Education level"> Elementary</field>
   <field name="Education level">Secondary</field>
   <field name="Education level">Higher Ed.</field>
   <field name="Education level">Graduate</field>
    <field  name="fulltext">  S&amp;E  Indicators  2006  -
Chapter   1:   Elementary   and   Secondary   Education   -
Student   Coursetaking   in   Mathematics   and   Science
Home   |   PDF   |   Help   |   Contact  Us  Search  Table  of
Contents Overview Chapter 1 Highlights
   . . .
```

5. Future Work

We plan to evaluate the features of the new RB07 using a method similar to the one we used in our JCDL 2007 paper [2] to evaluate the previous RB version. That is, we plan to compare it to a "vanilla" faceted interface, an hand-crafted web site, and a basic search engine interface. We are also interested in evaluating the usefulness and usability of the facet cloud and related visualizations of the facet space.

6. Acknowledgments

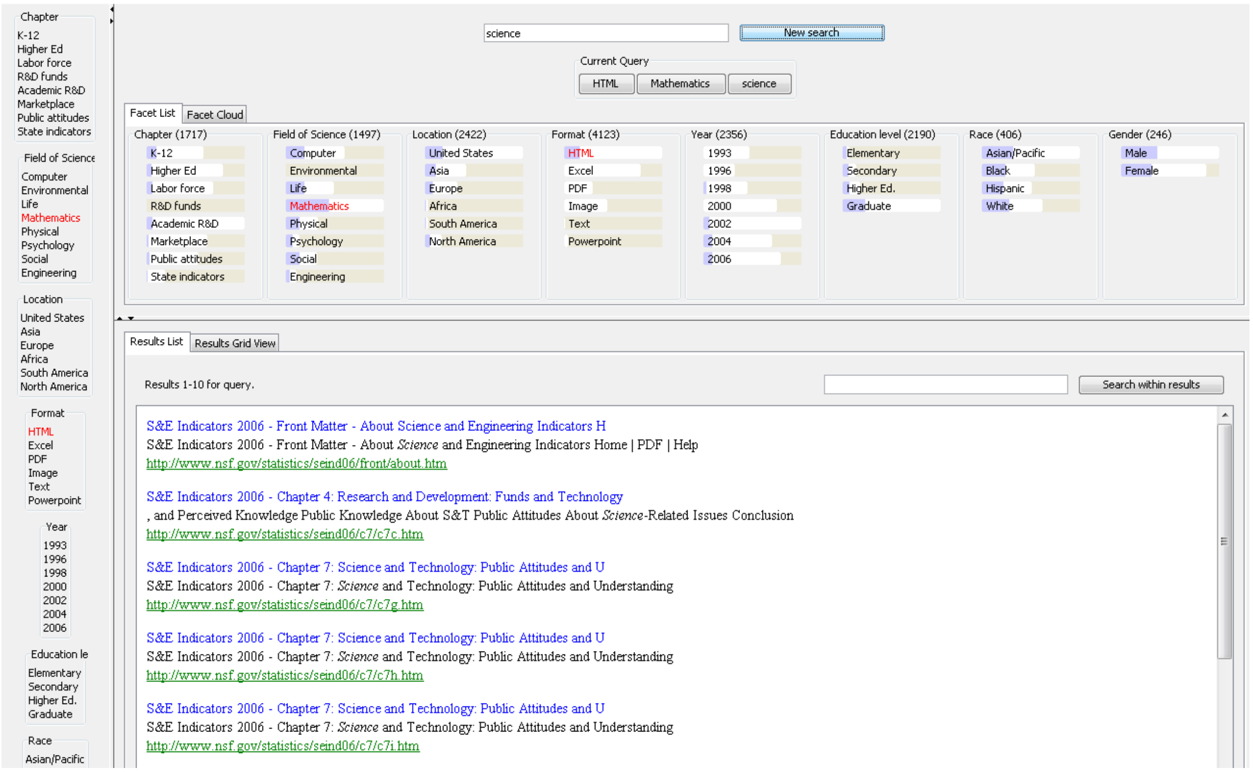
This work was supported by a contract with the NSF Division of Science Resources Statistics, Information and Technology Services Program.

7. References

1. Capra, R., and Marchionini, G. (2008). The Relation Browser Tool for Faceted Exploratory Search. Proceedings of the 2008 Conference on Digital Libraries (JCDL '08), Pittsburg, Pennsylvania, June 16 - 20, 2008.
2. Capra, R., Marchionini, G., Oh, J. S., Stutzman, F., and Zhang, Y. (2007). Effects of structure and interaction style on distinct search tasks. In Proceedings of the 2007 Conference on Digital Libraries (JCDL '07), Vancouver, BC, Canada, June 18 - 23, 2007.
3. Capra, R., and Marchionini, G. (2007). Visualizing Science and Engineering Indicators: Transitioning from Print to a Hybrid World. Technical Report TR-2007-07, School of Information and Library Science, University of North Carolina, Chapel Hill, NC.
4. Marchionini, G. (2006) Exploratory search: From finding to understanding. *CACM*, 49(4), 41-46.
5. Marchionini, G. and Brunk, B. (2003) Towards a general relation browser: A GUI for information architects. *Journal of Digital Information, 4*(1).
6. Shneiderman, B. (1994). Dynamic Queries for Visual Information Seeking, *IEEE Software*, 11(6): 70-77.
7. Zhang, J. and Marchionini, G. (2005). Evaluation and evolution of a browse and search interface: relation browser. In *Proceedings of the 2005 National Conference on Digital Government Research* (Atlanta, Georgia, May 15-18, 2005). ACM International Conference Proceeding Series, vol 89. Digital Government Research Center, 179-188.

Relation Browser Tool for Information Seeking

Robert Capra



The Relation Browser (RB) is a tool developed by the Interaction Design Lab at the University of North Carolina at Chapel Hill for understanding relationships between items in a collection and for exploring an information space (e.g., a set of documents or webpages). Facets and keyword searching are supported to allow users to easily move between search and browse strategies. The bars in the facet list indicate the number of results for the current query (purple) and the overall number of items in the collection that have this facet (white). Elements of the interface are coordinated and dynamic, meaning that as users brush the mouse over a facet, the elements update to show what the results would be after including the moused over item in the search. This feature allows users to quickly and easily explore an information space. Additional views are supported for both the results (display as a list, or in a grid), and the facets (display in a list, or as a 'facted cloud' similar to a tag cloud). The current query is shown at the top of the display and items in the current query are highlighted in red in other areas of the display. The example above shows an RB instance using data from the NSF Science and Engineering Indicators publications from 1993 to 2006.

Building Blocks For Rapid Development of Information Seeking Support Systems

Gary Geisler
School of Information
University of Texas at Austin
1 University Station, D700
Austin, TX 78712-0390
512 232-2980
geisler@ischool.utexas.edu

ABSTRACT

This paper proposes that to help accelerate progress in our understanding of how to best develop information seeking support systems in an increasingly wide range of domains, a coordinated effort to develop, publicize, and use *building blocks*—tools, services, system frameworks, data sets, sample collections—would make it easier for system developers to more quickly create and evaluate new systems and allow these developers more time to focus on potentially useful novel features. An explanation of this rationale is followed by a preliminary list of potential building blocks and a discussion of the challenges associated with encouraging use and development of building blocks.

1. INTRODUCTION

As Marchionini [2] suggests, systems designed primarily to present ranked retrieval results—no matter how effective the underlying retrieval mechanism might be—are not sufficient for many information seeking tasks. There are clearly many situations in which finding specific information resources might be part of the user's workflow but not the central activity. The learning and investigation tasks described in Marchionini [2] suggest a wide range of activities beyond entering search terms and evaluating the returned results, as do the "design principles" described by Shneiderman [5] in his analysis of how to better develop creativity support tools. These activities include exploring information by different means, such as graphical visualizations, results tightly-coupled to user interaction (dynamic queries), clustered and flexible categorization; personalizing and storing resources found useful through user histories, annotations, tagging, and persistent user-based lists and collections; and communicating and sharing information and discoveries with other people.

The April, 2006 special issue of the *Communications of the ACM* focusing on exploratory search highlights some of the systems and tools developed to support a broader range exploratory search tasks. White et al. [7] point out in their introduction to this issue that while most of these systems and tools are prototypes restricted to specific domains, information technology has advanced to the stage where we can begin to develop these types of systems and tools to support exploratory search more broadly. This paper builds on that premise, along with the ideas and suggestions for creativity support tools put forth by Shneiderman [5], and argues that one factor in stimulating progress toward better information seeking support systems might be a coordinated effort to encourage the production and use of freely-available, easy-to-use, *building blocks*—tools, services, system frameworks, data sets, sample collections—that will enable system developers to more rapidly assemble and evaluate new information seeking systems in new domains.

2. WHY BUILDING BLOCKS?

Consider a simple but recent, real-life example: The author teaches a graduate-level course on digital library principles and development to information and library science students, the large majority of whom have neither a computer science background nor programming expertise beyond, perhaps, a scripting language such as PHP. Figure 1 shows a timeline students developed for the Spring, 2007 course, in which the class project was to develop a digital library of archival materials from the production of the film *Gone With the Wind*. Figure 2 shows a timeline developed by the students in the course the following year (this time focusing on archival materials associated with the Alfred Hitchcock film *Rebecca*). The first example shows part of one of a series of linked but static graphics; somewhat useful for users of the digital library, but limited. In the second example, the students used SIMILE [6], an open source timeline tool developed at MIT. Because SIMILE is free, well-documented, and designed to be used with minimal effort, creating the *Rebecca* timeline required no more time and effort than the static timeline that was done for the *Gone With the Wind* digital library. It is clear, however, that as a prototype system to be evaluated and used by potential users, a digital library with the sort of functional, interactive timeline provided by SIMILE will be much more useful as a realistic demonstration and for obtaining constructive feedback in an iterative system development process.

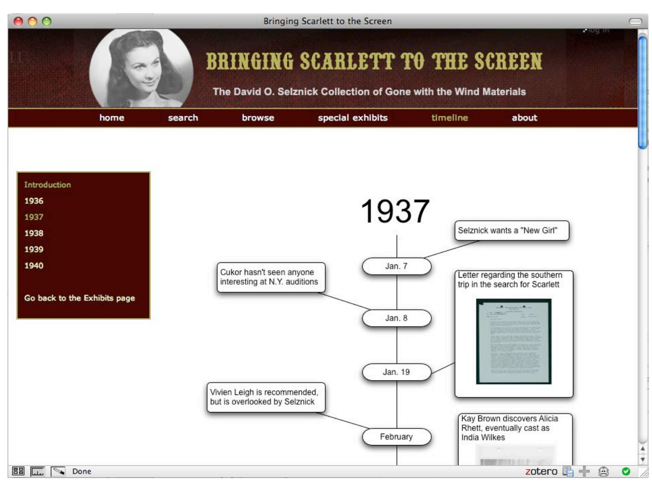


Figure 1. Static timeline from 2007 course

Figure 2. SIMILE-based timeline from 2008 course

While the example above is admittedly simple and anecdotal, it is important to recognize that as information systems increasingly become Web-based, the developers of these systems are increasingly more likely to resemble the students in the digital library course example than expert programmers with computer science degrees. The growing quantity of information resources becoming available in digital form naturally leads to a corresponding interest in creating focused collections of these information resources in a broadening range of disciplines and domains. The growth of digital libraries and digital media collections is one example of this, as are the efforts by publishers to provide increasingly feature-rich systems around their content. While, for instance, the New England Journal of Medicine might have the resources to hire computer science graduates to develop useful new features for their medical resource-based information seeking system [3], there will be many other situations (in archives, museums, and libraries, for example) where the people responsible for creating new information seeking systems do not have particularly strong programming backgrounds and are, in some cases, hired more for their domain expertise. Building blocks could help these developers better contribute to the creation of useful systems in new domains.

Building blocks for information seeking support systems could benefit more than those without sophisticated programming expertise, however. Even for experienced programmers, developing new system tools and features takes considerable time and effort. As more digital information resources and resource collections become available, there will be a broadening range of domains for which we might create useful information seeking support systems. While the literature in HCI and related fields shows no shortage of exemplary systems created for specific domains, it seems likely that most of these were developed over extended periods of time with relatively good grant or industry funding. As we face unprecedented opportunities to work with new collections of resources and new types of users and tasks in new domains, it is imperative that we be able to develop and evaluate systems for these domains more quickly than in the past. To create truly useful information seeking support systems in novel domains, and indeed to stimulate progress towards a general understanding of how we can create better systems for information seeking, we need to be able to rapidly and iteratively develop and test new ideas. Building blocks would help even experienced programmers by enabling them to rely on existing tools where appropriate and devote more of their creative energies to potentially useful new and novel features.

3. ISSS BUILDING BLOCKS

The SIMILE timeline tool described earlier is considered to be an example of a "building block" in this paper because it has been expressly provided to be easily used and adapted by others in their own systems. A developer of a new information seeking support system who believes a timeline might be beneficial to the intended users of that system does not have to design and write original code to provide this feature. A small investment in time to install and configure the timeline tool enables the developer to add a potentially useful feature to the system and move on to the next. Although the ideal scenario where the developer can pick and choose among a wide range of building blocks for a new system is not without significant challenges, some of which are discussed later, a growing availability of similar tools, along with a favorable climate towards open source development and improvements in supporting technologies, suggests that a much broader set of building blocks is not an unrealistic goal.

After many years, it seems safe to say that the open source software movement has made significant inroads into mainstream computing and that the open source model has led to rapid improvements in a range of software applications. Many of the existing tools and services that might be used as building blocks in information seeking systems are open source or at least made freely available. We've also seen funding agencies in recent years encourage, or in some cases require, products of grant-funded projects be made freely available, as well as programs that have directly funded projects to develop open source or free tools and services. In other words, there seems to be growing general support in the research community for the idea that new tools and services should be developed with the intent that they be freely shared and used by others.

At the same time, information technology has progressed to a point where an increased focus on creating tools that are written to be easily shared with and adapted by others is quite reasonable. A wide range of free, customizable communication tools such as wikis and blogs provide many possibilities for sharing information and collaborating on projects. We have easy-to-use methods for organizing and working with data, through powerful and free database management systems such as PostgreSQL and MySQL or well-defined standards for structured data, such as XML, RDF, and JSON. And lightweight scripting languages such as Python, PHP, and Ruby on Rails make it easy to work with databases and structured data to create sophisticated tools and services, making it much more feasible to develop tools and services that can be easily adopted and adapted by others.

So if the time is right to focus on building blocks for information seeking support systems, what might these building blocks be? A preliminary and incomplete list of ideas for the sorts of tools and services from which information seeking support systems could be more easily and rapidly built includes:

- **Sample data sets and content collections:** While systems are often created in response to a need to make more accessible an existing set of data or collection of information resources, there are also situations where a researcher or developer wants to explore the feasibility or effectiveness of an idea for a new system or feature. In these cases the developer often simply needs a structured collection of resources, any resources. Rather than forcing the developer expend time and effort to assemble a collection of content simply to have something to test with (which often results in the use of unrealistically small collections), a selection of pre-existing data sets and resource collections would not only save the developer time but enable the system to be developed with a more robust set of content.

- **Image, video, audio processing tools:** Multimedia content is increasingly central to digital libraries and other information resource collections. To be made useful in a system, however, this multimedia content often needs processing in various ways. Tools that can help automate and batch process tasks such as format conversion, resizing, preview and summary surrogate generation, and even feature extraction, would greatly increase the speed at which high-quality multimedia-rich systems can be developed.
- **Record cataloging and metadata:** Most information seeking systems rely on rich metadata to describe and distinguish resources and to use as a basis for search, browse, and other system features. Managing resource records—adding new records, modifying record attributes, deleting records—is often done in ad-hoc, cumbersome, and inefficient ways. Tools that provide simple cataloging interfaces and ways to bulk import, export, or convert between data formats could help make system development more efficient.
- **Search engines:** This is an active area of research with much potential for enabling the system developer to choose, based on appropriateness to the domain in which the system is being developed, from different engines with different strengths. The key is being able to easily install and configure an engine and connect it to the rest of the system.
- **Tools for creating browsing interfaces:** Flamenco [1] is a great example of a tool in this category; similar tools based on other organizational schemes would enable a wider range of possibilities for metadata in different domains.
- **Information visualization tools:** While there have been many excellent information visualization projects, examples, and commercial products, there is arguably a need for simpler, lightweight visualization tools that would enable a system developer to integrate visualizations—as a supporting feature, not the focus of the system—into an information seeking system by selecting a type of visualization and performing a simple mapping of visualization attributes to data set attributes. Prefuse [4] is a promising example, though its basis in Java arguably creates some drawbacks in terms of lightweight building blocks.
- **GIS and mapping tools and services:** Relatively easy-to-use mapping APIs from Google, Yahoo, and others have resulted in a plethora of recent examples showing the value of integrating geospatial views of information resources into systems. More polished open source versions of these would make useful building blocks for many systems.
- **Timeline tools:** In many systems chronological-based views provide users with useful ways of finding resources of interest and understanding relationships between resources. SIMILE has been mentioned as a good example of a building block for this purpose, but many other ways of presenting and interacting with resources based on time attributes are possible.
- **Web services:** There are a growing number of easy-to-use methods and APIs that enable a developer to integrate dynamic information from other Web sites or services, such as weather data, news headlines, or Amazon search results.

4. ENCOURAGING USE OF BUILDING BLOCKS

As mentioned above, there are already available excellent examples of tools and services that could be considered building blocks for the development of information seeking support systems. More are being developed all the time. However, because useful tools are developed in many different types of environments, by people with varied motivations, and publicized in different venues, it isn't easy for a system developer to know whether relevant building blocks might be available for a given project and which might be the most useful.

To facilitate the use of building blocks and thereby promote more rapid development of a wide range of information seeking support systems, it might be beneficial to develop a dedicated repository of information about and links to information seeking system related building blocks. This repository could include documentation, examples, and best practices related to information seeking support system development as a way to educate and help developers and encourage the growth of a cohesive research and development community around this topic. A rough example of how this repository might take shape is the Yahoo! Design Pattern Library [8], which provides documented design patterns and stencils for Web development, many of which are linked to code modules that provide dedicated functions.

Creating and maintaining this sort of resource has many challenges not addressed here, but the benefits of doing so are clear: system developers could visit this centralized resource when working on a project, read about available building blocks, determine which might be most useful and relevant for the particular project, and navigate to the source of the building block to download the code and documentation.

5. ENCOURAGING DEVELOPMENT OF BUILDING BLOCKS

The promise of building blocks hinges, of course, on the development of useful tools and services designed to be easily used and adapted by system developers. On the one hand, many tools and services that can be useful for information seeking systems are already being developed, without a focused effort to encourage or organize these efforts. On the other hand, there are many tools and services that could have much more practical value to the HCI, IR, and related research communities if there was an increased focus on making them better documented and sharable. And there are certainly a wide range of needed tools, both in areas discussed earlier and in areas not mentioned in this paper, that do not yet exist but would be very feasible to develop today.

Although most developers of tools and services, at least in the academic and industry research communities, are probably sympathetic to the idea of contributing the work they do to the community and seeing it used in systems developed by others, there are certainly obstacles that challenge the vision of information seeking system building blocks outlined in this paper. Two of these challenges—receiving recognition for non-traditional intellectual contributions and compatibility of developed tools and services— are discussed here briefly.

Developing useful code is time-consuming; developing code that can be adopted and adapted by others requires even more time, and providing illustrative examples and documentation of the code more time still. Quantifying and assessing the quality of this effort is perhaps more difficult than, for example, evaluating the contribution of a peer-reviewed publication, but it is clear that we are in a transition period where there are increasingly varied ways for those in academia and industry to make significant intellectual contributions to science and society. Better mechanisms for recognizing tool and code contributions in the tenure and promotion process, for example, would surely enable some researchers to spend more time developing useful, sharable tools and code without feeling pressure to spend that effort on more clearly publishable work.

If a selection of appealing building blocks are available to a system developer but these building blocks vary in the programming languages they are written in or the operating systems they support, the value of these building blocks to the developer is diminished. It is difficult to develop a system with tools written in different languages, and not feasible to mix tools that only run on incompatible operating systems. For example, there are useful, freely available tools available today that are based on Microsoft technology (often for good reasons); while these might be potentially good building blocks for systems that can reasonably require Windows, they have limited value for developers interested in creating information seeking systems for intended audiences that do their work on other OS platforms.

6. CONCLUSION

It is clear that systems that enable users to better explore a collection of information resources, to discover new ideas and make connections among those resources, to collect and save resources, and to communicate and share reflections and ideas about the resources with others will become increasingly valuable in many different domains in the near future. We are only in the earliest stages of understanding what might be the most useful components and features of effective systems that support these activities, and how these features might vary in systems intended for different domains. This paper suggests that a focused effort on developing, publicizing, and using building blocks—sharable, adaptable, and well-documented tools and services—could increase the rate at which we are able to develop and evaluate new systems and thereby help accelerate progress in our understanding of how to best provide information seeking system support in different domains.

7. REFERENCES

[1] Hearst, M. (2006). Clustering versus faceted categorization for information exploration. *Communications of the ACM 49*, 4, 59-61.

[2] Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM 49*, 4, 41-46.

[3] New England Journal of Medicine beta site. beta.nejm.org

[4] Prefuse Information Visualization Toolkit. www.prefuse.org

[5] Shneiderman, B. (2007). Creativity support tools: accelerating discovery and innovation. *Communications of the ACM 50*, 12, 20-32.

[6] SIMILE Timeline. http://simile.mit.edu/timeline/

[7] White, R.W., Kules, B., Drucker, S. and schraefel, m.c. (2006). Supporting exploratory search: Introduction. *Communications of the ACM 49*, 4, 36-39.

[8] Yahoo Design Pattern Library. http://developer.yahoo.com/ypatterns/

NeoNote: User Centered Design Suggestions for a Global Shared Scholarly Annotation System

Author: Brad Hemminger, UNC-CH School of Information and Library Science

Short Abstract for Information Seeking Support Systems NSF Workshop 2008

Introduction

Significant changes are occurring to scholarly communications due to technological innovations, primarily the advent of computers and the internet. Compared to thirty years ago, scholars use computers both to do their work and write their papers. They publish in journals still, but the articles are available and accessed more in digital format than in print format. Scholars increasing share their work with others by putting their work on their websites, in institutional repositories, or emailing to others. They use collaborative tools for writing papers, or creating shared content on wikis. Scholars are beginning to compile large digital collections of research papers instead of print collections. They are beginning to annotate these papers electronically, and to share these annotations. They save citation information digitally in citation managers and automatically incorporate this information when writing their papers. They start their searches more often with search engines that traditional library catalog resources, and they spend most of their time searching for and looking for information in web browsers.

There has been a proliferation of tools developed to help support scholars in performing this myriad of activities, but in most all cases, the tools are designed for only a specific task or environment. As a result, scholars are forced to utilize many different incompatible tools to perform their scholarly work and communication. This paper looks at the problem from the scholar's perspective and proposes a user interface well suited to scholarly work practices. It also proposes a paradigm for how scholarly annotations could be captured, stored, searched and re-used on both a global scale, and at the level of an individual research laboratory or group. It is our hope that the paradigms proposed in this paper will provoke discussions in the digital library community about shared global designs for digital library content, including annotations.

As part of a user centered design process to develop a global shared annotation system, information from surveys [Hemminger ASIST 2007], our interviews with scientists at UNC, and feedback from users of scholarly annotation tools were analyzed to generate a list of features scholarly researchers required. Currently, there are no systems designed to provide a single comprehensive system to address all user's needs. There are, however, many individual applications that provide excellent support for one or more features needed by the scholarly researcher. This paper describes the desired features of a single comprehensive system, gives examples of current tools that support these features, and then describes the architecture of our resulting design (NeoNote, [REF URL]). Overall, we believe the most effective way such an interface could be delivered, would be as part of a next generation interface to the web, closer to the original hypertext systems first proposed like Xanadu [http://en.wikipedia.org/wiki/Project_Xanadu]. The important change would be that annotation and sharing would be integrated into the interface and user experience, in the same way that browsing is in today's web browser interface to the web. Alternatively, a smaller incremental step would be to have today's browser incorporate such capabilities through plugins, such as are common in Firefox. To test these ideas, we have taken the latter course, and implemented these features as an integrated set of tools through a plugin to the Firefox browser. Our implementation,
NeoNote, is focused on global shared annotations, and is part of the more general NeoRef [URL] scholarly communications project at UNC. A video introducing the NeoNote system which visually

shows how the paradigm works is available on YouTube [http://www.youtube.com/watch?v=PUn09--HRaw] and at a higher resolution on the author's website [http://www.ils.unc.edu/bmh/pubs/NeoNote-For-YouTube.wmv]. Additionally, current challenges facing such global scholarly annotation repositories are discussed.

In this paper, for simplicity, discussion and examples about the information (content items) that scholars identify and want to save, will be scholarly journal articles and annotations on them. This is because this paradigm is familiar to most readers, currently better understood by researchers, and there are more examples of applications working with scholarly journal articles. However, this discussion should apply to all types of content items (multimedia (such as audio, pictures, video), data analysis, datasets (such as genetic sequences, observations from astronomy, weather, geography) etc. The proposed (NeoNote) design does support all data types.

I. User Interface Features Required for a Successful Global Shared Annotation System

Selection from Web Page. Searching and selection of content items should be done from within the single interface to the web that the users utilizes (currently this is web browsers). While in the past many literature database searches were performed on specific interfaces like Dialog, ISI (?), nowadays searching occurs of these same databases via web pages, and even more frequently via web-based search engines. As a result, researchers need to be able to add content items identified on a web page to their "personal digital library" with a single click.
Examples: Zotero, Connotea, RefWorks. Add descriptions and links. The search features should be an integrated part of web browser (or whatever future application provides the user access to the web), rather than distinct applications.

Multiple levels of annotation complexity should be supported. Annotations are done for multiple reasons. Some annotations are complex (examples, comment, tags, link) , some are very simple (save item, save item plus highlighted text). All kinds of annotations should be very easy to create and the user should be to specify them directly from the web browser environment.

Simple Item capture with one click. For simple types of capturing an item (e.g. to a citation database, or to highlight text as annotation), the user should be able to do it directly from the web browser page with a single click (to save). Examples: Google Notebook (although first have to select notebook). Fleck (but requires turning on Fleck for page, then saving bookmark).

Simple Item Highlighting. A users should be able to highlight text on a content item, and have this automatically be captured as an annotation on that content item. Multiple highlights of an article should be able to be done successively without requiring any other interactions (for instance to again select a highlight tool). For instance, this models how students highlight textbooks with a highlighter—they pick it up and then continue to highlight until they are finished.

Complex Item Annotation and Capture. In some cases, users will want to capture additional information about an item. They may want to add a text note, tags, links to other items, etc. This should be supported, easily and conveniently, while not complicating the ability to do simple item captures or highlights.

Concept of Current Folder or Notebook. This is critically important as users frequently are working on a particular topic, and they simply want all items they continue to select to all be captured and “stored” in the same “group”. The concept of the group is common, and supported as folders in most browser bookmarks, or as a “notebook” in Google Notebooks.
Examples: Google Notebook. Clipping an item places it in the current notebook (which is the most recently used one).

Organizing and Searching. Users expressed two preferences. First for research purposes they tended to prefer to organize materials by placing them in a folder. Second, they wished to be able to tag content items with tag words or phrases. They wanted to search by any of the tags including the folder label, by the full-text of the content item, and potentially by other metadata of the item.
Examples: Social Network Sites de.li.cious, Flickr, etc (Wikipedia entry).

Automatic capture of citation. When the user selects an item for capture, the citation information (all relevant metadata) should be automatically captured into the annotation, and available for future use.
Examples: Zotero, Connotea, RefWorks

Automatic capture of the item, not just the link. While for many web pages, users are content to just capture a link to the original resource, most scholars prefer to maintain their own copy of research articles. At some point in the future when content items are all easily retrieved via universally unique identifiers from long term stable archives, it may no longer be necessary to keep private copies. From our data, though, scholars currently clearly prefer to keep their own copies, just as they have made print or Xerox copies in the past.

Saved content and annotations should be universally accessible. Most users operate from many locations, and many computers. Users want to have their annotations and content items available on th web from any computer at any location.
Google Notebooks, Google Docs, Connotea, social bookmarking tools (many examples, see Wikipedia page http://en.wikipedia.org/wiki/List_of_social_software#Social_bookmarking).

Annotations and Content should be sharable. Scholars want to be able to share their annotations and content items as they see fit. In some cases this means with the world, but in most cases they wish to share it with their research group, department or collaborators.
Examples: Connotea, Social Networking Sites (Wikipedia page, URL form above)

Collaborative Sharing. Scholars wish to be able to save papers to shared digital library, collaboratively edit annotations, revise papers, work on documents. The documents need to be available from any computer at any location, and include provisions for handling multiple write accesses at the same time.
Examples: GoogleDocs, Web based File Systems (WebDAV, commercial systems (SharePoint, etc)).

Musings on Information Seeking Support Systems

Michael D. Levi
U.S. Bureau of Labor Statistics
2 Massachusetts Ave., NE Suite 4110
Washington, DC 20212
1-202-691-5900
levi.michael@bls.gov

1. A USE CASE

Even simple questions tend to get out of hand.

Let's say I have been following current events and I want to know "Is the US in a recession?" I start with the common sense notion that a recession is "hard times." A little searching on the Web gives me a more sophisticated definition: a recession is two consecutive quarters of negative growth in the gross domestic product. Further effort gives me the correct definition: in the United States, the National Bureau of Economic Research determines whether the economy has fallen into a recession, and they can only establish the beginning (and end) of business cycle phases after the fact. NBER hasn't made that call yet, so technically nobody knows whether the US is in a recession

That does not satisfy me. I keep poking around. I find plenty of economic pundits offering their opinions. Some say "yes," others say "no," some say "maybe," still others say "it doesn't matter what you call it, people are in trouble."

Being a data geek, I am intrigued by some of the economic indicators the pundits cite. I find the Bureau of Economic Analysis Web site and look at GDP for a while. Then I find the Bureau of Labor Statistics Web site and hunt for unemployment data. 5.5% in May, 2008. Is that good or bad? I locate a data retrieval tool and get a graph of unemployment rates for the last twenty years. I print it off, go back to the NBER site, and mark the dates of previous recessions onto the BLS graph.

During this exercise I remember having read in a magazine somewhere that BLS unemployment numbers are bogus. A little searching gives me lots of references to the article I kind of remember but the article itself is only available to subscribers. The charges strike me as overblown and possibly ill-informed. I go back to the BLS site to see if they have an explanation of how they calculate unemployment. I learn that BLS publishes six different unemployment rates, U1 through U6. I also learn that BLS publishes unemployment rates for the county in which I live.

Cool! I wonder if I can locate international data, too, and all of a sudden I want to know the principal export of Bolivia. The first page I find suggests it is probably coca. I don't know whether I can believe that. And I'm off in an entirely new direction…

2. CHARACTERISTICS OF EXPLORATION

The above is a plausible narrative of information seeking behavior. Starting with the same question, of course, I could have made a substantially different journey. If "Is the US in a recession?" really meant "Am I going to lose my job? Am I going to lose my house? Will I be able to take care of my children?" then I might have spent my time looking at job postings, financial advice columns, or the Web site of a political party that promises to save me from destitution.

Some characteristics of open-ended, discovery-oriented exploration emerge:

1) I may not know, at the beginning, whether a seemingly straightforward line of inquiry will expand beyond recognition. Sometimes it will, sometimes it won't. A lot depends on my mood at any given moment.

2) I can't predict when the exploration will end. It may be when I'm satisfied that I have learned enough (which also would vary from day to day and query to query.) It may be when I get tired or bored. It may be when I've run out of time. Or it may be when I get distracted by dinner or the allure of the swimming pool.

3) I can't determine, objectively, whether the exploration has been a success. There is usually no "right answer" against which I can measure my progress.

4) My exploration is not a linear process. I could get interested in a tangent at any time from which I may not return. I am also likely to backtrack, possibly with some regularity, either because a tangent proved unfulfilling and I want to resume my original quest, or because I thought of a new question (or a new way of formulating a previous question) to direct at a resource I visited previously.

5) I am likely to want to combine, compare, or contrast information from multiple sources. One of those sources is my memory – which may or may not be reliable in any given circumstance.

In addition, some characteristics of the on-line medium itself also emerge:

1) Though an enormous quantity of information exists on the Web, an huge volume of material is <u>not</u> on the Web or is not accessible to all users.

2) The material that does exist is often of unknown quality and reliability.

3. MANUAL TOOLS

In the pre-Web days, if I was in a rigorous frame of mind when I began an information seeking exploration, I typically assembled and used the following set of tools to assist me in organizing materials and my thoughts:

- A pen
- A pad of paper
- Index cards
- A highlighter
- Sticky notes
- A large flat surface (desk or table.)

The pen and paper allowed me to take notes. The pen, highlighter, and sticky notes allowed me to bookmark and annotate source material. The index cards became the repository of themes, ideas, and facts. The flat surface allowed me to view many index cards at the same time and arrange them into a sequence, into groups, into a tree or network, and into combinations of these structures.

4. HUMAN ASSISTANCE

In the pre-Web days I was also more prone to draw on human expertise: teachers, librarians, and subject matter experts. Though I was not hesitant to ask for specific answers or pointers to specific resources, the highest utility of these contacts lay in the questions they would ask, which helped me clarify my quest and raised interesting new directions to explore.

A valuable human resource typically exhibited some combination of the follow characteristics:

- Discernment ("What is he really trying to get at?")
- Judgment ("Can he do statistical analysis by himself?")
- Subject matter knowledge
- Research expertise
- Patience.

5. INFORMATION SEEKING SUPPORT SYSTEMS

A useful information seeking support system, then, would require the following minimum functionality:

1) It should not interfere with my behavior as listed under Characteristics of Exploration above.

2) It should give me capabilities at least as good as those listed under Manual Tools above.

3) It should positively assist my explorations by making them easier or faster or more comprehensive or less error-prone or…

In addition, an ISSS might give me capabilities that I never employed before because they were not possible or because I didn't think of them.

But, to be truly a leap forwards, an ISSS would need to exhibit at least elements of discernment, judgment, subject matter expertise, and research savvy.

Honestly, I am skeptical – though not utterly despairing – that such an ISSS is feasible. The readily automatable functionality is unlikely to be much better than the manual tools and processes with which I am intimately familiar. The "human" characteristics that would make a real difference to me are unlikely to be readily automatable.

But I have been pleasantly surprised before. And maybe I will be, again.

ContextMiner: Explore Globally, Aggregate Locally
Chirag Shah

ContextMiner is a framework originally devised to assist archivists to harvest and analyze video assets and associated metadata and contextual information to supplement their collections. It is based on the idea that in addition to the primary resources, contextual information such as annotations, popularity, and usage data are critical to preservation and future sense making.

ContextMiner helps an individual or group to: (1) run automated crawls on various sources on the Web and collect data as well as contextual information (e.g., Figure 1), (2) analyze and add value to collected data and context, and (3) monitor digital objects of interest over time.

Following is a typical flow of using *ContextMiner*:

1. Start a new campaign (project) based on some story, concept, or an object.
2. Choose the sources to harvest (Web, Blogs, YouTube, etc.).
3. Define search parameters such as frequency of probes, categories of context (e.g., comments, ratings, rank in result list), and data flow paths (what to get and where to put it).
4. Initiate and monitor resultant streams. Figure 2 shows an example status display.
5. Manipulate and annotate individual items, context, and sets.

ContextMiner is in use by several members of the National Digital Information Infrastructure Preservation Program (http://www.digitalpreservation.gov/) and can be used by teachers or others who wish to harvest content on specific topics. Further development providing access to more sources, and tools for information exploration is underway. *ContextMiner* is available as open source code or a web-based service (www.contextminer.org).

Biographic sketch

Chirag Shah is a doctoral student at UNC Chapel Hill working with Prof. Gary Marchionini. His interests are collaborative information seeking, exploratory search, social IR, and digital preservation. He received MS in Computer Science from UMass Amherst. He is a student member of ACM as well as ASIST. He can be contacted at chirag@unc.edu.

My Campaigns -> Elections 2008 -> **Vote Different**

Vote Different [YouTube Video]

Query: hillary clinton

Description: Make up your own mind. Decide for yourself who should be our next president. NOTE: This is a mashup of the famous Apple 1984 Super Bowl ad. Search for the original on YouTube.

Username: ParkRidge47

Category: News

Keywords: 1984, Apple, Barack, Clinton, Hillary, Ingsoc, Macintosh, Obama

Duration: 1.23 min.

In-links to this item: 246

Figure 1: Captured video from YouTube along with some contextual information

Select: | Action: | Go

☐	Title	Date Created	Collection	Status	Last Export	Manipulate
☐	Elections 2008	2008-07-08	YouTube: 343, In-links: 106729 Blogs: 4999	Active	N/A	Description | Parameters | Queries
☐	civil rights	2008-07-08	YouTube: 0, In-links: 0 Blogs: 0	Active	N/A	Description | Parameters | Queries
☐	commercial fishing	2008-07-08	YouTube: 959, In-links: 8449 Blogs: 0	Active	N/A	Description | Parameters | Queries
☐	library of congress	2008-07-08	YouTube: 149, In-links: 4389 Blogs: 0	Active	N/A	Description | Parameters | Queries
☐	Wisconsin Dairy	2008-07-10	YouTube: 223, In-links: 7715 Blogs: 0	Active	N/A	Description | Parameters | Queries
☐	Sarah Palin	2008-08-30	YouTube: 775, In-links: 88745 Blogs: 9352	Active	N/A	Description | Parameters | Queries
☐	Stock market	2008-10-10	YouTube: 517, In-links: 32288 Blogs: 11111	Active	N/A	Description | Parameters | Queries
☐	collaborative search	2008-10-15	YouTube: 844, In-links: 19271 Blogs: 8418	Active	N/A	Description | Parameters | Queries
☐	President-elect Obama	2008-11-10	YouTube: 711, In-links: 37371 Blogs: 7489	Active	N/A	Description | Parameters | Queries

Figure 2: Running campaigns

Research Agenda: Visual Overviews for Exploratory Search

Ben Shneiderman
Human-Computer Interaction Lab &
Department of Computer Science
University of Maryland
College Park, MD 20742
ben@cs.umd.edu

ABSTRACT
Exploratory search is necessary when users knowledge of the domain is incomplete or when initial user goals do not match available data or metadata that is the basis for search indexing attributes. Such mismatches mean that users need to learn more in order to develop a better understanding of the domain or to revise their search goals. Exploratory search processes may take weeks or months, so interfaces that support prolonged exploration are necessary. The attraction of exploratory search is that users can take on more ambitious goals that require substantial learning and creative leaps to bridge the gaps between what they know and that they seek.

Author Keywords
Information visualization, network visualization, exploratory search, user interface

ACM Classification Keywords
H.5. Information interfaces and presentation (e.g., HCI). H.2 DATABASE MANAGEMENT

INTRODUCTION

The success of information visualization stems, in part, from its capacity to provide overviews that offer deeper understanding of important phenomena [Card, Mackinlay & Shneiderman 1999]. These overviews reveal distributions, clusters, gaps, and outliers that may lead to insights by domain experts. These insights may indicate erroneous or missing data, important relationships, or surprising patterns. Determining the efficacy of visual overviews is difficult, but we can study human performance in the process of making known-item searches, exploratory information seeking inquiries, and insight discovery events [Shneiderman & Plaisant 2006].

The tools that support search, browsing, and visualization have dramatically improved in the past decade, so there is value for the information retrieval community to re-examine recent work and consider what future opportunities there are for integrating exploratory search technologies with interactive information visualization [Shneiderman et al., 2000; Kules & Shneiderman 2008].

As Turing award-winner Richard Hamming, wrote: "The purpose of computing is insight, not numbers." I might paraphrase with "The purpose of visualization is insight, not pictures." Successful tools support a process of information-seeking that leads to important insights for individual users, organizational teams, and larger communities. The term *insights* makes clear that exploratory search is a human experience, made possible by well-designed tools that support discovery [Saraiya, North & Duca 2004, 2005; North 2006].

This research agenda suggests that visual overview for web searches of many kinds could dramatically accelerate human performance in understanding the distribution of search results [Chen 2005; Thomas & Cook 2005]. This could lead to more effective search outcomes in a shorter amount of time, even by those with less domain knowledge and less search experience.

The emphasis here is on network visualizations for citations among results sets [Shneiderman & Aris 2006; Aris & Shneiderman 2007, 2008; Perer & Shneiderman 2006, 2008a, 2008b], but treemaps and other overview methods also offer opportunities [Kules et al., 2008]

Prepared as a working paper for National Science Foundation workshop on Information Seeking Support Systems (ISSS), June 26-27, 2008, Chapel Hill, NC. Workshop organizers: Gary Marchionini and Ryen White.

http://www.ils.unc.edu/ISSS/

ACKNOWLEDGMENTS
Thanks for partial support from National Science Foundation grant IIS-0705832) iOPENR--A Flexible Framework to Support Rapid Learning in Unfamiliar Research Domains (8/1/2007-7/31/2010), Principal Investigator Bonnie Dorr, also includes the author, Judith Klavans, Jimmy Lin, and Dragomir Radev.

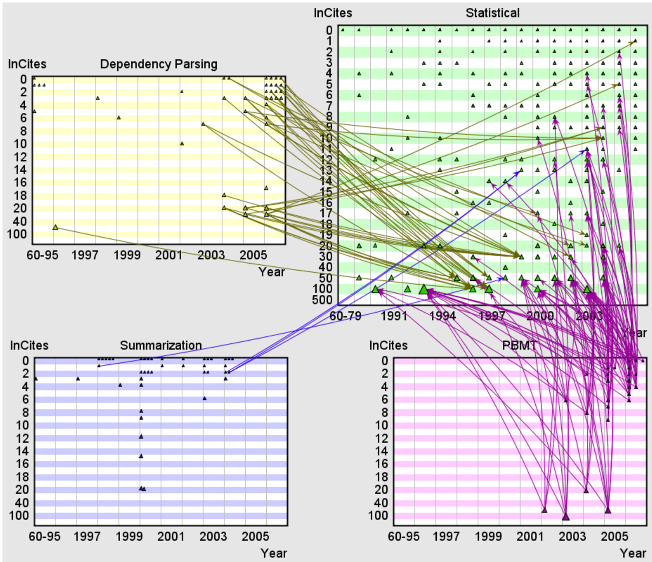


Figure 1: A semantic substrate shows regions for three small research topics (Dependency parsing, summarization, Phrase-based Machine Translation (PBMT)) and one large research topic (Statistical methods in natural language processing). The links show only citations that go from the small to the large topic, revealing strong linkage for some topics and weak ones for others. This figure was created using the Network Visualization with Semantic Substrates (NVSS) (http://www.cs.umd.edu/hcil/nvss).

REFERENCES

1. Aris, A. and Shneiderman, B., Designing semantic substrates for visual network exploration, *Information Visualization Journal 6*, 4 (2007), 1-20. doi: 10.1057/palgrave.ivs.9500162

2. Aris, A. and Shneiderman, B., A node aggregation to reduce complexity in network visualizations using semantic substrates, University of Maryland Technical Report, Dept of Computer Science (February 2008).

3. Card, S. K., MacKinlay, J. D., Shneiderman, B., *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann Publishers, San Francisco, CA (1999).

4. Chen, C., Top 10 Unsolved Information Visualization Problems, *IEEE Computer Graphics & Applications* (July/August 2005), 12-16.

5. Kules, B. and Shneiderman, B., Users can change their web search tactics: Design principles for categorized overviews, *Information Processing and Management* (to appear 2008).

6. North, C., Toward Measuring Visualization Insight., *IEEE Computer Graphics & Applications 26*, 3 (May-June 2006), 6-9

7. Perer, A. and Shneiderman, B., Balancing systematic and flexible exploration of social networks, *IEEE Symposium on Information Visualization* and *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (October 2006), 693-700.

8. Perer, A. and Shneiderman, B., Systematic yet flexible discovery: Guiding domain experts during exploratory data analysis, *Proc. ACM Conference on Intelligent User Interfaces* (January 2008a).

9. Perer, A. and Shneiderman, B., Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis, *ACM CHI2008 Conference*, ACM, New York (April 2008b), 265-274.

10. Saraiya, P., North, C., and Duca, K., An Insight-Based Methodology for Evaluating Bioinformatics Visualization, *IEEE Trans. Visualization and Computer Graphics 11*, 4 (July/Aug. 2005).

11. Saraiya, P., North, C., and Duca, K., An Evaluation of Microarray Visualization Tools for Biological Insight", *IEEE Symposium on Information Visualization 2004 (InfoVis 2004)* (2004), 1-8.

12. Shneiderman, B., The eyes have it: A task by data-type taxonomy for information visualizations, *Proc. Visual Languages* (Boulder, CO, Sept. 3–6). IEEE Computer Science Press, Los Alamitos, CA (1996), 336–343.

13. Shneiderman, B. and Aris, A., Network visualization with semantic substrates, *IEEE Symposium on Information Visualization* and *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (October 2006), 733-740.

14. Shneiderman, B., Feldman, D., Rose, A., and Ferre, X. A., Visualizing digital library search results with categorical and hierarchical axes, *Proc. 5th ACM International Conference on Digital Libraries,* ACM, New York (June 2000), 57-66.

15. Thomas, J.J. and Cook, K.A. (eds.), *Illuminating the Path: Research and Development Agenda for Visual Analytics*, IEEE Press (2005).

Evaluation Challenges and Directions for Information-Seeking Support Systems

Diane Kelly, University of North Carolina at Chapel Hill
Susan Dumais, Microsoft Research
Jan O. Pedersen, A9.com

Introduction

You are thinking about booking a vacation and would like to go someplace new. What process would you follow to gather candidate destinations and what systems, if any, would you consult? Would you think of this as a search task or something more general? What would constitute a successful outcome? Systems that support these sorts of open-ended tasks are referred to as Information-Seeking Support Systems (ISSS). Central to the development of ISSS is the question of evaluation. What does it mean for an ISSS to perform well, how do we measure this, and how do we use this information to build a more successful system?

Evaluating information-seeking support systems poses several challenges which we outline in this article. We first review current approaches for evaluating traditional information retrieval systems. We then discuss why these models are insufficient for evaluating systems that support information-seeking and present suggestions for creating a more inclusive and diverse evaluation framework for ISSS.

Current Evaluation Models

In the area of information retrieval (IR), evaluation has a long history that can be traced back to the Cranfield studies (Cleverdon, 1997/1967). The basic evaluation model has been extended by efforts associated with the Text Retrieval Conference (TREC) (Voorhees & Harman, 2005). In this model, researchers share common resources and guidelines for conducting system evaluations, which allow researchers to compare their search systems and to make progress in improving search algorithms. In the typical evaluation model, researchers use shared test collections which contain a corpus, queries and relevance assessments that indicate which documents are relevant to which queries. Particular evaluation measures are used to evaluate how well a search algorithm performs with respect to the number of relevant documents retrieved as well as the position of these documents within a ranked list. Common evaluation measures include precision, recall, mean average precision, mean reciprocal rank and discounted cumulative gain (see the Sidebar by Tunkelang). While researchers focus on different problems and different search strategies, the basic objective of IR system evaluation is to assess search performance which is usually tied directly to how effective the system is at retrieving and ranking relevant information objects. This evaluation model abstracts away the human elements and focuses primarily on the topical relevance of documents to queries.

Although most IR evaluations focus on search performance, the area of interactive IR (IIR) focuses on how people use systems to retrieve information (Ruthven, 2008). IIR is informed by many fields including traditional IR, information and library science, psychology and human-computer interaction (HCI). The IIR evaluation method was formalized by the TREC Interactive Track and has become the standard for laboratory evaluation of systems designed to support interactive information retrieval (Dumais & Belkin, 2005). In a typical IIR evaluation, searchers are asked to use one or more experimental IR systems to find information described by a small number of prescribed topics. Often the same test collections used in IR evaluation are used. Searchers' interactions with systems are logged and at various points they are asked to provide feedback via questionnaires and other self-report techniques. Typical outcome measures are usability and performance. While usability measures might be computed based on searchers' responses to questionnaire items or their interactions with the system, performance measures are usually computed based on the number of relevant documents searchers find and the time it takes them to do so. Very often performance is computed by comparing searchers' relevance assessments with baseline relevance assessments obtained from the test collection.

Both the IR and IIR evaluation models are based on laboratory evaluation, but researchers in these fields also conduct evaluations in more naturalistic environments. Researchers working at search engine companies, for instance, are able to analyze search logs that contain billions of records. In some cases, researchers can conduct live trials of experimental algorithms or search interfaces by making them available to a subset of searchers. While large-scale log studies allow researchers to observe a large number of searchers with a diverse range of interests and information needs, these observations are limited to what can be captured in a search log which primarily consists of queries and clicks. Important information about searchers, their information needs and the basis for their actions, are missing from such logs. It is possible to conduct ethnographic-style evaluations to obtain a more detailed understanding of searchers' real information needs, but these types of studies are less common. Moreover, data can only be collected from a small number of searchers about a small number of tasks which limits the generalizability of such studies. However, the use of complimentary methods, such as laboratory studies, log analyses and ethnographic observations can be used together to provide a richer picture of how systems support searchers in the search process (Grimes, Tang & Russell, 2007).

Why Current Models are Insufficient

Regardless of the location of the evaluation – laboratory or real world – current frameworks are insufficient for evaluating ISSS for a number of reasons. First, the user and task models used in traditional IR system evaluation do not capture all types of information-seeking tasks, activities and situations. Second, information-seeking tasks are often complex and evolve throughout the information-seeking process. Such tasks may not have stable, definable end points. Very often, the information-seeking process is just as important – if not more important – than the end point. Third, the dynamic and multimedia nature of the web continually changes the base of objects available for retrieval over time. Finally, information-seeking takes place over sustained periods of time and this implies longitudinal evaluation designs that measure change.

User and Task Models

Underlying all IR evaluation is some *user model*, which is an abstract representation of target searchers, and one or more *task models*, which represent user goals. The user and task models help define the particular behaviors and activities the system is intended to support and help determine the appropriateness of particular evaluation methods, measures, and participants.

User models in IR evaluation have often been limited to experienced searchers who have clearly-defined search tasks. One user model that has commonly been used is that of a librarian or other search intermediary. Other examples include intelligence analysts, patent searchers or novice searchers. The general Web user can also function as a user model, although when this type of searcher is studied researchers often narrow the possibilities by including contextual characteristics such as how much a person knows about a topic or how much time is available to complete a task. However, there are few Web user models, but a wide range of Web searchers, so the development of more nuanced models will be necessary for ISSS evaluation.

Task models in IR have included a wide range of search goals such as finding documents for a survey article, navigating to a key resource or homepage, checking a fact and answering a question (see the Sidebar by Kules and Capra). Most traditional IR evaluation is designed around a single search task in mind that can be resolved in a single session. It is important to emphasize that IR evaluation has focused on search tasks, not information-seeking tasks. While the user or task model might specify how the desired information is to be used, retrieval evaluation focuses on information finding and not information use. In information-seeking tasks, search is often not an end unto itself, but a task that is conducted in order to resolve a larger goal.

Task models in IR evaluation have also been somewhat self-contained. It is assumed that tasks are solvable in a single search session and that information needs can be represented and addressed with a single query. Most evaluation measures also make these assumptions. Relevance feedback incorporates the idea that an initial search query can be improved using explicit or implicit feedback from the searcher about which results are relevant. Using these judgments, a new search query is generated by modifying the original query by adding, removing or reweighting terms. But again, a single query and set of results are used for evaluation. More recently, measures like session-based discounted cumulative gain have been developed to summarize performance in search tasks where multiple queries are issued, results are of different quality and novelty, and stopping criteria vary (Järevlin, et al., 2008). Evaluation measures like this are important for characterizing information-seeking tasks.

People engage in information-seeking tasks for a number of reasons including to investigate curiosities, to learn about some topic of interest, to make connections between topics, to stimulate creativity and even for entertainment purposes. When tasks such as these are the goals, it is difficult to measure task outcomes. How can we determine if someone has learned something by using an ISSS? How much learning is required for us to say the ISSS is effective? What does it mean for an ISSS to help a person satisfy their curiosities or stimulate creativity? Appropriate outcome measures will be variable and are tied directly to the task the user is trying to accomplish. Thus, the development of richer task models for ISSS and corresponding evaluation measures are important research directions.

A related issue to user and tasks models is test corpora. A *corpus* is a set of documents, or information objects, which searchers access during a study. In traditional IR evaluations, test corpora have been fixed and stable. There has been a great emphasis on the use of static corpora because it facilitates evaluation, since all systems are working with the same collection of information. Moreover, it is possible to create topics that can be searched successfully and the researcher has some information about the number of topically-relevant documents. Traditional test corpora have consisted of mostly newswire text. Additional corpora that contain hyperlinked text and alternative types of information objects such as web pages, intranet pages, blog postings or images have also been developed, but test corpora typically contain only one type of information object. This is quite different from typical information seeking environments, where searchers are likely to encounter a large variety of document types and genre, which are of varying quality and constantly evolving over time.

Process and Outcome

There are many models of the information-seeking process, but for the most part these models have not made their way into IR system development and evaluation (c.f., Bates, 1989; Kuhlthau, 1993). Most of these models characterize information seeking as a *process* that takes place over time across many search episodes using many different resources. Information seeking is usually interwoven with a number of other activities and it is common for searchers to be engaged in multiple information seeking tasks simultaneously. These models also depict searchers as employing a number of information-seeking strategies, not just typing text into a query box and reviewing a list of search results, but also browsing related documents. For instance, Bates' (1989) berrypicking model presents evolving queries that change as the user engages in information seeking. This is different from the static and unitary query model assumed in IR evaluation. Bates further posits that information needs are often not satisfied by a single final retrieved set, but by a series of queries, navigations and selections that occur throughout the information-seeking process.

Information-seeking tasks are often complex and evolve during the information-seeking process. While there might be objective, definable solutions for traditional IR search tasks this is not necessarily the case for information-seeking tasks. Furthermore, the information-seeking process itself is just as important – if not more important – than the final state. This suggests that traditional evaluation measures based on system performance, and more specifically based on how many relevant documents are returned in response to a single query at a single point in time need to be extended. This also suggests that the notion of topical relevance, which measures whether the document is topically related to the query, will need to be extended. Other types of relevance – situational, cognitive and motivational – will become increasingly important in evaluation (Saracevic, 2007). The practice of evaluating performance using benchmark assessments based on objective, topical relevance will also no longer be sufficient. It has been known for some time that benchmark relevance assessments based on topical relevance do not generalize across searchers and it is difficult to create benchmark judgments based on other types of relevance since they are even more individualistic by nature. Finally, the practice of asking searchers to make absolute relevance judgments of information objects becomes less useful since relevance assessments will change throughout the course of information seeking.

Information seeking often takes place over time. One problematic aspect of the traditional IIR evaluation model is that only a small slice of the search process – that which can be captured during a short experimental session – is studied. For some types of tasks this is not too problematic. For example, many high-precision search tasks can be completed in a short period of time. These tasks might require searchers to find answers to specific questions, for example the current weather, show times, or the birth date of a celebrity. The resolution of these tasks might be motivated by some other larger task (e.g., a user might search for show times because they plan to attend the movies), but these larger tasks usually do not require synthesis and integration of information from multiple sources; information needs are temporarily fixed and their resolution immediate. A system's ability to resolve such tasks is also clear to searchers. Searchers can look at a search results list and determine whether the information is there and, if so, where it is located in the ranked list of search results. What is problematic about these types of tasks is that the answers to such questions often change frequently and their correctness may depend on when the searcher asks the question (e.g., the current weather conditions).

When we consider more open-ended information-seeking tasks, the temporal constraints of the traditional IIR evaluation model are more problematic since such tasks may not be resolved in a short time period and may be part of an on-going quest for information that has no easily identifiable point of completion. This suggests longitudinal study designs which allow researchers to observe series of information-seeking activities that occur over time. These types of study designs are more time consuming and require researchers to give up some experimental control. In laboratory IR evaluations, many of the variables that are not of immediate interest are controlled. However, in ISSS, searching is situated in a much richer context, making control more difficult and less desirable in many cases. While traditional evaluation has focused on component analysis which allows isolation and control of variables, holistic evaluation models are needed which capture more of the variability that exists in ISSS. Longitudinal evaluation models will require more sustained engagement with study subjects, the development of a wider range of less intrusive instruments for data collection, and richer analysis methods for identifying important information seeking behaviors and outcomes (see Sidebar by O'Brien & Toms for combining complex metrics).

Studies of web search logs represent another example of longitudinal, sustained engagement and use of unobtrusive data collection techniques. However, many of these logs contain only partial information about user behavior. Those that rely on server-side logging capture the user's communication with one particular search service, but not what searchers do after they navigate away from the search service. Client-side logging captures a more complete picture of searchers' behaviors and can be instrumented via Web browser toolbars, but researchers have the added challenge of identifying information seeking behaviors amidst the many other actions that are recorded. Regardless of what type of logging is used, there is a need to collect additional data that provide a context for interpreting searchers' actions found in these logs.

Evaluation Directions for ISSS

Information system development and evaluation is complex and usually requires multi-disciplinary teams including engineers, computer scientists and behavioral scientists. The analysis of information seeking and exploration and the development of systems to support this can be informed by many different scholars with a wide range of expertise. Thus, the creation of a diverse and interdisciplinary research community is necessary for significant advancement.

While research and development of information-seeking models have paralleled IR research, these two areas have developed largely independently. IR evaluation has been driven by Cranfield-style retrieval experiments and user studies most often conducted in laboratory environments which require a certain amount of abstraction and control. The development of information-seeking models has primarily been through naturalistic, qualitative studies that have involved small sets of searchers. ISSS represents an opportunity to bring together the development of information seeking models and retrieval systems to create a new framework for evaluation. ISSS also represents an opportunity to broaden and extend information-seeking models through large-scale evaluation. For instance, these models have typically excluded representation of the information environment; future models will need to accommodate the variety of information environments in which information seeking occurs.

One way to create a research community that includes participation from the broader range of disciplines needed to develop and evaluate information-seeking systems, is to create sharable resources that facilitate and enable participation. Such resources should contain datasets, search components, interface tools, instruments and measures, and operational search environments in which to try new ideas. Datasets that are likely to be valuable include both large-scale Web log data and data from more controlled laboratory studies. Datasets from smaller, focused studies often has richer contextual data and much more information is available about what searchers are trying to accomplish. It is often the case that interviews and video recordings of the screen accompany these types of datasets. A repository of datasets (with appropriate consent from and privacy protection for searchers) would provide a point of collaboration among researchers and allow for data to be examined and analyzed using a wider array of methods and techniques.

The development of search components that can be easily plugged into experimental systems or that can be used to examine behavioral issues is needed. ISSS will likely perform a variety of functions that extend beyond search, although search will still be an important component. However, it is very difficult for individual researchers, or even small teams of researchers, to develop an end-to-end system that is stable, robust, effective and efficient. Moreover, because system components work together, if one component does not work adequately then it will impact searchers' experiences with the entire system. At present, a stable working system is needed for inclusion in IIR research. This requires substantial engineering effort and presents a high barrier to entry which greatly reduces the number of researchers who can participate and the number of iterations that can be performed. This is especially true when we consider researchers in traditional social science disciplines, who may not have expertise in search technology or in building and deploying large-scale systems, but may have a great deal of expertise about human behavior and information processing. Developing resources that support richer collaborations and contributions from the wide range of relevant disciplines is a key enabler for improving information-seeking support and for developing a richer theoretical basis for our work.

Other types of shared resources that are needed to facilitate research are those that allow researchers to collect data from searchers. This includes loggers that monitor user actions (in interacting with Web search engines, vertical search engines, browsing, etc.) as well capture page contents. In addition to traditional loggers, instruments are needed that collect data that enriches log data. These instruments might elicit information from searchers about their goals, needs and states at various points in time. This data could supplement log data and provide more information information-seeking context.

The development of new evaluation methods and measures are also top priorities for ISSS evaluation. Because searchers may not seek to solve a specific problem but rather to sustain an interest, their expectations of how the system should support them may be different from searchers of traditional document retrieval systems who have more clearly defined goals that are tied closely to finding relevant documents or answers. Understanding these expectations will likely lead to the identification of additional evaluation criteria that reflect success. It is also the case that more process-specific measures will need to be developed as well as those that measure learning, cognitive transformation, confidence, engagement and affect. Performance, of course, will still be important, but measures are needed that do not depend on benchmark relevance assessments and which consider multiple query iterations and search sessions.

Studies that seek to describe the range of information-seeking tasks, processes and search strategies in which searchers engage are also necessary. These studies can help establish user and task models that can be used in more focused evaluations, such as those that might occur in a laboratory, and help developers understand the range of behaviors and activities that information-seeking support systems need to accommodate. Conceptualizing task and creating task models, in particular, are very important activities since they determine appropriate evaluation measures.

Finally, the evaluation of ISSS will require the development of longitudinal research designs that involve larger numbers of more diverse searchers. One way to address this is to create a *living laboratory* on the Web that brings researchers and searchers together. Such a lab might contain resources for facilitating evaluation, but also infrastructure for allowing searchers to participate in research studies. Today, many researchers in IR are using crowdsourcing via Mechanical Turk as a way to obtain relevance assessments (e.g., Alonso, Rose and Stewart, 2008). Crowdsourcing can also be used to solicit research participants for ISSS evaluation. The development of an infrastructure for facilitating recruitment, retention and participation is needed, so that researchers can broaden the range of participants in their studies and ultimately broaden what is known about ISSS.

Conclusion

The evaluation of ISSS provides an exciting opportunity to extend previous IR and IIR evaluation models and create a research community that embraces diverse methods and participation. An opportunity also exists for incorporating more aspects of the information-seeking process into ISSS development and evaluation and for developing a richer theoretical foundation. Community participation and shared resources are keys to leveraging the expertise that exists within the community, as well as attracting additional participants who can broaden our perspectives and enhance our understanding of the information-seeking process and the systems needed to support it.

References

Alonso, O., Rose, D. E., & Stewart, B. (2008). Crowdsourcing for relevance evaluation. *SIGIR Forum, 42*(2), 10-16.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13*(5), 407-424.

Cleverdon. C. W. (1997/1967). The Cranfield tests on index language devices. In K. Spark Jones & P. Willett (Eds.), *Readings in Information Retrieval*, San Francisco: Morgan Kaufman Publishers. (Reprinted from *Aslib Proceedings*, 173-192.)

Dumais, S. T. & Belkin, N. J. (2005). The TREC Interactive Tracks: Putting the user into search. In E. M. Voorhees & D. K. Harman (Eds.) *TREC: Experiment and Evaluation in Information Retrieval* (pp. 123-153), Cambridge, MA: MIT Press.

Grimes, C., Tang, D. & Russell, D. (2007). Query logs alone are not enough. In *Proceedings of WWW 2007 Workshop on Query Log Analysis: Social and Technological Changes.*

Järvelin, K., Price, S. L, Delcambre, L. M. L., & Nielsen, M. L. (2008). Discounted cumulated gain based evaluation of multiple-query IR sessions. *Proceedings of the 30th European Conference on Information Retrieval*, Glasgow, Scotland.

Kuhlthau, C. C. (1993). *Seeking meaning: A process approach to library and information services*. Norwood, NJ: Ablex.

Ruthven, I. (2008). Interactive information retrieval. *Annual Review of Information Science and Technology, 42*, 43-91.

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology, 58*(13), 1915-1933.

Voorhees, E. M. & Harman, D. K. (2005). *TREC: Experiment and Evaluation in Information Retrieval*, Cambridge, MA: MIT Press.

Author Biographies

Diane Kelly (dianek@email.unc.edu) is an assistant professor in the School of Information and Library Science at University of North Carolina at Chapel Hill. Her research and teaching interests are in interactive information retrieval, user behavior and evaluation methods and metrics. She received a PhD in information science from Rutgers University.

Susan Dumais (sdumais@microsoft.com) is a principal researcher and manager of the Context, Learning and User Experience for Search group at Microsoft Research. Her research interests are in information retrieval algorithms, interfaces and evaluation. She received a PhD in cognitive psychology from Indiana University. She is an ACM Fellow and a member of the CHI Academy.

Jan O. Pedersen (jpederse@yahoo.com) is currently Chief Scientist at A9, an Amazon company. His research interests span the range of Web search technologies, including matching, ranking and evaluation. He has held a number of positions at Web search companies, most recently Yahoo where he was Chief Scientist for the Search and Adverting Technology group. Dr. Pedersen received a PhD in Statistics from Stanford University and is an ACM Distinguished Scientist.

Designing Exploratory Search Tasks for Evaluation of Information Seeking Support Systems

Bill Kules and Robert Capra

Exploratory search tasks arise from information needs in which users "lack the knowledge or contextual awareness to formulate queries or navigate complex information spaces, the search task requires browsing and exploration, or system indexing of available information is inadequate" [3]. They comprise an important class of information problems that share uncertainty, ambiguity and discovery as common aspects.

Whether conducting usability studies or controlled experiments, researchers and practitioners need to know that the tasks they use are ecologically valid and representative of real-world information needs. Designing tasks to study exploratory search can be especially difficult because of the need to induce an exploratory rather than directed style of search. At the same time, the tasks need to be constructed in such a way that the results can be compared – between subjects in a single study and across multiple studies by different research groups.

Kules & Capra [2] identifies a set of desirable characteristics for exploratory search tasks and proposes a formal procedure for constructing tasks. It draws task topics from query log data, integrates them into a high-level work scenario [1], and addresses practical issues encountered in controlled or semi-controlled evaluations. An experimental evaluation of four tasks created using this procedure suggested that it led to well-grounded, realistic tasks that did elicit exploratory search behavior. The characteristics propose that an exploratory task:

- Indicates uncertainty, ambiguity in information need, and/or a need for discovery.
- Suggests a knowledge acquisition, comparison, or discovery task
- Provides a low level of specificity about: the information necessary and how to find the required information
- Provides enough imaginative context in order for the study participants to be able to relate and apply the situation.

The following example task meets these criteria and is based on a topic extracted from actual query logs from an online library catalog.

> *Imagine you are taking a class called "Feminism in the United States". For this class you need to write a research paper on some aspect of the U.S. feminist movement, but have yet to decide on a topic. Use the catalog to find two possible topics for your paper. Then use the catalog to find three books for each topic so that you might make a decision as to which topic to write about.*

The task gives participants some direction (two topics, three books on each), but leaves the main aspects of the information need open for the participant to explore.

The TREC conferences have demonstrated the value of well-constructed, comparable tasks in the evaluation of information retrieval systems. With growing interest in exploratory search from both researchers and practitioners, there is a need to develop such tasks that can be used in the study of exploratory search behaviors and systems.

References:

1. Borlund, P. (2003). The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3), paper no. 152. Retrieved September 19, 2008, from http://informationr.net/ir/8-3/paper152.html.
2. Kules, B., & Capra, R. (2008). Creating exploratory tasks for a faceted search interface. Paper presented at the Second Workshop on Human-Computer Interaction (HCIR 2008).
3. White, R. W., Kules, B., Drucker, S. M., & schraefel, m.c. (2006). Supporting exploratory search. Communications of the ACM, 49(4), 36–39.

Making Sense of Search Result Pages

[Position Paper]

Jan O. Pedersen
Yahoo! Inc.
2811 Mission College Blvd
Santa Clara, CA 95054
jpederse@yahoo-inc.com

1. OVERVIEW

Search engine result pages are presented hundreds of millions of times a day, yet it is not well understood what makes a particular page better from a consumer's perspective. For example, search engines spend large amounts of capital to make search-page loading latencies low, but how fast is fast enough or why fast is better is largely a subject of anecdote. Another example; search engines deploy large teams to continuously improve ranking quality, yet ranking quality does not fully characterize what is valued by the consumer. To see this, consider a feature referred to as *site collapse*, the custom of clustering results from the same site, subresults indented below the main (most relevant) result. Common measures of ranking quality, such as discounted cumulative gain (DCG) [1], are not optimized by this feature, a seeming contradiction.

Much of the contradiction comes from imposing a optimization criterion that does not account for perceptual phenomena. Users rapidly scan search result pages and often make a decision as to what action to take next within a few seconds. Presentations optimized for easy consumption and efficient scanning will be perceived as more relevant even if the content is identical to (or worse than) other, more awkward displays.

Eye-tracking experiments have shed considerable light on these issues [2]. It has been noted that users consume content from top to bottom cycling through quick scan-cull-decide cycles. Presentation details are extremely determinant of whether a consumer will notice a result. For example, the presence or absence of bolding can impact click rates significantly. Once a result is noticed, the decision to click is driven by ancillary information in the result summary, such as the contextual snippets and the url.

Search result pages are becoming more complex; they can no longer be characterized as a ranked list of homogeneously presented results. It is quite common to see content modules, such as the Yahoo! movie direct display, inserted above, below or between results. The result presentations are sometimes enriched with additional data, such as quick links for homepages and in-line players for video content. In addition, interventions in the process of constructing a query are possible via search assistance technologies such as Yahoo! Gossip.

As the search industry moves forward, a deeper understanding of user searching behavior will be required to better shape increasing complex products. The industry is fortunate to have an abundance of user-behavior data; every aspect of the search experience is carefully metered. However, this data is very complex in structure and captures only a slice of Internet activity. An effective model of user interaction with the search result page, validated through qualitative surveys and eye tracking experiments will be necessary to make full use of this data.

2. RESEARCH TOPICS

Yahoo! currently conducts research into user search behavior through qualitative ethnographic studies, eye tracking experiments, analysis of query log data and live experiments. Unfortunately, the largest scale, and hence richest, data source — the log data — is the hardest to interpret since it is a partial, retrospective record of user actions. In particular, simple, aggregate click-through data is not very diagnostic because it averages over a huge range of phenomena. Insights from user modeling are critical to both developing finer-grained, more revealing, user feedback measures and in developing tracking metrics that allow Yahoo! to monitor the health of its services.

The process Yahoo! search uses to design, validate, and optimize a new search feature includes the following steps:

1. Propose a design and build a mock-up or prototype
2. Translate the stated goals of the feature into expected user behaviors
3. Conduct a usability test of the proposed feature; often this includes a eye tracking experiment
4. Validate if the design achieves the desired goals; if not iterate
5. Develop proxy measures for the desired behaviors that can be measured in the user feedback logs

6. Conduct an online test of the feature to quantify effects
7. After launch, track the health of the feature through a feedback metric

User modeling is critical for steps 2 and 5 and, of course, reflects strongly on step 1. The next few sections describes some of the the data and methods used to develop the Yahoo! search user model.

2.1 Session Analysis

One rich source of insights into consumer search behavior is query log session analysis. A session is a sequence of user actions, including query reformulations and url clicks, over a relatively short span of time associated with the same user goal. For example, one finds in the Yahoo query logs the following session fragments:

Lead query	Follow query
samsung lcd tv	best buy store
samsung lcd tv	consumer reports
37 inch lcd tv	best 37 inch lcd tv
shelving units	costco wholesale
shelving units	home library

The user's intention is made clearer in the reformulation. Joining this data with user clicks actions is even more informative. Yahoo! actively mines session data to build search assists, spell correct queries, and to automatically expand queries (in sponsored search). A newer area is mining this data to measure the effectiveness of ranking and presentation algorithms, with the underlying assumption that better selection and ranking algorithms will anticipate user reformulations. For example, one study examines conditions under which users switch from one search engine to another [3]. However, the surface has only been scratched in the analysis of session data.

2.2 Toolbar data

Search engine query logs only reflect a small slice of user behavior — actions taken on the search results page. A more complete picture would include the entire click stream; search result page clicks as well as offsite follow-on actions. This sort of data is available from a subset of toolbar users — those that opt into having their click stream tracked. Yahoo! has just begun to collect this sort of data, although competing search engines have collected it for some time. We expect to derive much better indicators of user satisfaction by consider the actions post click. For example, if the user exits the clicked-through page rapidly then one can infer that the information need was not satisfied by that page. User satisfaction indicators can then be correlated with search result page context to shed light on usefulness.

2.3 Eye Tracking

Ultimately a direct measurement of what the user perceives on a search results page would be most useful. A closely proxy of this ideal is measuring eye movement and fixations across the page followed by user interviews. Aggregate measures reveal which page elements get the greatest attention, while tracking individual sessions is very revealing about scanning behavior. One challenge is scaling up these experiments, which are quite labor intensive, so that they can be used not only to generate hypotheses but also to statistically confirm them.

3. CONCLUSION

We are still very early in the development of a comprehensive model for user interaction with the search results page. A this time, we use operating principles, such as the hypothesis that users approach a search task with a fixed cognitive effort budget or that bolding is critical is gaining user attention. We will need more quantitative models to take the search experience to the next level.

4. REFERENCES

[1] Kalervo Järvelin and Jaana Kekäläinen, *IR evaluation methods for retrieving highly relevant documents*, SIGIR 2000, pages 41–48

[2] A.T. Duchowski, *Eye Tracking Methodology: Theory and Practice*, Springer-Verlag, New York, Inc., 2003

[3] Yun-Fang Juan and Chi-Chao Chang, *An Analysis of Search Engine Switching Behavior using Click Streams*, WWW 2005, Chiba Japan, pages 1050–1051

The Measurement Dilemma for ISSS

Elaine Toms and Heather O'Brien

For more than half a century, precision and recall, and its variants were the standard norms for evaluating IR systems. With the emergence of HCI came an imperative to view systems through users' eyes to assess efficiency, and effectiveness, as well as an assortment of user perceptions. In the collision of these two worlds emerged a collection of objective and subjective metrics that relate to users, their context and tasks, as well as aspects of the resources and systems. However, all tend to be measured independently (e.g., size of query, time on task, satisfaction) despite the fact that search systems have multiple inter-related components that control, manage and affect user interactivity. The dilemma is how to assess an ISSS given the complex nature of both the system and the human environment in which it operates.

The first metric to assess relationships among aspects of a complex system was the "Informativeness" measure which combined a subjective user response regarding usefulness of the information examined, with a system penalty derived from the system's ability to present relevant items in the most useful (to the user) order [2]. Notably, this metric was based on sound mathematical principles and information search theory, and the first to measure interactivity in search systems, rather than aspects of it.

With the limited availability of complex metrics, we need methods that examine inter-relationships among multiple simple metrics. This requires techniques that are not commonly used to evaluate search systems: factor analysis (FA) and structural equation modeling (SEM). Both enable different types of data (e.g., user attitudes, observed behaviours, system performance) to be examined simultaneously for a more holistic approach. Since search systems are tools, the success of the tool is tied intrinsically to its human use. A system cannot be assessed independent of its user; this calls for integrated metrics that reflect interactivity.

FA looks for simple patterns in the associations among variables and extracts the most parsimonious set. For example, FA was used to map metrics to dimensions of relevance, deducing that three core factors or variables (system, user and task) could be measured with eight objective or subjective metrics of user and system performance [3]. SEM takes the analysis a step farther, combining confirmatory factor analysis and path analysis to confirm factors and to build predictive models about their relationships. Since SEM models are theoretically derived, data analysis tests the "fit" between the hypothesized and actual relationships in the data, indicating which variables are independent and which are directly or indirectly related to a larger set. SEM was used to evaluate a six-factor scale of user engagement [1], confirming both the presence of the factors (Aesthetics, Novelty, Involvement, Focused Attention, Perceived Usability, and Endurability) and the predictive relationships among them.

ISSSs are complex systems with multiple features, and enable multiple types of interactivity. Techniques such as FA and SEM facilitate the assessment of varied, multiple, simple measures. Core to both techniques is a clear theoretical focus for measurement selection and the interpretation of the output, much like any other statistical technique. Both techniques will fail if the user "dumps" in a set of data and makes miraculous conclusions about the results, independent of the theoretical foundation that informs the phenomenon under examination. Used appropriately, these techniques may lead to the creation of complex metrics for a more holistic evaluation of ISSSs.

[1] O'Brien, H. (2008). *Defining and Measuring Engagement in User Experiences with Technology.* (Doctoral dissertation, Dalhousie University, 2008) (available from first author).

[2] Tague-Sutcliffe, J. (1992). Measuring the informativeness of a retrieval process. In *Proceedings of the 15th Annual international ACM SIGIR Conference on Research and Development in information Retrieval* (Copenhagen, Denmark, June 21 - 24, 1992). N. Belkin, P. Ingwersen, and A. M. Pejtersen, Eds. SIGIR '92. ACM, New York, NY, 23-36.

[3]Toms, EG, O'Brien, H, Kopak, R & Freund, L. (2005). Searching for relevance in the relevance of search. In F. Crestani and I. Ruthven (eds.)*, Context: nature, impact, and role : COLIS5: Proceedings of the Fifth International Conference on Conceptions of Library and Information Sciences (Glasgow, UK, June 5-8),* pp. 59-78. *Lecture Notes in Computer Science 3507*

Elaine Toms is Canada Research Chair in Management Informatics and Associate Professor, Faculty of Management, Dalhousie University, Halifax, Nova Scotia, Canada. She received a PhD in library and information science from the University of Western Ontario. Contact her at etoms@dal.ca

Heather O'Brien is an Assistant Professor at the School of Library, Archives and Information Studies, University of British Columbia, Vancouver, British Columbia, Canada. She received a PhD in Interdisciplinary Studies from Dalhousie University. Contact her at hlobrien@interchange.ubc.ca.

Resolving the Battle Royale between Information Retrieval and Information Science

Daniel Tunkelang
Endeca
dt@endeca.com

ABSTRACT
We propose an approach to help resolve the "battle royale" between the information retrieval and information science communities. The information retrieval side favors the Cranfield paradigm of batch evaluation, criticized by the information science side for its neglect of the user. The information science side favors user studies, criticized by the information retrieval side for their scale and repeatability challenges. Our approach aims to satisfy the primary concerns of both sides.

Categories and Subject Descriptors
H.1.2 [Human Factors]: Human information processing.

H.3.3 [Information Systems]: Information Search and Retrieval - Information Filtering, Retrieval Models

H.5.2 [Information Systems]: Information Interfaces and Presentation - User Interfaces

General Terms
Design, Experimentation, Human Factors

Keywords
Information science, information retrieval, information seeking, evaluation, user studies

1. INTRODUCTION
Over the past few decades, a growing community of researchers has called for the information retrieval community to think outside the Cranfield box. Perhaps the most vocal advocate is Nick Belkin, whose "grand challenges" in his keynote at the 2008 European Conference on Information Retrieval [1] all pertained to the interactive nature of information seeking he claims the Cranfield approach neglects. Belkin cited similar calls to action going back as far as Karen Spärck Jones, in her 1988 acceptance speech for the Gerald Salton award [2], and again from Tefko Saracevic, when he received the same award in 1997 [3]. More recently, we have the Information Seeking and Retrieval research program proposed by Peter Ingwersen and Kalervo Järvelin in *The Turn*, published in 2005 [4].

Information Seeking Support Systems Workshop, June 26–27, 2008, Chapel Hill, North Carolina, USA.

2. IMPASSE BETWEEN IR AND IS
Given the advocacy of Belkin and others, why hasn't there been more progress? As Ellen Voorhees noted in defense of Cranfield at the 2006 Workshop on Adaptive Information Retrieval, "changing the abstraction slightly to include just a bit more characterization of the user will result in a dramatic loss of power or increase in cost of retrieval experiments" [5]. Despite user studies that have sought to challenge the Cranfield emphasis on batch information retrieval measures like mean average precision—such as those of Andrew Turpin and Bill Hersh [6]—the information retrieval community, on the whole, remains unconvinced by these experiments because they are smaller in scale and less repeatable than the TREC evaluations.

As Tefko Saracevic has said, there is a "battle royale" between the information retrieval community, which favors the Cranfield paradigm of batch evaluation despite its neglect of the user, and the information science community, which favors user studies despite their scale and repeatability challenges [7]. How do we move forward?

3. PRIMARY CONCERNS OF IR AND IS
Both sides have compelling arguments. If an evaluation procedure is not repeatable and cost-effective, it has little practical value. Nonetheless, it is essential that an evaluation procedure measure the interactive nature of information seeking.

If we are to find common ground to resolve this dispute, we need to satisfy the primary concerns of both sides:

- Real information seeking tasks are interstice, so the results of the evaluation procedure must be meaningful in an interactive context.
- The evaluation procedure must be repeatable and cost-effective.

In order to move beyond the battle royale and resolve the impasse between the IR and IS communities, we need to address both of these concerns.

4. PROPOSED APPROACH
A key point of contention in the battle royale is whether we should evaluate systems by studying individual users or measuring system performance against test collections.

The short answer is that we need to do both. In order to ground the results of evaluation in realistic contexts, we need to conduct user studies that relate proposed measures to success in interactive information seeking tasks. Otherwise, we optimize under the artificial constraint that a task involves only a single user query.

Such an approach presumes that we have a characterization of information seeking tasks. This characterization is an open problem that is beyond the scope of this position paper but has been addressed by other information seeking researchers, including Ingwersen and Järvelin [4]. We presume access to a set of tasks that, if not exhaustive, at least applies to a valuable subset of real information seeking problems.

Consider, as a concrete example, the task of a researcher who, given a comprehensive digital library of technical publications, wants to determine with confidence whether his or her idea is novel. In other words, the researcher want to either discover prior art that anticipates the idea, or to state with confidence that there is no such art. Patent inventors and lawyers performing e-discovery perform analogous tasks. We can measure task performance objectively as a combination of accuracy and efficiency, and we can also consider subject measures like user confidence and satisfaction. Let us assume that we are able to quantify a task success measure that incorporates these factors.

Given this task and success measure, we would like to know how well an information retrieval system supports the user performing it. As the information scientists correctly argue, user studies are indispensable. But, as we employ user studies to determine which systems are most helpful to users, we need to go a step further and correlate user success to one or more system measures. We can then evaluate these system measures in a repeatable, cost-effective process that does not require user involvement.

For example, let us hypothesize that mean average precision (MAP) on a given TREC collection is such a measure. We hypothesize that users pursuing the prior art search task are more successful using a system with higher MAP than those using a system with lower MAP. In order to test this hypothesis, we can present users with a family of systems that, insofar as possible, vary only in MAP, and see how well user success correlates to the system's MAP. If the correlation is strong, then we validate the utility of MAP as a system measure and invest in evaluating systems using MAP against the specified collection in order to predict their utility for the prior art task.

The principle here is a general one, and can even be used not only to compare different algorithms, but also to evaluate more sophisticated interfaces, such as document clustering [8] or faceted search [9]. The only requirement is that we hypothesize and validate system measures that correlate to user success.

5. WEAKNESSES OF APPROACH

Our proposed approach has two major weaknesses.

The first weakness is that, in a realistic interactive information retrieval context, distinct queries are not independent. Rather, a typical user executes a sequence of queries in pursuit of an information need, each query informed by the results of the previous ones.

In a batch test, we must decide the query sequence in advance, and cannot model how the user's queries depend on system response. Hence, we are limited to computing measures that can be evaluated for each query independently. Nonetheless, we can choose measures which correlate to effectiveness in realistic settings. Hopefully these measures are still meaningful, even when we remove the test queries from their realistic context.

The second challenge is that we do not envision a way to compare different interfaces in a batch setting. It seems that testing the relative merits of different interfaces requires real—or at least simulated—users.

If, however, we hold the interface constant, then we can define performance measures that apply to those interfaces. For example, we can develop standardized versions of well-studied interfaces, such as faceted search and clustering. We can then compare the performance of different systems that use these interfaces, e.g., different clustering algorithms.

6. AN ALTERNATIVE APPROACH

An alternative way to tackle the evaluation problem leverages the "human computation" approach championed by Luis Von Ahn [10]. This approach uses "games with a purpose" to motivate people to perform information-related tasks, such as image tagging and optical character recognition (OCR).

A particularly interesting "game" in our present context is Phetch, in which in which one or more "Seekers" compete to find an image based on a text description provided by a "Describer" [11]. The Describer's goal is to help the Seekers succeed, while the Seekers compete with one another to find the target image within a fixed time limit, using search engine that has indexed the images based on tagging results from the ESP Game. In order to discourage a shotgun approach, the game penalizes Seekers for wrong guesses.

This game goes quite far in capturing the essence of interactive information retrieval. If we put aside the competition among the Seekers, then we see that an individual Seeker, aided by the human Describer and the algorithmic--but human indexed--search engine--is pursuing an information retrieval task. Moreover, the Seeker is incented to be both effective and efficient.

How can we leverage this framework for information retrieval evaluation? Even though the game envisions both Describers and Seekers to be human beings, there is no reason we cannot allow computers to play too--in either or both roles. Granted, the game, as currently designed, focuses on image retrieval without giving the human players direct access to the image tags, but we could imagine a framework that is more amenable to machine participation, e.g., providing a machine player with a set of tags derived from those in the index when that player is presented with an image. Alternatively, there may be a domain more suited than image retrieval to incorporating computer players.

The main appeal of the game framework is that it allows all participants to be judged based on an objective criterion that reflects the effectiveness and efficiency of the interactive information retrieval process. A good Describer should, on average, outscore a bad Describer over the long term; likewise, a good Seeker should outscore a bad one. We can even vary the search engine available to Seekers, in order to compare competing search engine algorithms or interfaces.

7. CONCLUSION

Our goal is ambitious: we aspire towards an evaluation framework that satisfies information scientists as relevant to real-world information seeking, but nonetheless offers the practicality of the Cranfield paradigm that dominates information retrieval. The near absence of collaboration between the information science and information retrieval communities has been a greatly missed opportunity not only for both researcher communities but also for the rest of the world who could benefit from practical advances in our understanding of information seeking. We hope that the approach we propose takes at least a small step towards resolving this battle royale.

8. REFERENCES

[1] Belkin, N. J., 2008. Some(What) Grand Challenges for Information Retrieval. *ACM SIGIR Forum 42, 1* (June 2008), 47-54.

[2] Spärck Jones, K. 1988. A look back and a look forward. In: SIGIR '88. In *Proceedings of the 11th Annual ACM SIGIR International Conference on Research and Development in Information Retrieval*, 13-29.

[3] Saracevic, T. 1997. Users lost: reflections of the past, future and limits of information science. *ACM SIGIR Forum 31, 2* (July 1997), 16-27.

[4] Ingwersen, P. and Järvelin, K. 2005. *The turn. Integration of information seeking and retrieval in context*. Springer.

[5] Voorhees, E. 2006. Building Test Collections for Adaptive Information Retrieval: What to Abstract for What cost? In *First International Workshop on Adaptive Information Retrieval (AIR)*.

[6] Turpin, A. and Scholer, F. 2006. User performance versus precision measures for simple search tasks. In *Proceedings of the 29th Annual international ACM SIGIR Conference on Research and Development in information Retrieval*, 11-18.

[7] Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology 58(3)*, 1915-1933.

[8] Cutting, D., Karger, D., Pedersen, J., and Tukey, J. 1992. Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections. In *Proceedings of the 15th Annual ACM SIGIR International Conference on Research and Development in Information Retrieval*, 318-329.

[9] Workshop on Faceted Search. 2006. In *Proceedings of the 29th Annual ACM SIGIR International Conference on Research and Development in Information Retrieval*.

[10] Von Ahn, L. 2006. Games with a Purpose. *IEEE Computer 39, 6* (June 2006), 92-94.

[11] Von Ahn, L., Ginosar, S., Kedia, M., Liu, R., and Blum, M. 2006. Improving accessibility of the web with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 79-82.

Daniel Tunkelang

Precision AND Recall

Information retrieval research today emphasizes precision at the expense of recall. Precision is the number of relevant documents retrieved by a search divided by the total number of documents retrieved, while recall is the number of relevant documents retrieved divided by the total number of existing relevant documents which should have been retrieved.

These measures were originally intended for set retrieval, but most current research assumes a ranked retrieval model, where results are returned in order of their estimated likelihood of relevance to a search query. Popular measures, like mean average precision (MAP) and normalized discounted cumulative gain (NDCG) [1], mostly reflect precision for the highest-ranked results.

For the most difficult and most valuable information-seeking problems, however, recall is at least as important as precision. In particular, for tasks that involve exploration or progressive elaboration of the user's needs, a user's progress depends on understanding the breadth and organization of available content related to those needs. Techniques designed for interactive retrieval, particularly those that support iterative query refinement, rely on communicating to the user the properties of large sets of documents, and thus benefit from a retrieval approach with a high degree of recall [2].

The extreme case for the importance of recall is the problem of information availability, where the seeker faces uncertainty as to whether the information of interest is available at all. Instances of this problem include some of the highest-value information tasks, such as those facing national security and legal/patent professionals, who might spend hours or days searching to determine whether the desired information exists.

The information retrieval community would do well to develop benchmarks for systems that consider recall at least as important as precision. Perhaps researchers should revive the set retrieval models and measures such as the F1 score, which is the harmonic mean of precision and recall.

Meanwhile, information scientists could use information availability problems as realistic tests for user studies of exploratory search systems, or interactive retrieval approaches in general. The effectiveness of such systems would be measured in terms of the correctness of the outcome (i.e., does the user correctly conclude whether the information of interest is available?), user confidence in the outcome (which admittedly may be hard to quantify), and efficiency (i.e., the user’s time or labor expenditure).

Precision will always be an important performance measure, particularly for tasks like known-item search and navigational search. For more challenging information seeking tasks, however, recall is as or more important, and it is critical that the evaluation of information seeking support systems take recall into account.

[1] Järvelin, K. and Kekäläinen, J. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems* 20, 4 (Oct. 2002), 422-446.

[2] Rao, R., Pedersen, J. O., Hearst, M. A., Mackinlay, J. D., Card, S. K., Masinter, L., Halvorsen, P., and Robertson, G. C. 1995. Rich interaction in the digital library. *Communications of the ACM* 38, 4 (Apr. 1995), 29-39.

Industry-Academic Relationships

Daniel M. Russell

Industry and the academy have always worked together—sometimes working well, occasionally at cross-purposes. In the area of information seeking, the relationship has been historically very clear: the academic environment gave early support, instruction and a nurturing environment to the early search engines. Without the academy, the search world would be very different today. Perhaps search engines would have arrived on the technology landscape without the university, but the critical role university research has played in the exploration of ideas in indexing, crawling and user-interface design is without question.

The landscape is very different now. In 1998 when Google was launched, not long after Yahoo and AltaVista, research on an academic scope and scale could still be deeply influential in many of the basic research topics. However, as the commercial search engines grew, it quickly became difficult for non-commercial entities to do certain kinds of research. As a consequence, those areas of research—ranking algorithms, user-behavioral responses to subtle UI changes, very-large-scale distributed systems—rapidly became areas of inquiry where the research tools in computer science took on the nature of high-energy physics; the tools and datasets required are large, expensive and essential. Just as no individual physicist could easily build out and own a Large Hadron Collider, no individual computer or small academic research science team could explore research issues that requires thousands of processors, large teams of programmers to implement and maintain codebases, and large pools of support infrastructure.

Just as physics didn't stop being an area of intense inquiry as ever-larger instruments became essential to research, so too ISSS will have areas that are uniquely susceptible to study in the academy, not requiring the computational equivalent of the LHC.

But the physics analogy only goes so far. Unlike physics, where one of the primary goals of research instruments is to create data sets for collaborative study, the large search engine companies have deeply vested interest in ensuring that their use data remains private and secure. In the wake of the AOL click-stream data release incident [AOL], industry remains extraordinarily cautious about the release of potentially identifying data. Thus, data-sharing between industry and academic research has to remain very carefully prescribed. As a side-effect, certain kinds of research (such as investigating the effect of mining very large data sets of querystream data) will remain a topic of very carefully monitored and prescribed studies. To respect privacy and security considerations, the data involved can't be just simple open-access, but must be necessarily limited.

Where then can the relationship between academic research and industrial institutions go?

The industrial-scale ISSS companies are deeply aware of the value of research. In Google, it's part of the culture of the place in the form of the 20% time allocation, where individual employees have 1 day/week to investigate topics and ideas on their own. The vast majority of software engineers come out of an academic computer science background, many with advanced training in research. It's just in the genome of the search companies to do research as an essential part of moving forward in the industry.

Working with academic researchers is an important outcome of that history. Active industrial/academic research grants programs exist at each of the large search companies. These outreach programs not only support individual research projects, but also actively engage with researchers to jointly understand what research areas are amenable to academic research, and which are best done in places with appropriate instrumentation to perform the experiments.

Programs linking university research centers directly with search engine systems and data include Microsoft's "Accelerating Search" program [Microsoft], the partnership between Google and IBM to provide a cluster of processors for cloud computing research [Google & IBM], and Yahoo's collaboration with CRL to provide another testbed platform for cloud computing built on Apache / Hadoop. [Yahoo & CRL]

In truth, both academic research and industrial research fit into an ecosystem of extended collaboration, with people, research grants, shared resources and ideas flowing back and forth from universities to the large search companies. Academic computer science isn't a pejorative term at the search companies, but an essential part of the cycle of ideas and their transformation from concepts to working systems. As in many other academic/industrial relationships (think electrical engineering, chemical engineering, molecular biology, etc.), both the academic world and the industrial world naturally split into hemispheres of influence based on their unique tooling, size, interests and capacities. The academy generates and tests ideas through the vehicle of graduate students, professors, grants, and small-to-medium-scale experimentation. The industrial world scales ideas that work into fully operational systems. No one sphere of research holds a lock on novel ideas or innovation, but both mutually profit by having an active set of relationships allowing ideas and people to add to the vigor of our field.

References

[Yahoo & CRL] http://yhoo.client.shareholder.com/press/releasedetail.cfm?ReleaseID=301187

[Microsoft]
http://research.microsoft.com/ur/us/fundingopps/RFPs/Search_2006_RFP_Awards.aspx

[Google & IBM] http://www.google.com/intl/en/press/pressrel/20071008_ibm_univ.html

[AOL] http://query.nytimes.com/gs

A Call to Action

The adoption of web search technologies have led people have come to expect immediate and easy access to information in all aspects of life. Web-based search services have become fundamental components of the cyberinfrastructure that supports economic growth and social advance. Although we have built a base set of expectations for search, human needs for information go beyond search to filtering, assessing, sensemaking, synthesizing, and using information to meet our needs. Our nation and our world depend on citizens who are able to seek, assess, understand, and use diverse kinds of information. Much of the information we need is complex with different components held in disparate electronic sources. Additionally, much of the information we need is not discretely anticipated, but rather emerges as seeking and reflection continues over time. Information seeking in the digital age is a kind of problem solving activity that demands agile and symbiotic coordination of human and cyber resources. The skills and tools of information seeking are as important to children in school as they are to our best scientists and government leaders. Computation has expanded our ability to do scalable what if thinking that leverages the best capabilities of humans and machines to abstract, synthesize, and iterate intellectual actions, and today's search engines are the primitives on the technical side of information seeking. We must rise to the challenge to move information seeking from search engine support that provides discrete items in response to simple queries to tools and services that support reflective and interactive search over time and in collaboration. This report demonstrates some of the challenges such support presents and points to a sample of the work that is underway to move us beyond search. We require much broader and intensive efforts on the part of the academy, government, and industry if we are to meet the grand challenges of usable and ubiquitous information seeking support systems that empower us to solve problems and increase participation. National efforts are urgently needed to support research that leads to understanding information seeking as computationally augmented learning and problem solving, better seamless and ubiquitous systems for supporting information seeking, methods for training people to practice effective and efficient information seeking, and techniques and measures for assessing the tools and practices.

Workshop Schedule

Thursday June 26, 2008
Workshop will be held in the Pleasants Family Assembly Room in Wilson Library on the UNC campus

8:00 Breakfast on-site

8:30 Introductions and overview

9:30 Breakout Session 1 (assigned based on project statements)

11:00 Reports back in plenary

12:30 Working lunch and Breakout Session 2 (lunch on-site)

2:00 Break

2:30 Plenary discussion: Key research themes

5:00 Close

6:00 Group dinner Carolina Inn

Friday June 27, 2008
Workshop will be held in the Pleasants Family Assembly Room in Wilson Library on the UNC campus

8:00 Breakfast on-site

8:30 Plenary discussion: Refine themes and map projects to themes

10:30 Break

11:00 Breakout Session 3. Aggregate and refine projects and plan writing

12:30 Working lunch (lunch on-site)

1:30 Plenary discussion of final report and task assignments

3:00 Wrap-up