4.2.4 Broader impact

This proposed research has the potential to empower people through the attainment and mastery of better information search skills and the provision of search environments that support deeper learning. This should lead to better work products, more fulfilling leisure pursuits, increased opportunities for life-long learning, and greater self-actualization. The work of this proposal also brings together researchers from many different areas including IR, interactive IR, human-computer interaction, information-seeking behavior, psychology, education, and library science.

4.2.4.1 Obstacles and risks

There are several obstacles and risks. First, users must make an explicit choice to use these tools. These tools are likely to interrupt and disrupt a comfortable searching style. We will need to make tools that will lead to meaningful and positive outcomes to motivate adoption. Another risk is that the development of tools that highlight particular characteristics of content, such as authority, might lead to adversarial web page authoring and/or introduce other types of search bias. Such tools might also be viewed as overly paternalistic and controlling. Finally, such tools might lead to the establishment of new comfort zones that do not actually lead to higher levels of learning.

4.3 Finding What You Need with Zero Query Terms (or Less)

Future information retrieval systems must anticipate user needs and respond with information appropriate to the current context without the user having to enter a query — or even initiate an interaction with the system. In a mobile context such a system might take the form of an app that recommends interesting places and activities based on the user's location, personal preferences, past history, and environmental factors such as weather and time. In a traditional desktop environment, such a system might monitor ongoing activities and suggest related information, or track news, blogs, and social media for interesting updates. In any case, such systems must allow users to quickly act on the information and suggestions. While such systems would generally remain unobtrusive, waiting for the user to initiate an interaction (but providing "zero query terms"), sometimes the system might proactively interrupt the user to provide critical information (which we call "less than zero query terms").

4.3.1 Motivation

The need for these systems increases in mobile environments, where the user's ability to interact with the system is hampered by the physical limitations of the devices. On the other hand, development of these technologies is enabled by the context provided by mobile devices, which can provide a detailed account of the user's location, movements, activities and interests. Overall, much more of a person's life is online and always available, particularly through social media and other online interactions.

In contrast to traditional search engines, these systems must function without an explicit query, depending on context and personalization in order to understand user needs. In contrast to traditional recommender systems, these systems must be open domain, ideally able to make suggestion and synthesize information from multiple sources, involving multiple people, objects and actions.

In one form, we imagine a personal assistant who provides a key document at just the right time, sends a meeting summary to someone's mobile device just as they sit down, or even "whispers in one's ear" short biographical facts about the people at a meeting. While few people can afford to hire a personal assistant to perform these tasks, core technologies are now in place to automate them. For example, the DARPA CALO project — and Apple's Siri, its iPhone spin-off — has already examined core tasks in this area. Here, we propose a stronger, IR-oriented focus on automating the search process in the context of current activities.

In another form, we imagine a system that automatically gathers information related to an upcoming task. For example, if someone were planning to write a report during a long plane trip, they might find the necessary background information already available in a folder on their laptop. In order to achieve this goal, the system would need to be aware of current and planned activities, automatically gathering and organizing information in a forward looking fashion.

In an extreme version, we imagine someone's phone ringing as they walk down the street, interrupting their thoughts with the message that the love of their life is sitting in café they are just walking past. In this case, the urgency of the information need is judged to outweigh the annoyance of the interruption. In order to reach this level of performance, deep insights into personality and preference are required.

4.3.2 Proposed research

Many core IR issues are related to this problem, particularly given the increasingly rich, personal and heterogeneous signals and domains involved in these systems. Research in this area requires new representations of information and user needs, along with methods for matching the two and presenting the results.

Other problems include methods for modeling person, task, and context; methods for finding "objects of interest", including content, people, objects and actions; and methods for determining what, how and when to show material of interest.

4.3.3 Research challenges

This research requires efforts to study users, interpret user behavior, prototype systems, and develop appropriate evaluation methodologies, and there are substantial challenges in all of these areas.

User-related challenges include: methods for dealing with rich interaction sequences; use of multi-modal sensor data; open domain, time- and geo-sensitivity; trust, transparency, privacy; determining interruptibility; summarization – e.g., why the system made this suggestion for this person, now and here; amount of time or information required – e.g., 5 minutes to kill vs. 15 hours on a plane; and log and interaction analysis.

Prototype development challenges include data gathering and synthesis; power management in mobile contexts; user interface/interaction, particularly in mobile contexts; and deployment and logging of data.

Evaluation challenges include the development of methodologies to assess the quality of specific systems and suggestions, as well as the creation of appropriate test collections and methodologies to allow results to be compared across research groups.

4.3.4 Broader impact

We foresee three broad areas of impact:

1. Filling information gaps *on demand*: For example, a desktop app might populate a start screen with personalized information and updates, or a mobile app might suggest ideas to fill free time. Neither app would require the user to enter a query; interaction may be limited to browsing information and rejecting ideas.
2. Proactively *whispering in one's ear*, perhaps through a screen on mobile device, or by literally whispering in one's ear through a headset. Proactive information gathering might also take place on a non-real time basis, such as gathering information for a forthcoming plane trip.
3. *I Really Mean It Now!* Identifying when and how a user can and should be interrupted to provide essential information, either because of a negative event, such as an emergency, or a positive event, such as finding the love of one's life.

4.3.5 Obstacles and risks

Achieving success requires a close interaction with numerous fields of computer science, including information agents, data mining, ubiquitous computing, NLP, and HCI.

4.4 Mobile Information Retrieval Analytics (MIRA)

During the last decade people have begun carrying mobile devices and using them for a variety of communication, social interaction, information seeking, and other routine tasks. In spite of their ubiquity, no company or researcher has an understanding of mobile information access that spans a variety of tasks, modes of interaction, or software applications. This lack of understanding is an obstacle to scientific study and the development of new tools that provide more effective information access.

The first stage of this project develops a methodology and tools for large-scale collection of data about mobile information access. The information gathered provides the foundation for a second stage, in which we will develop benchmark tasks, test collections, and evaluation methodologies for community-wide research initiatives. The project will be a success when the tools and resources developed are used by subsequent research projects that develop improved information access technologies.

4.4.1 Motivation

Mobile devices are an important source of information for much of the public. Tools and technologies developed for desktop and fixed computing platforms can't capture much of the information about how people use mobile devices. A company that provides mobile devices, software, or services can capture some types of information, but usually is unable to construct a view that spans multiple applications. For example, a search service provider might know that a query was issued, but not know whether the results it provided resulted in consequent action. The lack of a more comprehensive understanding of a person's mobile information seeking and usage prevents progress on a variety of important research questions.

How a person's information need (e.g., a query) interacts with contextual features, such as the person's location, platform, and behavioral pattern is an important topic for research, one that has been studied as an issue of interaction design, but is largely unexplored within the IR community.

People have large social networks, however the role and value of each individual is context-specific. The mobile environment is ideal for studying social network activation dynamics and how a person's global social network is refined into task- or setting-specific social networks.

The identification of common *types* of web search queries led to query classification and algorithms tuned for different purposes, which improved web search accuracy. A similar understanding for mobile information seeking would focus research on the problems of highest value to mobile users.

The mobile environment enables study of how information seeking spans apps and services. For example, a person may check foursquare to find restaurants, search for restaurant reviews on yelp, and then phone to make a reservation. Understanding cross-app interaction patterns enables development of context-specific authority metrics, study of cross-modal information seeking (e.g., text, voice), and research on how online activities lead to user action (and vice versa).

Mobile devices have small screens and can be difficult to hear in noisy environments, thus they are a unique environment in which to study *what* information, what *kind* of information, and what *granularity* of information to deliver for different tasks and contexts.

4.4.2 Proposed research

The project consists of a first phase that develops a methodology and tools for collecting data about mobile information access, and a second phase that develops benchmark tasks, test collections, and evaluation methodologies that enable reproducible research.

4.4.2.1 Developing a Holistic View of Mobile Information Access

The first project component is a methodology and tools for doing large-scale collection of data about mobile information access. The methodology uses software applications installed on a person's mobile device to capture information about how the device is used. A toolkit is developed to provide basic logging capabilities. The toolkit can be deployed within different types of applications, for example, a passive monitor, a game, or an application that provides a useful service. The toolkit can also be used to capture different types of information, to support different research agendas. The goal is to develop applications that might be installed on several thousand devices.

There is precedent for this activity in other settings. Commercial search engines use browser toolbars to collect information that is used to improve search services; the public generally considers this acceptable and benign. Spyware is embedded within software without a user's knowledge to capture information for unknown purposes; the public generally considers this unacceptable and a threat. Capturing broad mobile information in a socially-acceptable manner requires research on several topics.

Research on incentive mechanisms is required to understand situations in which people are willing to allow their behavior to be monitored. For example, small monetary rewards, free games, social good, and a useful free service are all incentives that are used successfully in other settings. Research on acceptable practices is required to understand what features and practices provide a sense of transparency for consumers, for example, periodic opt-in, and easy removal. Research on privacy is required to understand what can be protected by dataset licenses alone, what must be anonymized, and tradeoffs between anonymization and data utility.

The result of the first phase of the project is a methodology, a set of tools, and a set of best practices that support the collection of useful data about mobile information access in a variety of situations.

4.4.2.2 Community-Wide Evaluation and Participation

A large data gathering effort is only worthwhile if it enables high-quality research, thus the second project component is the development of well-defined information seeking tasks and supporting data collections that represent important real-world mobile information seeking situations. The tasks and data collections would be designed to support quantitative evaluation in well-defined evaluation frameworks that lead to repeatable scientific research. They would be deployed in large-scale community-wide evaluations of information retrieval research such as TREC, CLEF, NTCIR, or FIRE.

Annual evaluations such as those attract many of the best researchers from around the world. They focus the attention of a broad and high-quality research community on a small set of specific problems. They also involve that community in establishing well-defined problems and evaluation methodologies that produce repeatable science and become standards for the scientific community. Engaging these evaluation forums to shape scientific research and establish a long-term research agenda for mobile information seeking is of the highest priority.

4.4.3 Broader impact

This project is enabling technology that is a catalyst for groundbreaking research on mobile information access. It develops a data-gathering framework, a software toolkit, and a set of "best practices" that enable data collection about information seeking on mobile platforms in a manner that university internal review boards (IRBs) will find acceptable. It uses the collected data to develop a set of representative and well-defined tasks and data collections that can be used in community-wide research forums, thus enabling and supporting research by a broad scientific community.

4.4.4 Obstacles and risks

The project will need to address four important obstacles. First is developing incentive mechanisms that will cause enough people (several thousand people) to install software that allows their activity to be monitored. Second is developing data collections that are sufficiently detailed to be useful while still protecting people's privacy. Third is collection of data in a manner that university internal review boards will consider acceptable ethically. Fourth is collection of data in a manner that does not violate the Terms of Use restrictions of commercial service providers. None of these obstacles are insurmountable.

4.4.5 Related initiatives

There are several potentially-related initiatives. The TREC 2012 Contextual Suggestion Track will study how context and user interests affect web search. The *zero-query* research task (Section 4.3) also studies how to deliver information proactively, using only information about a user's interests and context. The Center for Embedded Network Sensing (CENS) at UCLA has developed projects and tools that capture data from mobile phones, and thus might have expertise and resources that would contribute to the project.

4.5 The Structure Dimension

A key research question to be addressed by IR researchers in the next decade is: How do we move beyond simple document retrieval? Better integration of structured and unstructured information to seamlessly meet a user's information needs is a promising, but underdeveloped area of exploration. Can we take advantage of the synergies between linked data, information extraction, collaborative editing, and other structured information to improve search breadth and quality?

4.5.1 Motivation

All data has structure, whether it is explicit or implicit. Even classic document retrieval assumes a structure where full-text documents are delineated. However, the structural dimension stretches beyond plain document layout to include type information identifying entities, user profiles, contextual annotations, as well as (typed) links between information objects ranging from web pages to social media messages. Users routinely access and amalgamate information from multiple inputs while interacting continuously within a virtual environment. While harmonizing various heterogeneous inputs from a user's environment is a major challenge, new opportunities to improve the search experience arise – humans can take an active role in the information seeking processes. For example, human computation in a crowd-sourcing platform or "friend-sourcing" information requests in a person's own social network could be integrated into the search experience. Mixing structured and unstructured data representations is not a new research idea. However, recent changes in how users access information on the Internet are increasing the importance of moving beyond traditional ranked documents retrieval. The real challenge is that the underlying structure may be hidden in the data or even in the representation. Related fields are making progress in uncovering this structure – incrementally driven by human effort, as in the data spaces abstraction proposed for data management or the development of the semantic web; or, automatically created from natural language processing by machines reading the web and the heterogeneous information networks discovered with techniques originated in the data mining community. In spite of the increasing availability of structural information, considerable work must be done before we can fully utilize these new models to significantly improve the information seeking experience.

Clearly, the information retrieval community is well positioned to investigate and remove uncertainties arising in this process, whether these are caused by the selection of heterogeneous resources or by the unification of varying structure and quality of these resources and their annotations.

4.5.2 Proposed research

Are user information needs really answered by a list of documents without further processing? Consider a variety of everyday contexts that may trigger information seeking: pruning an apple tree, going on a trip, assessing a job applicant, or deciding whether the beach one is visiting is safe for swimming. In all of these scenarios, structural information may help scope the information need to the most relevant subset of resources to consider, and improve the results presented by giving more focused answers.

Modern web search engines can already identify verticals to provide better answers for queries involving products, locations, restaurants, movies and artists. E-commerce systems can dynamically select and create facets to support interactive exploration. Domain-specific websites, like IMDB and Rotten

Tomatoes for movies, consider a rich result list constructed from different answer types. Entities are key; relatively unambiguous pieces of information, as anchors for or pivots between the user's information need and the representations available to the system. Named entities have an equally important role in digital cultural heritage, where they are the key to provide access to multimedia artifacts.

4.5.3 Research challenges

A fundamental challenge in synthesizing structured and unstructured collections is the development of better approaches to represent information. Here, we consider three dimensions of representation: query, collection, and result presentation.

Move beyond simple keyword queries. For a system to accurately find and rank different information from disparate collections, a new approach to mixing term queries, Boolean operations, or other relational constraints more intuitively is needed.

Design storage representations capable of supporting efficient free form queries. Collections may be highly dynamic, privacy preserving, or contain various types of unrefined data. How can we represent the collection in a way that would allow imposing a desired structure at query time? Can we defer statistical modelling (or even support exact matching) efficiently to query time?

Improve result presentation. How do we construct a result that mixes aesthetic and functional aspects appropriately? What type of evaluation framework is needed to quantify the quality of the new integrated results?

Many related challenges arise. How do we conduct effective and efficient search in hybrid networks of structured and unstructured data, and apply constraints at query time to find, process, and synthesize multiple, loosely cooperating data repositories simultaneously? Query processing techniques such as SPARQL, XQuery, SQL, NEXI, XIRQL and the INDRI query language are useful for inspiration – but are these languages sufficient for our needs? How to deal with uncertain information within structured and unstructured data? Links between entities in knowledge sources in repositories such as Wikipedia may be incomplete and noisy. Sampling from heterogeneous and distributed sources inherently leads to uncertainties about the underlying structure. Structured data generated by information extraction components may be associated with confidence scores. It is important to design a framework that can naturally deal with the uncertain information. Inquery's inference networks and semi-structured relevance models provide a starting point, but we need new approaches to dealing with the uncertainty about the imposed structure; the (semantic) gap between structured data and unstructured data; and, efficient solutions that generate desired results in interactive time.

For all three dimensions - queries, collections, and presentation - we need new evaluation methodologies that allow us to determine how effectively and efficiently we are delivering the expected results to end-users. In this context, we need to go beyond the traditional evaluation paradigm and strive for developing benchmarks and tasks which are able to combine the assessment of structured retrieval, as they do in the semantic and database communities, with unstructured retrieval, as it is traditionally done in the IR community. Evaluation must address efficiency and effectiveness in concert, not independently as is the usual case in our field. Also, the capacity of handling heterogeneous sources would be a highly desirable feature to consider. The goal is to reduce the barrier-to-entry and better utilize structure in answering complex information needs. Can we design a system that is flexible enough to express the models used in approaches to TREC, INEX, and CLEF?

4.5.4 Broader impact

The potential impact is seamless support for complex information seeking tasks. Envision an example application "Sherlogue": a computer generated Wiki, where each user's search result is an interactive wiki page that presents multi-faceted answers. For example, the query "Design a new course offering on Information Retrieval" would produce a result page including a syllabus, lectures slides, assignment test banks, videos, tutorial write-ups, useful references, links to area experts, source code and tools; editing the wiki result interacts directly with both the retrieval engine (machine) and the user's social networks (human).

4.5.5 Obstacles and risks

Addressing heterogeneous structural views over multiple data collections will require advances in almost every sub-discipline of IR, varying from efficiency to understanding and evaluating multi-valued relevance.

4.6 Understanding People in Order to Improve Information (Retrieval) Systems

Despite widespread acknowledgement that the understanding of users is essential for the creation, improvement and evaluation of IR systems, there is still a large gap between the study of users and the study of IR algorithms. Hence, we propose the development of a research resource for the IR community, from which hypotheses about how to support people in information interactions can be developed, and in which IR system designs can be appropriately evaluated.

4.6.1 Motivation

All IR systems have the purpose of supporting people, through interactions with information, to achieve their goals and underlying intentions in work, academic, everyday life situations. In order to design and evaluate such systems, it is necessary to understand the goals that lead people to engage in various interactions with information, as the goals of systems must be commensurate with the goals of the people whom the systems are designed to support. Such systems will need to "understand" people's behaviors during their interactions with information, the problems they have in realizing their intentions, and the general nature of their information problems. Although there has long been a consistent call for IR research and practice to base their activities on understanding of the people for whom the systems are intended, both the theoretical arguments for this, and the empirical results which have resulted, have been largely ignored by the IR system design and evaluation community. As a consequence, both IR research and IR evaluation standards and methods are proving to be inadequate for the new types of interactions with information in which people engage, and the new types of support systems envisioned for the future. Thus, the goals of our proposed program are to: provide basic data according to which characteristics of goals, intentions, behaviors can be identified across a variety of contexts; develop a research resource for the IR community, from which hypotheses about how to support people in information interactions can be developed, and in which IR system designs can be appropriately evaluated; and, provide insight as to how search interaction characteristics are shared or differ among, for instance, different user groups, search tasks, cultures, and languages of searching.

What is required to achieve the goals of the program is the systematic investigation and characterization of the goals, intentions, and information interaction behaviors of people across a wide variety of contexts and situations, with specific reference to IR system design and evaluation. To date, there has not been such research; this program aims explicitly to address this gap.

4.6.2 Proposed research

In order to achieve these goals, we propose an integrated program of studies of people before, during, and after engagement with information systems, at a variety of levels, using a variety of methods. These should include (but not be limited to) a range of levels from the individual, the group, and the community; and a range of methods from ethnography, *in situ* observation, controlled observation, experiment and large-scale logging.

The basis of the proposed program is the establishment of a set of standard, minimum protocols for the conduct of studies and of data collection relevant to different levels and contexts of study, applicable to different types of methods.

We provide general descriptions of two such protocols, as examples of how such a program would work, and how the results of different types of studies could inform one another.

Controlled observation of people engaged in interactions with information. A standard protocol would include: detailed specification of the tasks that participants in the investigation are asked to perform, where these tasks are presented as "simulated work tasks"; detailed demographic description of the participants in the investigation, and instruments that elicit participants' prior experience with relevant systems; instruments which elicit the participants' knowledge of the tasks and their topics, their expectations of difficulty of the task and estimates of their likely success, prior to engaging in the interaction; complete client-side logging of the information interaction associated with the task; instruments which elicit participants' evaluation of task difficulty, of their success in the task, and of the value of system features (if relevant) for task accomplishment; instruments which elicit evaluations of the usefulness of information encountered during the interaction for task performance;

Large-scale logging of search session interactions. A standard protocol for server-side search logging would include: no task specification, just completely natural search behavior; logging the content of the search results page and clicks on that page; logging of limited contextual information, such as the user's location, time of day; logging of implicit indicators including: 1) did they click, 2) did they dwell, 3) did they return next week.

Large-scale logging on the server side gives a less rich record of individual actions than the detailed observation described before. However, because it incorporates many people's actions, it is extremely valuable for: 1) Distribution of queries, 2) Distribution of clicks for each query, indicating user "intent", 3) How query-click distributions vary according to other factors such as location or time, 4) Identifying overall use cases, based on patterns in the logs.

By following protocols as in the two examples, it will be possible to link large-scale, server-side logging, which has limited kinds of data, but lots of examples; to smaller-scale, client-side studies, which have very rich data, but very few examples.

A result of this program will be establishment of a research resource consisting of the records of information interactions collected by many groups, at many sites, in a large variety of specific contexts and situations. This resource will be made available to the IR research community at large, and will enable the principled study of similarities and differences in goals, intentions, associated behaviors, and success in task performance for many different types of tasks, in many different situations. This resource will also provide an infrastructure for evaluation of IR systems, at both traditional levels of evaluation (e.g. relevance-based), and more especially, evaluation of support for information interaction episodes (e.g. search sessions) as a whole, which has not been possible before.

An integral aspect of this program will be sharing tools that help to implement the protocols among all groups who participate in this endeavor. It must be noted that our proposal requires a site that is responsible for maintaining and distributing the protocols, receiving and curating the data from the cooperating sites, integrating the reports into a single database, and affording access to the resulting database to the IR research community.

4.6.3 Research challenges

Developing a framework for the research resource that is easy to understand and use will be a major issue.

Challenges facing the research program include the following: agreement on standard protocols amongst the research community; construction of the research resource, and its maintenance; cost of data collection; dirtiness and sparseness of data; coordination of data collection for at least minimally compatible data; instrumenting logging; and anonymizing.

4.6.4 Broader impact

A systematic categorization of the goals, tasks and intentions of common information interaction tasks that people actually undertake is essential for building and evaluating systems for information retrieval in an increasingly diverse information ecology. Furthermore, the establishment of a research resource of records of information interaction sessions will enable such research in ways that have previously been impossible.

More widespread adoption of user investigation will not only support and inform the development of specific approaches; having a common base of data that is collected will also enable a wider examination of information seeking behavior that can contribute to the development of the field as a whole.

4.6.5 Obstacles and risks

IR community inertia, lack of expertise in appropriate methods, cost of relevant studies, finding a site and people to maintain the research resource, funding for such a site.

5 Topics Discussed Briefly

This section includes short discussions of research directions, project ideas, or challenge areas that were discussed in less detail at the workshop. In particular, these represent the 21 areas that were nominated as interesting by the first-round breakout groups but were not voted as "most" interesting (see Section 2.2).

It is important to highlight that this list is not expected to be exhaustive, even when combined with the previous section. There are many ideas that were presented within individual breakout sessions that are exciting and interesting, but did not receive enough support from that session to be nominated to the larger group. Nonetheless, this represents an exciting group of proposals.

The topics are listed in alphabetical order.

5.1 Abstracting Information Retrieval Evaluation

This project aims at abstracting the constituents of IR evaluation to allow for easier understanding, comparison, re-use and application of experimental results. It will develop methods, algorithms, and an open infrastructure to address the diversity of different evaluation tasks, activities, and systems.

Motivation. IR evaluation is challenged by variety and fragmentation in many respects – diverse tasks and metrics, heterogeneous collections, different systems, alternative approaches for managing the experimental data. Not only does this hamper the generalizability and exploitability of the results but it also increases the effort and cost needed to produce such experimental results and to further exploit them. Currently, the development of new data sets, tasks, and metrics requires large overhead for organizers and participants. Abstracting over these constituents as well as over the obtained results is crucial to scale-up evaluation. While defining these abstractions is not new, the problem has not been addressed systematically as a community and the existence of partial or overlapping solutions favors fragmentation rather than shared understanding and re-use.

Proposal. Abstracting evaluation infrastructure requires new data models, modular architectures, scalable solutions, and interoperability to manage, make accessible, curate, and enrich ever increasing amounts of experimental data. Abstracting across evaluation tasks requires shared representations of information units across tasks and their associated metrics as well as efficient assessment of a (presumably) larger set of information units than documents. Abstracting across evaluation runs requires dealing with sampling bias, non-stationary collections and relevance, and re-using data from historic interaction sequences.

Challenges. The proposed abstractions require both (1) cross-disciplinary competencies, coming from information retrieval, databases, and statistics, and (2) combining them in an effective way which ranges from creating unifying models to designing and developing infrastructures supporting them. Furthermore, openness and community involvement are fundamental to ensure that a consensus is reached and approaches are shared. Finally, funding is crucial for the generation of ideas, for the development, and for the sustainability over the time of the resulting infrastructures.

Related efforts. Nugget-based test collection creation represents an example of abstraction over tasks and information units. Evaluation infrastructures, such as the DIRECT system used in CLEF, community repositories, such as EvaluatIR.org, or the Open Relevance Project by the Apache Software Foundation are cases of systems trying to abstract over collections, evaluation activities, and evaluation runs.

5.2 Adapting to Various Sites, Tasks, and Contexts

Numerous tasks and applications cannot be served by standard off-the-shelf search engines. We lack the methodology and tool support to adapt an engine to its usage context, be it automatic or through expert intervention. We need approaches for the design, evaluation and deployment of adaptive systems.

Motivation. The presumption that a general purpose search engine can fulfill all needs of a specific site, a specific user group, or a specific collection without parameter tuning is wrong. Search as encountered in its most general form on the web is highly effective and convenient for a majority of search transactions. However, for the numerous specific needs and tasks in various organizations and self-selected user groups and communities, information seeking can be a cumbersome process which is only partially supported: multi-lingual and cross-cultural issues, quality assurance requirements, in-house jargon, etc., interact to make site-specific and adaptable search technology a necessity. Since users nowadays expect similar convenience and effectiveness from in-house system that they are used to in a web context, many organizations outsource their search needs to web search site-level indexes. In practice however, a tailored enterprise search solution would be most effective, if not too costly.

Proposal. This activity aims to formulate a design, testing, evaluation, and application framework for the intersection of task models, use cases, dynamic knowledge representations, and structured information, including existing installed systems.

As an example evaluation of the resulting work, consider pointing a set of adaptable enterprise search systems at a new site. Compare their representation of the site with respect to index terms, inferred concepts, identified and highlighted divergence from general language use, relation of inside information to outside information such as other known sites, conceptual models, and relevant data streams.

Challenges. There are countless ways in which sites, tasks, and contexts can vary, and settling on one or more to explore will require the availability – preferably broad availability – of data and users. It is not clear in advance how to evaluate the success of adaptive systems given the interplay between the variables.

5.3 Axiometrics – Foundations of Evaluation Metrics in Information Retrieval

Already around 100 IR effectiveness metrics exist, and more keep appearing. This project aims at understanding the relationships among them, in terms of both axiomatic properties and statistical relations, for both metric science (understanding of metrics) and engineering (development of metrics).

Motivation. The choice of the effectiveness metric that we use in our evaluation experiments depends on the current fashion. From a practical point of view, the current situation is that many researchers simply use the most popular metric, without further investigation into its suitability for the problem at hand. There is also the temptation for researchers to choose, among all available metrics, those that help to corroborate their claims. To avoid this situation, we need a better understanding of metrics. In addition, new metrics are being proposed continuously, and many of them rely on an arbitrary choice of parameters (for example, the logarithm discount function in (n)DCG). A more complete understanding of these factors would help researchers to make more informed choices.

Both companies and stakeholders will benefit from Axiometrics because of the better understanding of the properties and suitability of a metric for the task at hand. Companies will not waste time and money in tuning their systems according to the wrong metric for their purposes. Stakeholders and customers will be able to better interpret and understand whether the performance reported according to a given metric are really representative of their needs and whether a reported improvement will have an actual impact on their business.

Proposal. There have been some studies in this direction, which are the natural starting points: some apply formal constraints to define suitable metrics, a few present their evaluation metrics using a probabilistic framework. Measurement theory is also a useful tool. The ultimate goal would be to devise a general framework capable of taking into account the various notions of relevance (binary, categories, etc.), retrieval (binary/set, ranking, categories, etc.), and metric requirements (axioms, formal constraints and empirical properties that any metric must/should satisfy). Models of users and tasks can be used as a starting point for designing the axioms. Such a framework would help to increase the transparency and understandability of metrics.

5.4 Before and After the Mobile Query: A Holistic Approach to User Context

Understanding a user's offline and online context is the key to increasing user engagement with mobile information and media applications. Furthermore, in mobile devices, user engagement is a better metric for the success of a search application than traditional precision/recall-based approaches.

Motivation. Because of the limited real estate on a mobile device, metrics based on precision and recall are of limited utility. We propose that the user's engagement with the application is a more reliable metric for relevance and success in this domain. Users engage with an application because it provides them a satisfying experience in their interaction with information and media. Because mobile devices are carried with the user on the go, the user's engagement with an application is just as dependent on their offline context as it is on their online context.

Proposal. We propose a study of offline context and its relation to the online context of a user's interaction. This involves employing a wide range of sensors in the device (gathering ambient sounds, spatial directionality, temperature, etc.) as well as cross-application data gathering, and ethnographic studies to know what a user was doing at the time of interaction. We propose to use this knowledge to design better experiences by understanding the user's engagement with the device as a whole, and with a given application in particular.

Challenges. Ethnographic studies are difficult to do on a large scale. Users themselves may not be aware of why they find an application engaging, which will make it difficult to elicit information from them about this. As for engagement, there are many variables that affect engagement (the user's offline context, the application interface, whether the user has an information need or an entertainment need) which will need to be decoupled in order to be understood. Finally, we need to understand the relationship between engagement and satisfaction.

Related efforts. Related initiatives include the context-aware cell phone within MIT's MIThril project, the University of Helsinki's freely available ContextPhone platform, and the "reality mining" effort of Eagle and Pentland that seeks to mine data collected from cell phones. This topic is also strongly related to the Mobile IR Analytics (Section 4.4) and Zero Query Terms (Section 4.3) topics.

5.5 Community Evaluation Service

This project will provide an "experimental IR evaluation service" available to the community for research and educational purposes. The critical challenge is to provide funding resources and on-going support.

Motivation. Evaluation using large data sets is often desirable in Information Retrieval in order to provide corroboration of claimed improvements in search effectiveness, search efficiency, or both. But running experiments on terabyte-sized datasets requires sophisticated software tools, non-trivial computing hardware, and access to suitable data sets. Large research groups acquire these resources over multi-year multi-project spans, and experimental momentum and know-how is passed from one cohort of students and postdocs to the next; often with ongoing professional software engineering support, and extensive codebases. But small research groups lack these resources, and face ever-mounting barriers to entry. Even the largest of research groups face occasional "drought" periods when key personnel leave, or when funding is temporarily unavailable.

Proposal. This project will provide an "experimental IR evaluation service" that has the following desiderata: low barriers to entry for standard collections and tasks, recording and validation of experimental runs (for future baselines), incorporation of new data sets as quickly as possible, need for only limited local resources for researchers, and flexibility of processing. The project will be a success if a new graduate student is able to do their first experimental run against a ClueWeb-sized collection within a month of commencing.

Challenges. Provision of ongoing funding for such a service is clearly a challenge. We suggest that the project be funded via a combination of modest subscription fees or membership fees from researchers making use of the services, of the order of perhaps US$1-2k per year per user; specific research grant funding, of the order of perhaps US$10-50k per year, from national or international funding bodies; and philanthropic donations from major search services, of the order of perhaps US$50-100k per year from each of three or more companies.

Related efforts. Search systems and toolkits such as Lemur and Terrier provide related support. A workshop on open-source IR will occur at SIGIR 2012.

5.6 Exploring the Intersection of Social and Algorithmic Search

Global growth in Internet connectivity and participation is driving a renaissance in *human computation*: use of people rather than machines to perform certain computations for which human competency exceeds that of state-of-the-art algorithms (e.g. AI-hard tasks such as interpreting text or images). While integrating human computation with algorithms has potential to greatly improve state-of-the-art performance over today's IR systems, realizing this potential will require new research to ascertain how this emerging computing paradigm can be most effectively employed in the field of IR.

Motivation. Just as cloud computing enables us to utilize vast Internet computing resources on demand and at-scale, *crowdsourcing* lets us similarly call upon the online crowd to perform human computation tasks on-demand and at-scale. This newfound ability to integrate human computation alongside algorithms greatly expands traditional accuracy-time-cost tradeoffs and represents a disruptive shift in design and implementation of computing systems. New hybrid, socio-computational systems can harness the *collective intelligence* (or *wisdom)* of the crowd in concert with automation to better tackle large and/or difficult processing tasks. While the IR community already has a rich understanding of systems and user-centered design issues, crowd-based computing represents a significant departure from existing knowledge and practice. How can we innovate design, implementation, and evaluation IR systems in order to effectively incentivize, incorporate, and benefit from crowd participation?

Proposal. New IR research is needed to determine how to most effectively blend algorithms and crowd participation to quickly, accurately, and affordably: (i) collect IR-specific data, such as topical relevance judgments or search logs; (ii) conduct human subjects research, such as interactive IR studies; and (iii) deliver enhanced search solutions which blend traditional search algorithms with social search via community portals, social networks, and/or dynamically-formed online collaborations. In comparison to current approaches, we anticipate crowd-based solutions will yield superior data-collection, research infrastructure, and search systems that are faster, more accurate (or diverse), and/or cheaper.

Challenges. Effectively integrating human computation in IR systems represents a broad interdisciplinary challenge, spanning information and computer science, psychology, sociology, and economics. How can we facilitate productive cross-disciplinary knowledge dissemination and collaborations? While human computation for simple tasks like image tagging is well-established, how can we effectively accomplish more complex or specialized tasks with crowds (where anonymity may limit our knowledge of participants' relevant skills and experience)? How do we effectively recruit, train, and retain participants for effective, sustainable hybrid computing models? How can we best scale up crowd participation? How can we best address quality assurance issues in data collection, differentiating objective errors vs. legitimate data diversity? As online human computation creates new potential for worker exploitation and digital sweat shops, we must also develop conscientious practices which mitigate legal and ethical risks.

Related efforts. Crowdsourcing workshops were organized at SIGIR 2010 and 2011, WSDM 2011, and WWW 2012. A crowdsourcing NIST TREC track will run for the 2nd year in 2012.

5.7 Getting Your Life Back: Capturing, Searching, and Synthesizing Personal Data

This project will design tools for capturing and indexing the online data streams encountered by users, including data currently stored in external silos; for searching this data; and for making sense of it.

Motivation. An increasing proportion of peoples' lives is lived online, sending and receiving data through personal digital computers. This data is, in principle, readily capturable, but is not currently captured in any unified, user-accessible way. Capturing these data streams offers users an aid and extension to their own memory, providing opportunities for re-finding, reflection, interpretation, and self-presentation. The importance of such tools is emphasized by the documented tendency for our online life to dissipate our attention, keeping us in shallow working memory, and distracting us from the deep processing necessary for understanding and retention of information.

Proposal. We will develop tools that users can run on their own devices that will capture and index the data they receive and send. Initially, this may be as a browser plugin, recording only textual data. Further tools will be provided for searching and exploring this data. We believe that the novelty of these tools, and the opportunity for self-exploration they offer, will drive widespread adoption. Immediate research-engineering questions will be posed about how to collect, index, and selectively preserve this information. Observational and interview-based studies will be performed on how users employ the tools we provide. Retrieval methods will be developed across this heterogeneous but personal data, and user-based experiments performed. Users will be invited to perform data deposits of redacted personal data for collaborative experiments.

Challenges. There is a growing body of work exploring how best to use and evaluate personal data, but when we consider complete "life logs" the problems are only more complicated. A first challenge, of course, is attracting users to our tools and encouraging them to adopt them. Protocols must be extended for performing experiments on unshared personal data. For data collection and aggregation, anonymization and semi-automated redaction will be critical.

Related efforts. There has been recent work on desktop search and personal information management; this project goes beyond this to include all personally experienced digital information. Collaboration with researchers in HCI, and in cognitive and neurological theories of memory and recall, will be sought.

5.8 *Information* Retrieval

This project will provide effective algorithms for providing relevant *information* to the user, as opposed to relevant *documents*. Systems will be evaluated based on “information units.”

Motivation. Information needs are often not about a ranked list of *documents* or snippets. The user wants *information*. The information may be scattered across documents. It may be explicit or implicit. Providing just the right information to the user in the right format is especially important in a mobile environment.

Proposal. This project will first provide a clear guideline for systematically defining information units. The guideline will not be about blindly tearing a given document into pieces – it will define information units based on specific information needs. The information units will serve both as the basis for building models for *information* retrieval, and as the basic unit of evaluation. The *information* retrieval models and the evaluation framework should aim to handle redundancy/novelty, diversity, dependency between information units, and even information discovery through inference.

Challenges. The evaluation based on information units is challenging. Relevant information units would differ depending on users (e.g. background knowledge and interests) and contexts (i.e. anything besides the query such as physical surroundings that may help specify the user’s information need). Also, relevant information changes over time. Handling dependency across information units is difficult both for the retrieval model and for the evaluation. Moreover, if a system is able to infer and synthesize information from multiple sources, how can it convince the user?

Related efforts. The NTCIR-10 One Click Access task evaluates system output based on information units. The TREC 2012 Contextual Suggestion track asks systems to respond to complex information needs with “suggestions” rather than documents. This topic is also strongly related to the Conversational Answer Retrieval (Section 4.1) and Structure Dimension (Section 4.5) topics.

5.9 IR4ALL: Addressing Digital, Physical, Cognitive and Cultural Divides to Search

This project is about offering specific search support for a diversity of physical, cognitive, and cultural user profiles that require specific search assistance. In other words, the goal is breaking information access barriers and making search facilities available to everyone.

Motivation. Most IR research to date has implicitly assumed an archetype of user that has no physical, cognitive, economic or contextual (age, culture, native language) handicaps to engage in an information-seeking task. But that leaves a substantial proportion of potential users out of the loop: for instance, the billions of illiterate people across the globe or the 10-15% of the US population with dyslexia. Making search facilities available to everyone has, then, an obvious and measurable impact.

Proposal. This project topic focus on providing and evaluating dedicated search support facilities for a diversity of physical, cognitive and contextual user profiles that require specific search assistance, including: (i) user’s education level and background knowledge: in particular, illiterate users and children (IR for education); (ii) people with impaired physical abilities –vision, hand movement, etc.; (iii) people with cognitive disabilities (such as aphasia or dyslexia); (iv) people without access to a computer –but with access to a phone; (v) people with cultural and linguistic backgrounds that do not match those of the target information (such as an immigrant trying to find his way in the regulations of a country in a foreign language and a foreign culture).

Challenges. The basic challenges associated with IR4ALL are related to interaction design broadly:

- *Resource selection, adaptation and presentation*. Besides relevance and authority, an essential factor is adequacy of the source to the user: a document suitable for an expert or an adult is possibly not suited for a child (particularly if engaged in a learning process). If the relevant information is only present in a document which is not suited for the target set of users, it must be enriched, or simplified, or in general transformed. A document for children, for instance, might need to be simplified in terms of vocabulary and enriched with links to basic explanations of the concepts in the document. A legal document for an immigrant might need to be simplified, summarized, translated, and linked to relevant external regulations.

- *Assistance in query formulation and refinement*. There are a wide range of challenges here depending on the type of users, many of them hard to tackle. In the immigrant example, consider translating a non-expert spoken query in Chinese into the vocabulary of the Spanish legal system and back; this requires sophisticated term suggestion technologies (together with educational style concept explanations) that are beyond current IR capabilities.

5.10 Information Retrieval for the Ecosystem of Apps

There is a growing trend towards the use of apps ranging from entertainment to government and financial services. Some describe this new dynamic world of apps via an ecosystems metaphor. What constitutes IR in such an ecosystem?

Motivation. Propelled by smart devices and service marketplaces (e.g., apps stores) and the growing number of services that are being more generally exposed (e.g., whole-of-government service frameworks), there is a growing number of process objects to be retrieved, rather than classic documents.

Proposal. We suggest a project that aims to explore indexing and retrieval of services or apps. Complex needs would involve the retrieval of a number of relevant apps that would need to be appropriately choreographed to satisfy the need. Consider a complex goal like opening up a coffee shop. Pertinent services span a spectrum of issues such as occupational health and safety, tax, employment etc. IR in this environment is exploratory. A target “app IR system” should help the user navigate the ecosystem by providing suggestions for apps which they may not be able to formulate themselves.

Such a project will require investigation of how to classify the functional aspects of services or apps, to explore whether and how service needs are related to information needs, how to model the “ecosystem” of apps, and how to evaluate all of this work.

Challenges. Exploratory search for complex service agendas is interactive and so brings with it all of the complications of interactive user evaluation. Ideally such a system would support users as they learn about apps and how to search them, raising new problems of adaptation and user context.

5.11 Information Seeking Stage Aware Search Tools

A search session for a non-trivial search task consists of stages with different sub-goals (e.g., problem identification) and specific search tactics (e.g., reading introductory texts, familiarizing with terminology). Can we make a retrieval system aware of the searcher’s stage in the information seeking process, tailor the results to each stage, and guide the searcher through the overall process?

Motivation. Making a system aware of a searcher’s information seeking stage has the potential to significantly improve the search experience. Searchers are stimulated to actively engage with the material, to get a grasp on the information need and articulate effective queries, to critically evaluate retrieved results, and to construct a comprehensive answer. This may be of particularly great help for those searchers having poor information or media literacy. This is of obvious importance in many situations: e.g., education, medical information, and search for topics “that matter.” Some special domains, such as patent search and evidence based practices in medicine, have clearly prescribed a particular information seeking process in great detail. Here building a systems to support (and enforce) this process is of obvious value.

Proposal. Likely a coarse-grained model (problem identification stage, information collection stage, search closure stage) is sufficient. There are obvious resources that are particularly suitable in a particular stage (e.g., Wikipedia as introductory), also demanding different evaluation measures (precision versus recall). Related genre/reading level classification methods have been proposed. Use cases can inform specific progression models. System design may vary from lightweight (unobtrusive grouping/labeling of search results on a search result page) to systems that guide a searcher through the relevant stages.

Challenges. Information seeking processes have attracted much attention over the last decades in information science, resulting in a range of theories and models that identify different stages in the process. Although there is family resemblance between the different models, there is no universally agreed upon form.

5.12 Protecting Users' Privacy in Search

The research topic develops effective and efficient solutions in protecting users' privacy in search applications.

Motivation. Imagine that we could create a machine into which many people would type their hopes, dreams, aspirations, and interests. How would we protect this intensely personal information? We have indeed built such a machine; we call it a "search engine." And we have seen how remarkably easy it is to associate search behavior with a specific individual, from news accounts of users being identified in AOL query log data and in Netflix logs. Attribution of specific queries to specific users can have serious consequences, for example in personal health and national security applications.

Two broad lines of research have been pursued to protect users' privacy in search applications. One approach is to allow knowledge of the query but to obscure the identity of the searcher. For example, it is possible to forward each user's query through a randomly generated path of cooperative network nodes before reaching a search engine. The alternative approach is to obscure the content of the query. Examples of this include (1) "private information retrieval", which encodes users' information requests using security protocols of high computational complexity, and (2) "plausibility deniable search", which substitutes each actual user query with a set of standard queries that together cover the actual information need and some other topics. Some other research builds distributed and private user log for helping search engines to improve search results without disclosing sensitive user information.

Challenges. We need to better understand the ecosystem of users, and in particular how values involving privacy, effectiveness and cost are balanced in different applications. We need to develop systems that can help users to understand and to control those tradeoffs in different ways when working on different tasks. We need ways of combining strategies that protect users' privacy and confidentiality of searchable contents in a unified setting.

5.13 Search among Secrets

Massive collections of informal interactions (e.g., email or recorded meetings) are now accumulating in personal collections, business information systems, and government archives. Leveraging this new cornucopia requires not only that we find what should be found, but that we not find what should not be found!

Motivation. Segregating sensitive documents was a 20th-century concept in business and government that relied on skilled administrative staff; today's information systems often contain an undifferentiated mishmash of sharable and sensitive user-generated content. The result is a conundrum: we can easily retain this intermixed content for its potential value, but we will fail to leverage that value because we lack the technology to find the sharable content among the comingled thicket of sensitive information.

Proposal. Requests for archived information closely follow a power law distribution, with most content never being requested. Thus, tremendous efficiencies can be achieved by rapidly segregating sensitive content only in response to specific access requests. We can view this as first finding relevant content, and second finding which among those relevant items are sensitive. For example, relevance search might be fully automated, while sensitive content removal might be performed by experienced professionals using advanced tools that help them to make rapid and accurate decisions.

Challenges. Holders of sensitive information are naturally risk averse, so broad adoption will require very accurate, and demonstrably very accurate, detection of sensitive information. New evaluation designs that reflect the nature of sensitive content detection (automated or human-in-the-loop) will be needed. New work on information seeking behavior is also called for, since search strategies developed in an era of instant response may not work well if on-demand review imposes latency constraints. Potential attacks on system integrity (e.g., collusion among searchers) will need to be modeled and mitigated. Since costs (e.g., for review) and benefits (of access to otherwise unavailable content) may accrue to different parties, economic models that result in rational resource allocation will also be needed.

5.14 Simulation of Interaction

The simulation of interaction provides a powerful and flexible way to study the interaction between a user and a retrieval system towards the construction of an evaluation framework that can potentially bridge lab-based and user-based evaluation. The challenge is to develop the appropriate abstractions given the complexities of the environment to be simulated.

Motivation. Search in the real world is inherently interactive. Yet much of the research conducted in IR follows the Cranfield paradigm, where interaction is typically restricted to a single search. This lab-based approach suffers from being static in nature, but it provides cheap, controlled, and re-usable evaluation resources. In contrast, studies directly involving users show the rich sequence of interactions between users and IR systems; however, they are much more costly, and less generalizable because they are conducted in particular contexts under particular conditions. Simulation of interaction offers a way to improve lab-based research and open it up to the study of interaction, and also offers user-based research a way to transfer and encode their understanding of users within simulated users.

The major outcome of this project will be a portable test-bed that can dynamically solicit, and react to, system actions. This will give the ability to researchers and practitioners to build systems that optimize for multi-step, multi-session, multi-task, multi-mission information-seeking episodes.

Proposal. The proposed directions in this project are: (a) identifying and modeling the interactions, behaviors, and factors that affect the search process; (b) formalizing evaluation methodologies for the different kinds of simulation of interaction (what-if studies); (c) developing an Evaluation-through-Simulation Framework, which paves the conceptual and logical path for undertaking and integrating user-based, systems-based and simulation-based research; (d) determining the limits of simulation and how applicable it is in practice and theory for quantifying, generalizing, repeating, reducing bias, and (e) implementing a simulation toolkit.

Challenges. Interaction is a complex process and its simulation is even more complicated. Some degree of understanding and recognition of the limits of simulation will be needed for its results to be accepted.

5.15 Spoken Information Retrieval

This project aims to extend IR systems such that they can accept spoken queries and generate spoken results, thereby helping the visually impaired as well as anyone in a setting where speech is a more natural interface.

Motivation. An estimated 285 million people are visually impaired. At present, these people must search using information retrieval systems built for sighted users, brought into their world using so-called “screen readers.” We can, and we do, today build systems that accept spoken input and that generate spoken output, but the potential exists to build systems that manipulate speech with a facility similar to the way in which present search engines manipulate written text. Fully actualizing this potential can also benefit sighted users in hands-free/eyes-free situations such as driving, low-literacy users who have access only through cell phones, and speakers of some minority languages for which spoken content may be far more easily collected than written content (e.g., Australian Aboriginal languages)

Proposal. This research aims to build spoken IR conversational models, for deployment on a mobile platform. The input is speech; the output is audio. In addition to help everyone in settings where speech is a preferred interaction mode, this is a “public good” project to facilitate information access to a large section of the population that is currently disadvantaged by visual impairment, low literacy, or a primarily spoken language.

Challenges. There are numerous scientific and engineering challenges: HCI models for spoken IR; extracting short meaningful snippet suitable for spoken interaction; user interaction with spoken IR interfaces; focused IR – delivering the best entry point to documents (BEP); text/audio summarization; text-to-speech; speech storage, indexing, search, and delivery; building the software with restrictive and with innovative mobile devices; mobile technology challenges (noisy environment, bandwidth, power, cost); incorporating into the spoken IR models the global positioning and derived contexts that are accessible on mobile devices; building collections of spoken IR for end users (e.g. consider text-to-speech of a Wikipedia document: what should a spoken Wikipedia document look like for delivery on a mobile device?); as well as building test collections and evaluation platforms. User interaction over a sequential and slow speech communication channel presents new challenges.

5.16 Super Models of Information Retrieval Interaction

This project aims to develop powerful and rich formal models of IR that incorporate users, systems, and the interaction between them.

Motivation. Information Retrieval has had a long history in the development of formal models. While such models have been developed mainly for ranking information objects, significantly less attention has been paid to developing models (and theories) for interactive Information Retrieval. In recent keynotes at the European Conference on Information Retrieval, Belkin (2008) and Jarvelin (2011) have stressed the growing need for and grand challenge aspect of formal models and theories that include and cater to the user and the interaction between users and systems. This is because having theories and models of IR is intrinsically important to IR as a science. Furthermore, such models would provide the basis to understand, explain and predict what is likely to happen during the course of interaction and how it is likely to happen. Also, such models and theories would enable the generation of testable hypotheses that help to guide and direct our science. We hope that super models of Information Retrieval Interaction would lead to scientific laws, which enable the relationships between observables to be characterized and better understood.

Proposal. We propose a program driven by these questions: (i) How can we develop models and theories that adequately capture retrieval, search and information-interactions: going beyond just ranking? (ii) What steps should we take to nurture and cultivate theories and models within IR, and how can we progress towards theoretically driven research? (iii) Will it be possible to develop an over-arching framework in which emerging theories can co-exist or be combined into a grand unifying theory? (iv) What kinds of empirical research, evidence and validation of super models of Information Retrieval Interaction will be required to provide adequate support for such theories? (v) And will such super models and grand theories manifest useful and intuitive laws that govern Information Retrieval, and enable the prediction of information interactions between users and systems?

Challenges. Being able to answer these difficult research questions presents many pragmatic challenges such as: (i) theory and model development are very time consuming, require a lot of effort to construct, with an unknown pay-off (i.e. may be very successful, but likely to be rather unsuccessful). (ii) finding the right level of abstraction and catering for a large number of variables may lead to overly simple models which lack predictive power (iii) there are few forums for theory and models to be presented or the routes to publications are not very clear, and (iv) the nature of the models may be too complex to represent, or rather, to illustrate the complexity will require a serious and dedicated research program.

5.17 Supporting Complex Search Tasks

This project aims to build an information access tool that actively supports a searcher to articulate a whole search task, and to interactively explore the results of every stage of the process.

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

Proposal. A complex task could be formulated as a complex query or search strategy expressed in a structured query language. This query could be interactively constructed based on an initial plan and further feedback, combining several constraints on content as well as on structure (collection structure and annotations). Different task constraints offer also different points of entry into this process. We can exploit implicit and explicit contextual information on the specific search request, the specific location/time/IP, the searcher and her preferences, etc. Mobile devices are the natural setting for rich contextual information, with a preference for easy interaction rather than entering text. Often, the process will be iterative, with the task leading to another (sub)task.

Challenges. Search with task and person context requires a novel mixture of search and recommendation methods, novel retrieval models, and evaluation methods (beyond topical relevance). Structured querying and semantic annotation become crucial, but can be hidden from the searcher. Interactive search requires user-centered evaluation or grounded simulations. Storing personal profiles can invade privacy and hence this type of data requires special care (e.g., TREC 2012 Contextual Suggestion Track).

5.18 Time Changes Everything: Dynamics and Time in Search

Time is a crucial parameter for understanding human-generated information and information needs. Yet few information access systems represent information dynamically or leverage time for their services.

Motivation. An information system is affected by time in many ways. The information it processes changes continuously both in content and form, the world that information references evolves, and information needs and usage scenarios change and evolve. In a big data context, modeling the character, content and evolution of a steadily changing immense information stream requires a perspective of information as something dynamic over time, not as something constant to be extracted. Awareness of time in information access scenarios can be of many types: real-time issues ("I need to know what is happening right now!"), semantic drift ("Are there new ways of referring to my topic of interest?"), temporally specific data ("Give me information about what happened at this specific date"), and temporal references ("What documents mention this specific point in time?"). *Temporal awareness* requires systems to process time references on multiple scales in both content and queries. *Temporal dynamics*—the volume and content churn in the data stream under analysis—necessitates a processing model and a representation that can handle changing meaning of items in the index, changing associations between index keys and the items being indexed, and changing information needs. Systems need to be aware of time and leverage that awareness to enable entirely new services, not only to serve documents in response to a query but to model and present the state of the world as it changes.

Proposal. Search and retrieval must have time as a core dimension. In a dynamic environment, every single ingredient of the retrieval pipeline needs to be on-line modifiable, without lengthy test-train-deploy-update cycles. To make this happen systematically we need a framework to talk about time and to test system performance vis-à-vis time. There are excellent first steps taken along these paths and several systems for time-aware information retrieval are in the market but there is no collected research effort to talk about them.

Challenges. The traditions of the IR field are not well equipped to deal with experimentation and evaluation of dynamic data. Making the shift from set and collection based testing to evaluating situation awareness will be a major challenge.

5.19 Understanding and Evaluating Rich Aggregated Answers

This project aims to move research beyond evaluating ranked lists of documents, in order to understand the usefulness of richer answers. Rich aggregated answers are search results that might include multi-granular information objects: combinations of paragraphs, summaries, tables, graphics, images, and so on.

Motivation. Search systems are becoming more complicated and are presenting richer results (for example, combinations of documents, images, and videos). Simple summaries (title, snippet, and URL) are no longer sufficient for emerging application areas. While the academic IR community has a good understanding of rigorous evaluation approaches, these need to be expanded to encompass new answer modes. The risk of not doing this is that IR research will remain focused on a particular search evaluation approach that does not reflect important new directions that search is taking. Out-dated evaluation models may lead to incorrect conclusions about which approaches are better. Comparing the usefulness and impact of new rich complex answer systems is particularly important in non-traditional domains, for example satisfying deep information needs in education, providing screen space-efficient answers for mobile search.

Proposal. We seek to study the way in which rich answers support different information tasks and to develop a taxonomy of information needs. Subsequently, this project could model how systems can exploit this taxonomy (for example, integrating structured and unstructured data) and develop appropriate task-oriented evaluation paradigms.

Challenges. How to conduct meaningful, rigorous, and repeatable experiments in this environment is a challenge. It is not clear that test collections with traditional relevance judgments are suitable for evaluating rich aggregated answers. To be useful, any new evaluation approach will need to gain widespread acceptance. The ultimate outcome will be to build a practical test-bed that supports repeatable experiments.

Related efforts. This research cuts across a wide range of new search approaches and tools that generate answers beyond simple ranked lists of text snippets, including mobile search, integrating structured and unstructured data, and dialogue systems. It is also related to the broader issue of understanding users.

5.20 Understanding Opinion Engineering

This project aims to study mass online opinion-influencing mechanisms such as "viral advertisement campaigns" and "computational astroturfing" (aka fake grass-roots campaigns).

Motivation. Many powerful entities in society such as large corporations and political interest groups have a vested interest in influencing public opinion. The rise of online social media, which facilitates interactions between large groups of relative strangers, has resulted in an unprecedented opportunity for these entities to influence and possibly misinform opinion. By paying individuals to comment in forums and social networks, these entities can influence the opinions of large percentages of the population. Furthermore automated “bots” can be used for similar purposes, creating “fake users” who strategically place comments in such a way as to maximize influence over the desired population. As this technology matures, campaigns will likely become more focused on certain target demographics, and given the importance of influencing individuals’ opinions, this progression toward targeted opinion engineering may result in a social engineering arms race.

Proposal. The project will tackle important questions such as whether it is possible to detect opinion engineering automatically, and if so, what signals are the best indicators of it and how reliable can such detection systems be? Having detected influence, another important question is whether it is possible to measure the amount of influence that each particular entity is having. Then on an individual level, a further question is whether it is possible to identify those users who are “fake” or have been "compromised"? Having identified fake users, what would the ethical ramifications (if any) be of attempting to discredit such individuals online?

Challenges. Building an opinion engineering test collection will be necessary to compare automated systems.

5.21 Understanding Search in the Workplace

The project will attempt to characterize the search problems typical of modern workplaces, and to develop models which can underpin realistic evaluations and lead to more effective search tools.

Motivation: Traditional IR tools have not been designed or optimized for use in the diverse information environments found in knowledge-rich workplaces. Workers performing their duties may now need to search across personal, departmental, corporate, outsourced, and public information sources and content types ranging from database records, office documents, email, and webpages, to microblog posts, instant messaging, and continuous media clips. This heterogeneity of data and the wide range of search-mediated tasks are not well modeled in current IR test collections.

Proposal: This project involves an initial phase of ethnographic study of a representative range of workplaces, leading to the development of test collections which more faithfully model real world search scenarios. The availability of test collections is expected to promote community research into retrieval systems which better serve workers and promote the productivity and competitiveness of their organizations.

Challenges: The ethnographic study needs to be extensive and well-planned to ensure that its findings are indeed representative of the full range of workplace search problems. This may best be achieved through a large-scale collaboration between research groups. The non-public nature of the majority of documents in most workplaces and of the queries and answers poses a huge challenge for creation of public data sets. However, if the ethnographic study is able to precisely characterize the data and the information needs, it will hopefully be possible accurately simulate the interesting problems using non-sensitive data.

6 Conclusion

It is 62 years since Calvin N. Mooers first defined the term “information retrieval” in the academic literature. In each of the six decades that followed, IR researchers have produced substantially improved search systems. As demonstrated by the volume of ideas proposed during this workshop, here at the start of the seventh decade of IR research, the field of information retrieval continues to be both a strong and vibrant research area. The themes in the proposals show a worldwide research community identifying topics of future work that go far beyond the abilities of existing commercial search providers. The opportunities to extend the abilities of IR systems are wide ranging and diverse.

While heterogeneous in their views of where future trends in IR lie, delegates were consistent in their view that the SWIRL 2012 workshop was a stimulating opportunity to discuss and debate new ideas as well as to interact with other leading researchers.

The organizers and participants would like to thank and acknowledge the following sponsors who provided partial support for arrangements and travel: the University of Melbourne, the ELIAS network (funded by the European Science Foundation), the National Science Foundation by grant number 1216764, and RMIT University.