People, Interacting with Information[1]

Nicholas J. Belkin
School of Communication & Information,
Rutgers University, USA
belkin@rutgers.edu

Abstract

Recent recognition in IR that people engage in information seeking *sessions*; attempts to formally model search sessions; and, the move toward evaluation of IR systems over entire search sessions, are evidence of a new, broader understanding of IR's goals. This move takes IR's goal beyond identification of relevant information objects, to supporting people in achievement of the goal or task which led them to engage in information seeking. The emergence of the term, *interactive information retrieval*, the amount of research conducted under that rubric, and the lack of agreement about how to conduct and evaluate such research, testify to both the significance and the difficulty of addressing this new goal. In this new context, IR faces daunting challenges, including: understanding why people engage in information seeking; identifying aspects of that condition that need to be considered in supporting them in achieving their goals; designing support for evolving user goals and knowledge; and, perhaps most daunting, how to evaluate, in ways as rigorous and successful as IR evaluation has been to date, IR systems which attempt to address this broader goal. I, and others, have proposed that moving from relevance to *usefulness* as the main criterion for evaluation of IR system performance, could provide a basis for addressing many of these challenges. In this paper, I discuss how such a move might be implemented, some possible alternative means for addressing this broader goal for IR, and what serious attention to users of IR systems, and their interactions with information objects and systems, could mean for research and practice of IR.

1 Introduction

Let me begin by expressing how pleased, and humbly honored I am to be a recipient of the Gerard Salton Award. Gerry was a great man, and to receive the award named for him is very special. For me it is especially meaningful, given the sometime disputatious nature of our professional interactions, and what might have seemed, on the surface, to be quite different ideas about

[1] This is a modestly edited version of the lecture delivered on the occasion of presentation of the 2015 Gerard Salton Award, at SIGIR 2015, Santiago, Chile, 10 August 2015.

information retrieval. I say, on the surface, because in the end, I believe that he and I both shared the same goal for the field, although we approached it from quite different positions. In this lecture, I discuss at some length that goal, and on how I think it might be best addressed.

I am humbled also by the honor of having joined the ranks of the previous recipients of this award; founders, leaders and innovators in information retrieval (IR), from the earliest beginnings of the field to today. It has been my distinct good fortune to have known all of the previous recipients, to have collaborated with many of them, to have argued with all of them, to have learned from them, and, I hope, to have been able to appropriately incorporate their insights into my own work.
Today, following the example of many of my predecessors, I'd like to take the opportunity to, in Sue Dumais's words of 2009, "present a personal reflection on information retrieval." This will include an overview of my history in IR, a discussion of my personal take on its proper goals, and on how those might be best achieved, and proposing a specific approach to addressing a significant problem in information retrieval research.

2 Some personal background

The occasion of this Salton Award Lecture gives me the opportunity to consider the circumstances in which I began my research career, and about the nature of the path that my research has followed, and to talk about these a bit. For this, I beg your indulgence, as I ramble on.

I'd like to begin by pointing out that, whatever it is that I have been honored for, it is not my achievement alone, but rather a product of my good fortune in having been able to interact with, and be influenced by, many brilliant and supportive people. I came to the field of IR from a starting point in information science; specifically, with the concern of addressing general problems of information in society. I initially thought that the best way to do this would be to establish a firm framework for a science of information. Acting on my then understanding of what a science constituted, I began the project of defining information.

Fortunately for me, I did this at University College, London, with B.C. Brookes as my Ph.D. supervisor, and Stephen Robertson my office mate. It didn't take long for me to be disabused of a) the idea that information could be defined; and b) that defining its phenomena of interest was a necessary precondition to a "real" science. Instead, perhaps under the influence of the great pragmatist, Jeremy Bentham, founder of University College, whose body and brain I passed every day for two years, I turned to attempting to develop a concept of information which would lead to being able to predict its effect on a person's state of knowledge, which I took at that time as what IR systems should be attempting to do.

I had the great good fortune to begin my research career in the UK in the early 1970s. Beginning already in the late 1950s and throughout the 1960s, there was an extremely active, influential and close community of brilliant information science and information retrieval researchers in the UK.

At the time that I entered the field, as a lowly Ph.D. student, Robert Fairthorne and Brian Vickery, among the intellectual founders of these fields, were still active. Cyril Cleverdon, who, together with Michael Keene, founded the evaluation paradigm which information retrieval still follows, continued to be an influential figure. Jason Farradane, who founded the first academic unit of information science, was just reaching retirement at The City University, London. B.C. Brookes, was still a major contributor and integrator, just beginning his most productive period. Through Brookes, I was able to become a member of this community, and to meet and learn from all of these founders of our field.

But this community also included newer members, whom one might characterize as the next generation. Amongst these was, of course, Karen Spärck Jones, who began her career in the mid 1960s at the Cambridge Natural Language Research unit under the philosopher Margaret Masterman, and was, by 1973, one of the world's leading information retrieval researchers. Keith van Rijsbergen had just recently completed his PhD studies with Nick Jardine. Steve Robertson had just received the Royal Society Science Information Fellowship, and begun his studies with B.C. Brookes at University College London, after several years as a researcher at Aslib. I stepped into this milieu at about the same time as Bruce Croft and David Harper began their studies at Cambridge, and Bob Oddy his at Newcastle-upon-Tyne.

It is difficult to describe just how exciting and vital doing research in information science and information retrieval within this context was, and how important having been in that context was in influencing the careers of so many of us in it. I feel that I owe an incredible debt to the chance circumstances that landed me in it.

But of course there was also a world of information science and information retrieval beyond the UK. And Steve and I, who had been PhD students for only a couple of years, were encouraged and supported by B.C. Brookes and Brian Vickery, then the Director of the School of Library, Archive and Information Studies at University College, to organize a meeting of information science researchers that took place in London in 1975, which we ambitiously called the International Research Forum in Information Science. The goal of the meeting was to bring together an international group of people with interests in establishing theoretical foundations for information science. Figure 1 is a photo of some of the people whom the four of us invited as participants, with funding to Professor Vickery from the British Library Research and Development Department. This meeting then led to my becoming a member of the even broader international information science community, particularly with European colleagues. And the resulting interactions with such people as Peter Ingwersen at the Royal School of Librarianship, Copenhagen, Gernot Wersig and Thomas Seeger, at the Freie Universität, Berlin, and Marc de Mey at the University of Ghent, led me to developing my concept of the cognitive viewpoint in information science.

It still amazes me that we beginners in the field, Steve and I, had the audacity, and were given the opportunity, to be part of this. But this is the kind of support that one could find in this community at that time. And I strongly believe that this experience is the basis of whatever success or good I have since achieved.

Figure 1. Attendees of the First International Research Forum on Information Science, Westfield College, University of London, July/August 1975. Bottom row, first on right: Stephen Roberson; next to him Jose Marie Griffiths. Middle row, second from right, Manfred Kochen; fourth from right, Karen Spärck Jones; sixth from right, B.C. Brookes; next to him, Nick Belkin; next to him, Belver Griffith; top row, third from right, Marc de Mey. (For all names, see Appendix)

3 My research program

As I mentioned earlier, I began my research career thinking that I would "define" information, and thus establish a foundation for a science of information. But being exposed to Robertson's and Brookes's critical faculties, as well as doing some serious reading, thinking and talking with others, led me to something slightly more modest: to develop a *concept* of information which would lead to being able to predict its effect on a person's state of knowledge, which I took at that time as what information retrieval systems should be attempting to do. This reformulation led to considering two closely related problems. One was understanding what is meant by a "state of knowledge", how it could be represented, and how it develops and changes. This led to me cognitive psychology, and in particular, to the emerging field of cognitive science and what was known as "the cognitive viewpoint". The other was understanding *why* a person engages in information seeking, since what an information retrieval system should predict would not be only *some* effect, but specifically a *beneficial* effect. And *beneficial* would mean that which directly addressed that *why*.

Thinking about these two problems led me to propose the Anomalous State of Knowledge (ASK) hypothesis; that is, that what leads people to engage in information seeking, is some anomaly in their state of knowledge with respect to some topic or problem area, and that the goal of an information retrieval system is to help people in resolution of the anomaly. From this followed, that the most appropriate way for information retrieval systems to address this goal is developing means for understanding and "accurately" representing ASKs, and then to represent the information associated with documents in ways that are commensurable with the ASK representation (Belkin, 1977; 1980).

You will note that this way of considering how IR systems should operate is quite different from what was then (and still mostly is) the norm; that is, finding good ways for representing what documents are *about*, and then representing, in a commensurable manner, what a person's request to a system is *about*, in order to provide them with documents that are *about* what their request is *about*.

But our view of ASK-based information retrieval left the problem of how a system could construct an ASK. At the time that I was touting the ASK idea, Bob Oddy was completing his dissertation on information retrieval through man-machine dialogue (Oddy, 1977). His model suggested a way to progressively represent, during the course of a person's interaction with the information encountered through an information-seeking session, the person's initial and changing state of knowledge, and we collaborated for several years in developing the foundations and design for ASK-based IR systems (Belkin, Oddy & Brooks, 1982). For various reasons, including the inability of the technology available at that time to support the interactivity required, this line of research came to a gradual halt.

At about this time, I began a collaboration with a group at the Free University, Berlin, led by Gernot Wersig and Thomas Seeger. In the course of this research, the ASK idea became extended, primarily through the influence of Wersig's dissertation research (Wersig, 1972; 1979), to consider the impetus for information seeking to be the person's facing a *problematic situation*, a much broader concept than the ASK. This led us to investigate how people, in the position of helping others to resolve their problematic situations, went about accomplishing that goal. This work led then to a series of studies of the functions of human intermediaries, and of how information retrieval systems could incorporate those functions (Belkin, Seeger & Wersig, 1983). What followed was an active period of research on user modeling, and developing system designs based on complex models of the searcher and the searcher's context (e.g. Daniels, Brooks & Belkin, 1985; Brooks, Daniels & Belkin, 1986).

In 1985 I left the UK for US and Rutgers University, where I was again fortunate to engage with a supportive and innovative community, including Tefko Saracevic and Paul Kantor at Rutgers, Donald Walker and Susan Dumais at Bellcore, and Bruce Croft at University of Massachusetts, Amherst. To this I must add my PhD students, who kept me honest and from whom I learned so much, including how little I know. This cohort and community was essential to all that came later for me.

What then followed in my research was an increasing emphasis on studying and taking account of the information seeker's interactions not only with the system and its intermediary functions, but with the information itself. This led to the general idea of considering information retrieval as interaction with information (Belkin, 1993). This also formed the basis for a proposal of a model of information retrieval in which interaction was the central process, and the "user" was the central actor. In this model, the major concern of the information retrieval system was how best to support the type of interaction with information that the person was currently engaged in (or, in an alternative locution, how best to support the person's current search *intention*) (Belkin, 1996).

At about this time I also began to consider more directly the person's task or goal which initiated the problematic situation as the motivation for engaging in information seeking. These two moves led pretty directly to my work in personalization of information retrieval, which I believe actually incorporates much of what I learned in all of my research to this point (see, e.g. Cole, et al. 2011; Liu, C. et al. 2012; Liu, J. et al. 2010; Liu, J. et al., 2012).

Thus began, and developed, what I now understand as *the* unifying theme and goal of my research throughout my career:

understanding and supporting people's interactions with information.

The operationalization of this goal has changed over time, beginning with attempts to represent an information seeker's ASK, going on to designing ASK-based IR systems, to understanding the ASK as an aspect of a person's problematic situation, through mimicking the intermediary, to my current stance concerning the goal of IR:

> supporting people in accomplishing the task or goal which led them to engage in information seeking, by supporting appropriate interaction with information objects.

In all these somewhat different modes, there have been two constants in my research: emphasis on the person (often called the "user") as the central actor in the IR situation; and, understanding interaction as its central process. This position, for perhaps the first 2/3 of my career, has been in quite distinct contrast, not to say conflict, with "mainstream" IR research, which, until quite recently, focused almost exclusively on information object representation and retrieval techniques, and on models of IR which included the "user" as merely an input and feedback device, if at all, and which has no room for interaction as a process at all. But I can see that this situation is beginning to change, and I'd now like to consider why that used to be the case, and why it may be changing, and what that implies for research in information retrieval.

4 The "proper" goal of information retrieval, its problems and its implications

In her Salton Award Lecture in 1988, Karen Spärck Jones, one of the foremost contributors to our understanding of information object representation, and to the development of retrieval techniques and models, and arch experimenter, said the following, referring to IR research at least through the 1970s:

> But these concerns, though worthy, had unfortunate consequences. One was that, in spite of references to environmental parameters and so forth, it tested information systems in an abstract, reductionist way which was not only felt to be disagreeably arid but was judged to neglect not only important operational matters but, more importantly, much of the vital business of establishing the user's need. (Spärck Jones, 1988, p. 18)

And, referring to her own research in relevance feedback:

> More importantly, I felt that the required next step in this line of work was to carry out real, rather than simulated, interactive searching ... (Spärck Jones, 1988, p. 17)

Karen's example demonstrates that there has been, for some time, recognition that IR research would be wise to more directly address the user, as a significant actor in the IR system, and the interaction of the user with the system and with the information objects, as research foci.

Why then have the terms and concepts of users and people and interaction only so recently appeared in any significant number in venues such as SIGIR? Clearly, the inertia associated with reluctance to give up what seem to have been successful methods and practices has played a significant part in this state of affairs. But more important, I think, have been other factors, two especially.

One has been the reluctance, or inability of IR researchers to accept a broader and more realistic goal of their enterprise; that is, to go beyond identification of relevant, or even authoritative, information objects, to the goal which I think many would agree is what we should aspire to:

> the support of people in achievement of the goal or task which led them to engage in information seeking.

The other is more technical, but perhaps equally important. That is, that the theories, and research methods and techniques necessary for addressing the issues of users and interaction in IR, and the related evaluation methods and measures, are so very different from what we are familiar with, and also, so very difficult to both develop and implement.

I think that there's no question that the actual practice and performance of IR in contemporary systems demonstrates that its goal needs to be expanded to at least that which I propose, and that, therefore, information retrieval now needs to overcome these factors, and face up to the necessity of accepting this new goal, and adopting and adapting theories and methods appropriate to achieving that goal.

It is clear that there is a fairly recent trend in research in information retrieval, which usually goes under the rubric of *interactive* information retrieval, which has accepted these challenges. This move:

- recognizes that people engage in information seeking sessions;
- studies people in their interactions with information and with information retrieval systems;
- attempts to formally model search sessions; and,
- addresses the issue of evaluation of IR systems over entire search sessions.

The recent Dagstuhl Interactive IR (Fuhr, Belkin, Jose & van Rijsbergen,) and NII-Shonan Whole Session Evaluation (Belkin, Dumais, Kando & Sanderson, 2012) seminars, and the new SIGIR Conference on Information Interaction and Retrieval (CHIIR), and its immediate predecessors, IIiX (e.g. http://iiix2014.ur.de/) and HCIR (https://sites.google.com/site/hcirworkshop/), are also explicit demonstration of our field's newly found commitment to accepting and addressing these challenges.

But there is also substantial evidence of the difficulties faced by this move, including the disappearance of the TREC Session Track, and the lack of agreement in the field about how to conduct, and to evaluate, interactive information retrieval research.
So, information retrieval that accepts the goal which I've proposed, still faces daunting challenges, including:

- understanding why people engage in information seeking;
- identifying the aspects of that condition that need to be considered in supporting them in achieving their goals;
- how to design support for evolving user goals and knowledge;
- how best to support people's interactions with information objects and the other components of the IR system with respect to different and dynamic intentions; and, perhaps most daunting of all,
- how to evaluate, in ways as rigorous and successful as information retrieval evaluation has been to date, information retrieval systems which attempt to address this broader goal.

5 Usefulness and the evaluation of interactive information retrieval

The rigor and success of its evaluation methods and measures are without doubt one of the crowning achievements of information retrieval research. But evaluation of any system is always undertaken with respect to the goal, expressed or unexpressed, of that system. For many years, the field of information retrieval has accepted something like "the provision of relevant documents" as the goal of its systems. This goal has served us quite well, for quite some time, but it is quite different from the goal of "the support of people in achievement of the goal or task which led them to engage in information seeking".

In particular, the methods and measures associated with the standard goal, or variations of it, as have been developed for different information retrieval contexts, such as question answering, turn out to be inapplicable for evaluation of interactive information retrieval. Specifically, I would like to point out that the basic criterion for most information retrieval evaluation is *relevance*, and that this criterion refers to the relationship between one information problem and one information object. This condition is severely limiting when considering evaluation of interactive information retrieval, which takes place over one or more search *sessions*, each of which is a sequence of different interactions of person with information.

During such a session, people's interpretations of their information problems may change; their intentions in different parts of the session may vary, and may have nothing to do with the relevance of any document; and, the goal that led them to engage in information seeking may be satisfied in ways in which document relevance does not play a major role. Furthermore, is it even reasonable to think about the relevance of a search session as a whole? What then, could be an alternative to this criterion, and allow the development of measures and methods suitable for interactive information retrieval?

I, and others, have proposed that moving from relevance to *usefulness* as the main criterion for evaluation of information retrieval system performance, may provide an answer to this conundrum (e.g. Belkin, 2010; Cole et al., 2009). The basic reasons for this are that: usefulness is a more general concept than relevance, which it may, in certain circumstances, encompass; usefulness is directly applicable to the goal of supporting accomplishment of goal/task; and, usefulness can be applied as a criterion to evaluation of search sessions as a whole, and to the different intentions during a search session. I've provided argument supporting these statements elsewhere. Here, I'd like to support them through discussion of applicability of usefulness. First, I will discuss briefly how making this shift to usefulness could provide a basis for addressing many of the other challenges facing interactive information retrieval research that I've just mentioned. I'll then give an overview of how usefulness could be used in evaluation, and then provide some examples of how it might be implemented. Finally, I'll briefly describe some recent and current research we've been conducting at Rutgers, which attempts to actually use usefulness in understanding interactive information retrieval.

With respect to how starting from usefulness could provide a basis for addressing the challenges of interactive information retrieval research, consider the following progression:

- Evaluating an information retrieval system's usefulness in helping people to achieve a motivating task/goal necessarily entails the development of an empirically-based taxonomy of such tasks/goals;
- which in turn requires the development of methods for investigation of the behaviors and intentions of people prior to, during, and after engagement in an information retrieval system. Adapting methods of information behavior research to the specific requirements of information retrieval system design is a likely path to such outcomes.
- This, in turn, could lead, in one scenario, to
    - the ability to accurately simulate whole search sessions,
    - which might enable TREC-style evaluation of interactive information retrieval systems.
- or, in another scenario, to
    - associating loggable behaviors with specific search intentions, related to specific task types,
    - which might lead to the ability to analyze large-scale corpora of search sessions

according to identify and characterize task-specific behaviors in operational IR systems.

Furthermore, evaluation according to usefulness allows considering search sessions both as whole entities, and as coherent sequences of different types of interactions with information and system, each with its own specific support techniques and usefulness-based performance measures. Thus, establishing usefulness as the criterion according to which any specific aspect of interactive information retrieval should be evaluated could lead to principled research with respect to any of those aspects.

How could usefulness be applied in evaluation of support for achieving motivating task/goal? We can consider applying the usefulness criterion at three (possibly 4) levels with respect to an information seeking session.

- Level 1: how useful was the search session in helping the person to achieve the motivating goal or accomplish the motivating task;
- Level 2: how useful was the support offered by the system at each point during the search session in helping to respond to Level 1;
- Level 3: how useful was the support offered by the system at each point during the search session with respect to the specific goal of that support technique;
- Level 4: how useful was the sequence of search intentions to the accomplishment of the task/goal

We can then consider what one should or could measure at each level, and how such measurement could proceed. Here are some possible ways in which this could be accomplished, in an experimental or observational environment.

At level 1, we begin by first characterizing the motivating goal or task of the information seeking session. In an experimental mode, this could be done by controlling the task type, according to some well-defined taxonomy. In an observational or quasi-experimental mode, the task or goal which motivated the information seeker could be elicited and mapped to a taxonomy.

We are then in the position of being able to address evaluation of support over the whole search session, in several possible ways. One is to measure the extent to which the goal or task has been accomplished at the end of the search session. This measure of usefulness is both obvious, and obviously difficult to implement. Nevertheless, we already have examples of research which has been able to do this (e.g. Liu & Belkin, 2015). Another, also obvious method and measure is direct elicitation of judgments from the searcher (whether in controlled or uncontrolled situation) of the usefulness of the search session with respect to the goal/task. For this type of measure, we also already have some precedent, going back even to very early information retrieval research (e.g. Su, 1993; 1998; Tagliacozzo, 1977).

At level 2, the situation is a bit more complicated and difficult. A possible means for addressing usefulness evaluation at this level is to replay the record of a search session, eliciting judgments of usefulness of support at each interaction interval to the eventual accomplishment of the task/goal.

At level 3, it is first necessary to have a taxonomy of "within-search" intentions, such as: learning about a topic; finding good search terms; judging the usefulness of documents; finding good documents; comparing different documents, and so on. For each such intention, specific usefulness criteria and measures can be established, for instance relevance for the intention of finding good

documents, and nDCG as a measure. And, of course, explicit usefulness judgments from the searcher can be elicited.

Level 4 is really rather dicey, and I almost didn't include it. It seems at least possible that, for some specific combinations of goal/task type, search context and user characteristics, there might be "ideal" or "best" sequences of search intentions. If we are able to identify such situations, then comparison of an actual search sequence with the ideal, could be a possible measure of usefulness of system support. This would especially be the case if an aspect of system support were suggesting that the user carry out certain intentions at certain points in an information seeking session.

Obviously, there are many problems with the actual implementation of these suggestions. One is establishing methods that allow replicability and comparability between different studies and experiments. Some possible ways to address this problem is the development and use of consistent user models, using well-controlled task types, and possibly (eventually) simulation of search. Another obvious problem is the elicitation of usefulness judgments, which is difficult in many circumstances, and clearly subject to various types of bias. The most obvious way to deal with this problem is to rely only on "objective" measures. At the intention level, at least at level 3, it might be possible to establish ways to measure usefulness of different intention types, in just the way that measures have been established for relevance-oriented intentions.

But the real biggy is scale. All of the methods and measures which I've proposed do not scale well. One solution to this problem is simulation of interactive search, but we still can't do this at all well. However, it does seem to me that if a lot of small-scale studies are done, and done well and in commensurable ways, that the cumulation of such work could provide the basis for searcher and goal/task models which would allow relatively realistic simulation. Another, rather more likely approach, is to conduct relatively small-scale user studies in concert with analysis of large-scale search logs. The results of the detailed studies could be mapping of behaviors which can be logged in the large to the contextual characteristics identified in the small-scale studies. But I guess that my real message here is that a lot of work needs to be done first to overcome the problem of scale.

Now I'll briefly describe research at Rutgers (with Chirag Shah and Michael Cole), which attempts to understand and apply usefulness as an evaluation criterion. This encompasses two projects, which together have the aims of:

- Identifying coherent sequences of behaviors during an information seeking session;
- Relating behaviors to specific, different search intentions during an information seeking session;
- Identifying criteria for support for different search intentions;
- Testing measures and methods for evaluation of support for different intentions; and,
- Testing measures and methods for evaluation of support for controlled motivating task types

The first project, recently completed, addressed the first aim and has come up with some interesting results. One is the identification of a relatively small number (8-10) of coherent sequences of eye-fixation behaviors exhibited by participants in a controlled study in which they searched for information that would respond to four different task types. Of particular note is that different task types, and different facet values of task types, are associated with different sequences of eye-fixation behavior classes. These results give us hope that it is possible that observation of quite low-level behaviors will enable us to relate those behaviors to different search intentions, both with respect to the motivating task, and to search intentions associated search session segments. Aspects of these

results have been reported in Cole, Hendahewa, Belkin & Shah (2015). The second project builds upon the results of the first, and addresses the remaining aims just mentioned.

We are currently embarking on a new study, which attempts to relate observed behaviors to different search intentions within an information seeking session.[2] In this study, we have again four different motivating task types, constructed according to a task type classification scheme developed by Yuelin Li (Li & Belkin, 2008; 2010), somewhat modified by us. Participants are asked to conduct searches for information related to these different tasks, a whole bunch of their behaviors will be logged, including eye fixations, scrolling, clicking, copying, saving, query input, ..., and video of their search will be recorded. After each search, the video is replayed and participants are asked, with respect to each query interval in the search session (query interval is what takes place between one query and the next, including the initial and subsequent queries), to identify what they intended to accomplish during the query interval and with what success, and why they entered the new query. They also evaluate the usefulness, with respect to the overall motivating task, of any information objects they saved, in each query interval, and their reasons for that evaluation. They are also asked to evaluate their success in accomplishing the task, and to evaluate the usefulness of the system's support for accomplishing the task. These searches take place on the open web, using any web search systems the participants choose. The minimal outcomes hoped for in this stage of the research are: association of specific search intentions with specific sequences or classes of search behaviors; learning if elicitation of participant evaluation of system usefulness with respect to overall search task is feasible, valid and reliable; and, identification of some evaluation criteria specific to particular search intentions.
We'll then conduct a second study similar to the first, with a new set of task types. These results of these two studies will be combined and analyzed, in order to:

- Identify search intentions and criteria for evaluation of their accomplishment
- Identify relationships of behaviors to search intentions
- Identify "optimal" sequences of intentions/behaviors
- Characterize usefulness measures for accomplishing different task types

The final part of the project will take theses results, and use them in two different experiments, one in situ and another a controlled lab experiment, in order to test:

- Measures of usefulness of support for different search intentions, by comparing "objective" measures based on intention-specific criteria to usefulness judgments of those support techniques
- Methods for collecting data required for these measures
- Measures of usefulness of support for task accomplishment, by comparing "objective" measures of task performance to usefulness judgments of search system support

We are clearly in the very early stages of this investigation of usefulness as an evaluation criterion, but I think that what I've outlined demonstrates at least some possible ways in which that can be accomplished, and that the inherent problems can be addressed.

I hope that this overview of usefulness as the criterion for evaluation of interactive information retrieval has at least led you to think about the problem of how best to evaluate information retrieval systems under a goal of supporting people in achieving the goals/tasks which led them to engage in

[2] Characterizing and Evaluating Whole Session Interactive Information Retrieval: http://inforetrieval.org/iir/

information seeking. Even if you don't buy my proposal, I do hope that I've convinced you that what we have now in the way of evaluation methods and measures just will not do, and that we do need to find some alternate paradigm if information retrieval research is to remain, relevant.

6 In conclusion

I mentioned at the beginning of this talk that I thought that Gerry Salton and I, despite the differences in our approaches, may actually have shared the same overall goal for information retrieval. It seems to me that Salton was concerned with establishing and increasing what information retrieval techniques *can* do toward enhancing information seekers' interactions with information retrieval systems, and that in some ways limited how goals for information retrieval were set and addressed. I am sure, however, that he understood that these were indeed limits that were less than what would be the most desirable outcomes or goals. My position, in contrast, was that we should not feel constrained by these limits, and rather start with attempting to address the high level goal, that is, what we *ought* to do, or at least thinking about what starting at that point implies. I think that we are now at an intellectual and technical position in information retrieval where what we *can* do, if we're willing to take some risks and engage in some uncertainty, is much more than what we have been previously able to do, and come closer to what we *ought* to do, that is, achieving the real reason for doing information retrieval.

Finally, I can't help but feel that my being a recipient of the Salton Award is evidence of information retrieval's acceptance of including *people, interacting with information*, as a major research concern, and for this I sincerely thank the ACM SIGIR for bestowing this award, and for all those whose research has led us to this new point of departure for the field.

7 References

Belkin, N.J. (1977) Internal knowledge and external information. In: *CC 77: The Cognitive Viewpoint, International Workshop on the Cognitive Viewpoint*, Gent, 1977 (pp. 187-194). Gent: University of Gent.

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

Belkin, N.J. (1993) Interaction with texts: information retrieval as information seeking behavior. In *Information Retrieval '93: von der Modellierung zur Anwendung. Proceedings of the First Conference of the Gesselschaft für Informatik Fachgruppe Information Retrieva*l, Regensburg 1993. (Schriften zur Informationswissenschaft, Band 12) (pp. 55-66) Konstanz: Universitätsverlag Konstanz.

Belkin, N.J. (1996) Intelligent information retrieval: Whose intelligence? In: *ISI '96: Proceedings of the Fifth International Symposium for Information Science* (pp. 25-31). Konstanz: Universitätsverlag Konstanz.

Belkin, N.J. (2010) On the evaluation of interactive information retrieval systems. In: B. Larsen, J.W. Schneider & F. Åström (Eds.) *The Janus Faced Scholar. A Festschrift in Honour of Peter Ingwersen* (pp. 13-21). Copenhagen: Royal School of Library and Information Science. Retrieved on 7 January 2014 at http://www.issi-society.org/peteringwersen/

Belkin, N.J., Dumais, S., Kando, N. & Sanderson, M. (2012) *NII Shonan Meeting Report. Whole-*

Session Evaluation of Interactive Information Retrieval Systems. Tokyo: National Institute of Informatics.

Belkin, N.J., Oddy, R.N. & Brooks, H.M. (1982) ASK for information retrieval. Part I: Background and theory; Part II: Results of a design study. *Journal of Documentation*, vol. 38, nos. 2&3: 61-71; 145-164 . Part I reprinted in: *Readings in Information Retrieval*, K. Sparck Jones and P. Willett, eds. San Francisco, Morgan Kaufmann, 1997: 299-304.

Belkin, N.J., Seeger, T. & Wersig, G. (1983) Distributed expert problem treatment as a model for information system analysis and design. *Journal of Information Science*, vol. 5, no. 5: 153-167.

Brooks, H.M., Daniels, P.J. & Belkin, N.J. (1986) Research on information interaction and intelligent information provision mechanisms. *Journal of Information Science*, vol. 12, nos. 1&2: 37-44.

Cole, M.J., Gwizdka, J., Liu, C., Bierig, R., Belkin, N.J. & Zhang, X. (2011) Task and user effects on reading patterns in information search, *Interacting with Computers*, v.23 (4): 346-362.

Cole, M.J., Hendahewa, C., Belkin, N.J. & Shah, C. (2015) User activity patterns during information search. *ACM Transactions on Information Systems*, 33 (1): Article No. 1 (39 p.)

Cole, M., Liu, J. Belkin, N.J., Bierig, R., Gwizdka, J., Liu, C., Zhang, J. & Zhang, X. (2009) Usefulness as the criterion for evaluation of interactive information retrieval. In: B. Kules, D. Tunkelang & R. White (eds). *HCIR 2009, Proceedings of the Third Workshop on Human-Computer Interaction and Information Retrieval* (pp. 1-4). Retrieved on January 26 2010 at http://cuaslis.org/hcir2009/workshop_proceedings

Daniels, P.J., Brooks, H.M. & Belkin, N.J. (1985) Using problem structures for driving human-computer dialogues. In: *RIAO '85, Actes of the Conference: Recherche d'Informations Assistee par Ordinateur* (pp. 645-660). Grenoble, I.M.A.G. Reprinted as: Daniels, P.J., Brooks, H.M. & Belkin, N.J. (1997) Using problem structures for driving human-computer dialogues. In: *Readings in Information Retrieval*, K. Sparck Jones and P. Willett, eds. (pp. 135-142). San Francisco: Morgan Kaufmann.

Dumais, S.T. (2009) An interdisciplinary perspective on information retrieval (Salton Award Lecture). In: *SIGIR 2009. Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval* (pp. 1-2). New York: ACM.

Fuhr, N., Belkin, N.J., Jose, J. & van Rijsbergen, C.J. (2009) Workshop report [Dagstuhl] Seminar 09101 – Interactive Information Retrieval. Retrieved on 30 October 2015 at http://drops.dagstuhl.de/portals/index.php?semnr=09101

Li, Y. & Belkin, N.J. (2008) A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, v. 44 (6): 1822-1837

Li, Y. & Belkin, N.J. (2010) An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for Information Science and Technology*, v. 61: 1771-1789.

Liu, C., Cole, M.J., Baik, E. & Belkin, N.J. (2012) Rutgers at the TREC 2012 Session Track. In: *Proceedings of TREC 2012, the 21st TExt Retrieval Conference*. Gaithersburg, MD: NIST. Retrieved on 25 October 2013 at http://trec.nist.gov/pubs/trec21/papers/ruiiltrec.session.final.pdf

Liu, J. & Belkin, N.J. (2015) Personalizing information retrieval for multi-session tasks: Examining the roles of task stage, task type, and topic knowledge on the interpretation of dwell time as an indicator of document usefulness. *Journal of the Association for Information Science and Technology*, 66 (1): 58-81.

Liu, J., Cole, M.J., Liu, C., Bierig, R., Gwizdka, J., Belkin, N.J., Zhang, J., Zhang, X. (2010) Search behaviors in different task types. In *Proceedings of the 2010 ACM/IEEE Joint Conference on Digital Libraries (JCDL 2010)* (pp. 69-78). New York: ACM.

Liu, J., Liu, C., Cole, M., Belkin, N.J. & Zhang, X. (2012) Exploring and predicting search task difficulty. In: *CIKM 2012: Proceedings of the 21st ACM International Conference on Information and Knowledge Management* (pp. 1313-1322). New York: ACM.

Oddy, R. N. (1977). Information retrieval through man-machine dialogue. *Journal of documentation*, *33*(1), 1-14.

Spärck Jones, K. (1988) A look back and a look forward. In: *SIGIR 1988. Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval* (pp. 13-29). New York: ACM.

Su, L. T. (1992). Evaluation measures for interactive information retrieval. *Information Processing & Management*, *28*(4), 503-516.

Su, L. T. (1998). Value of search results as a whole as the best single measure of information retrieval performance. *Information Processing & Management*, *34*(5), 557-579.

Tagliacozzo, R. (1977). Estimating the satisfaction of information users. *Bulletin of the Medical Library Association*, 65, 243-248

Wersig, G. (1972) *Information – Kommunikation - Dokumentation*. Pullach bei München: Verlag Dokumentation.

Wersig, G. (1979) The problematic situation as a basic concept of information science in the framework of the social sciences; a reply to N. Belkin. In: *Theoretical problems of informatics: New trends in informatics and its terminology* (pp. 48-57). Moscow: International Federation for Documentation.

8 Appendix: Names and locations in Figure 1 of participants in the First International Research Forum in Information Science

Bottom row, left to right: Maria Dembowska, Polish Academy of Science, Warsaw; Bob Bottle, City University, London; John Martyn, Aslib, London; Ole Harbo, Royal School of Librarianship, Copenhagen; Ulrich Windel, Freie Universität, Berlin, FRG; Liz Wilkinson, Loughborough University; Josè Marie Griffiths, University College, London; Stephen Robertson, University College, London.
Middle row, left to right: Ferdinand Leimkuhler, Purdue University, West Lafeyette, Indiana; Belver Griffith, Drexel University, Philadelphia; Nick Belkin, University College, London; B.C.Brookes, University College, London; Barbara Wright (Forum Secretary) University College, London; Karen Spärck Jones, University of Cambridge; Mrs. Kochen; Manfred Kochen, University of Michigan, Ann Arbor; Mrs. Slamecka
Top row, left to right: Alan Gilchrist, Aslib, London; Silvio Ceccato, University of Milan; Verina Horsnell, Polytechnic of North London; Jean Meyriat, Fondation Nationale des Sciences Politiques, Paris; Marc de Mey, University of Ghent; Vladimir Slamecka, Georgia Tech, Atlanta; Slamecka son.
Entering, stage left: Michael Heine, University of Newcastle-upon-Tyne
Not pictured: Michael Brittain, Loughborough University; Bill Cooper, University of California, Berkeley; Robert Fairthorne, North Farnborough, England; William Goffman, Case Western Reserve University, Cleveland, USA; Bernard Houghton, Liverpool Polytechnic; Michael Lynch, University of Sheffield; Margaret O'Hare, British Library R&DD, London; Janet Rennie, City University, London; Kjell Samuelson, Stockholm University; Alina Vickery, University of London; Brian Vickery, University College, London; Isaac Welt, The American University, Washington, D.C.