Information Fostering – Being Proactive with Information Seeking and Retrieval

Chirag Shah
School of Communication and Information (SC&I)
Rutgers University
4 Huntington Street, New Brunswick, NJ 08901, USA
chirags@rutgers.edu

ABSTRACT

People often have difficulty in expressing their information needs. Many times this results from a lack of clarity about the task at hand, or the way an information or search system works. In addition, people may not know what they do not know. The former is addressed by search systems by providing recommendations, whereas there are no good solutions for the latter problem. Even when a search system makes recommendations, they are limited to suggesting objects such as queries and documents only. They do not consider providing suggestions for strategies, people, or processes. This Perspective Paper addresses it by showing how to investigate the nature of the work a person is doing, predicting the potential problems they may encounter, and providing help to overcome those problems. This help could be an object such as a document or a query, a strategy, or a person. This whole process is referred to as Information Fostering. Beyond crafting a general-purpose recommender system, Information Fostering is the idea of providing proactive suggestions and help to information seekers. This could allow them avoid potential problems and capture promising opportunities from a search process before it is too late. The current paper presents this new perspective by outlining desired characteristics of an Information Fostering system, envisioning application scenarios, and proposing a set of potential methods for moving forward. Beyond these details, the primary purpose of this paper is to offer a new viewpoint that looks at the other side of the information seeking coin, by bringing together ideas from human-computer interaction, information retrieval, recommender systems, and education.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Recommender systems**; **Task models**;

KEYWORDS

Information Fostering; Proactive IR; Intelligent Assistants; Task-based Information Seeking

CHIIR'18, March 11–15, 2018, New Brunswick, NJ, USA

ACM ISBN 978-1-4503-4925-3/18/03…$15.00
https://doi.org/10.1145/3176349.3176389

1 INTRODUCTION

Two of the fundamental problems for information seekers are: not being able to express their needs due to lack of understanding of the task/topic at hand, or the way a resource/system being used works; and not knowing what they do not know [3, 10]. The former is addressed by information retrieval (IR) systems by providing recommendations, whereas there are no good solutions for the latter problem. Even the recommendations made by IR systems are often limited to suggesting information objects only, and do not explore the possibilities of recommending a process/strategy, people, or other forms of suggestions.

In this Perspectives Paper, a novel idea is presented, which looks at the other side of the information seeking coin to address the issues mentioned above. This idea is termed as *Information Fostering*. In a nutshell, it refers to proactively identifying problems and opportunities to not only help information seekers in a more comprehensive way than typically offered by today's IR systems, but also recognize possible questions, answers, barriers, and help that a person may not be even aware of.

The idea of proactively helping in an information seeking situation is not new. It is often studied under the term *Proactive IR* (e.g., [16]). Search systems often pre-fetch information that is likely to be relevant during a search session. But there is one fundamental difference between proactively retrieving information that an information seeker may look for and the concept of *Information Fostering*. The essence of Information Fostering is in trying to understand the larger context of the task in which information seeking may or may not happen, and provide recommendations and help that are more than just queries or documents. This is inspired by the education domain, where learning involves not only getting to right answers, but also the processes that take one to those answers.

The rest of the paper is organized as follows. The idea of Information Fostering is further elaborated in the following section with examples, challenges, and possible ways to address those challenges. Section 3 presents some of the related and relevant works with respect to these challenges. This also sets a context in which the perspective of Information Fostering is described here. Section 4 provides some of the recent works done by various scholars, including the author, trying to address some of the challenges. These works also serve as preliminary investigations into different aspects of Information Fostering for what may come next. To guide the scholars who may want to take on this challenge of Information Fostering, Section 5 provides various methods and ideas for conducting new research in this area. Finally, the paper concludes in Section 6, summarizing this new perspective.

2 INFORMATION FOSTERING SCENARIOS AND EXAMPLES

This section will provide further explanation of what Information Fostering is, could be, and what it would look like implemented in real systems. It starts by showing some of the existing systems that try to cover at least some of the aspects of Information Fostering, and then provides scenarios for ideal systems with the support of Information Fostering.

2.1 Clarifying Information Fostering

Most models for information seeking start with a premise that a person lacks some kind of information [14], or has a need [44] that can be met by obtaining relevant information. Many IR systems are built to address such needs. These systems often even have various mechanisms to help that information seeker when there are problems in obtaining information.

Those interested in the user side have studied this under the names of information seeking/behavior, whereas those interested in the system side have contributed to IR aspects. Then there are scholars who study various interaction elements (information science, HCI) and recommendation aspects (data mining, machine learning). Of course, these are quite broad strokes to paint a picture that depicts human information interaction and retrieval. While definitions and conceptual framings of these terms and areas could be argued, it is clear that

- These studies of information seeking start with the assumption that a person has a need for information;
- When a problem appears in that information seeking process, an intervention/recommendation is presented; and
- These interventions or recommendations are often provided as objects of information, such as documents and queries.

The idea of Information Fostering relaxes all of these assumptions and constraints. A good metaphor is that of a teacher. A good teacher does not simply answer a question by a student; she also shows him how to ask a good question. She does not simply provide an answer; but illuminates a path that the student can follow to reach to that answer. The teacher does not wait for a struggling student to fail; but rather helps him early in the process to ensure that failure does not happen. The *Fostering* part in Information Fostering really emphasizes such characteristics.

2.2 Existing and desired support

While it may be tempting to say that there is no current system or support for Information Fostering, there are indeed several instances one could find in various information systems, albeit at a preliminary level and with shortcomings.

In the late 90s, Microsoft introduced the office assistant feature to its Microsoft Office suite, with the default and the most popular character being *Clippit* (also known as *Clippy*). This feature, through an animated character on screen, monitored a user's activity in an application such as a word processor, and offered proactive help. The reception of this feature was mixed and Microsoft decided to remove it a few years later.

Since the early years of this century, RSS (Really Simple Syndication) feeds have found their way in many applications – desktop-based and mobile – to aggregate and proactively provide content to a user based on his/her interests, preferences, and actions. Variations of RSS have continued their popularity as feeds and notifications of different kinds that push information to people without them actively trying to retrieve it. This is widely used in most social media services such as Facebook and Twitter, as well as news aggregators such as Ozmosys.

More recently, intelligent assistants such as Siri, Google Now, Cortana, and Alexa have provided smarter approaches to offering proactive suggestions to a user based on the given context (e.g., time, location), past needs and assessments, as well as personal preferences. The nature and the amount of proactive support vary greatly, with some of these systems tightly integrated into a person's overall digital life including smarthome and cars, whereas others are glorified IR systems with voice recognition.

In all of these systems or scenarios, the common element is being proactive. While this often helps address the problem of people not knowing what they do not know, there are several drawbacks and shortcomings of such systems. As is evident with *Clippy* example, a proactive suggestion could be perceived more of an annoyance than a help, turning people away from using it even at times when it could be quite useful. In the case of information aggregators or feeds, the recommendations are driven by a person's interests and past assessments, but not necessarily based on the current context or task. In other words, they are less dynamic and less tailored to micro-moments. Finally, even the newer systems, such as Siri, that offer proactive suggestions are limited to the kind of recommendations they make. Siri never suggests that you consult a friend or a colleague as a proactive recommendation.

These realizations lead the author to making a list of an ideal Information Fostering system as enumerated below.

(1) The system should find a good balance between being proactive and invading one's privacy. In other words, offer proactive suggestions in a way that does not hinder one's ongoing task or even take their attention away from it.
(2) The system should consider not only the broad strokes of one's past behaviors and preferences, but also take into account the ongoing activities, assessments, and performance.
(3) The system should make a wide range of recommendations, including information objects (e.g., documents, queries), processes and strategies, as well as people.

In the following subsection, an attempt is made to provide an outlook for such a system in different settings.

2.3 Future scenarios and mockups

This subsection provides a couple of scenarios and mockups to clarify what Information Fostering implementation could look like in typical information seeking situations.

The first one (Figure 1) describes a typical search episode, where an information seeker is looking for information using an IR system and encounters a problem. This is where the system can offer a suggestion. For instance, if the searcher mistypes a query, the system could offer a correction. If the searcher is lost for what would be a good way to complete a query, the system could offer suggestions based on what may be most relevant or how others completed such a query. Assuming the searcher takes such a suggestion, his/her search episode will now be on a new path going forward. An Information Fostering scenario, on the other hand, plays out differently. Here, the system is actively evaluating the current search path that the searcher has taken and tries to predict where it might lead. If it finds a potential problem down the road, rather than waiting for that moment to come, it proactively warns the searcher about this and offers a detour that avoids the problem. This is similar to a GPS navigation unit that is aware of the traffic conditions on the road ahead and offers an alternative route if it finds there to be a problem in the charted path, before the traveler reaches close to that point of congestion.

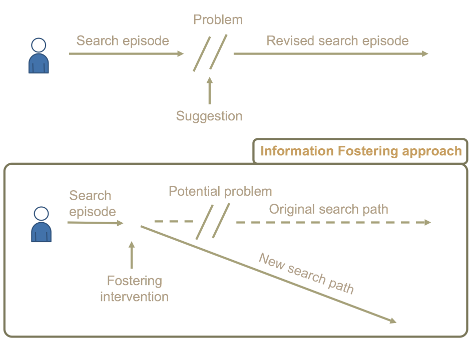

Figure 1: Information Fostering scenario involving an individual information seeker who encounters a problem during a search episode.

Figure 2 provides another scenario. Here, imagine two information seekers working on the same/similar task in the same timeframe. Normally they would go through their search episodes independently, even if they could benefit by working together. There may not be any problems with their individual search episodes, but they could achieve much more if they work with each other. This is an opportunity that is lost, unless someone or something helps make this connection. An Information Fostering system will do just that. It will recognize that there is a potential opportunity here in these two people collaborating, and points it out to them. Assuming they agree to work together, their new search episode is the one that involves collaboration. For an example – the Information Fostering system could recognize that two people working on the same task have different skills or abilities and suggest for them to work together. This asymmetric relationship could be teacher-student, expert-novice, or structured around a specific skillset.

Let us now examine the idea of an Information Fostering system through a real-life scenario. Alice is looking for ideas and things for her grandma's 90th birthday. She starts by searching online – first at a general-purpose Web search engine, and then at an e-commerce site. However, she does not have very clear notion of what this gift may look like, how much she would want to spend, and how she could make this something very special (and not just an off-the-shelf object). An Information Fostering system integrated in her browser quickly realizes this and starts offering suggestions, even as Alice keeps looking around. The system, based on Alice's

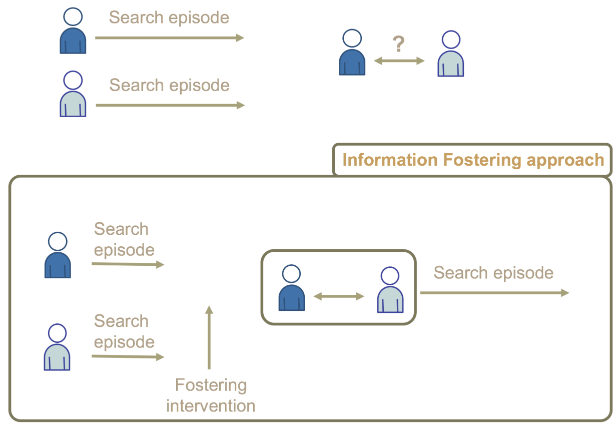

Figure 2: Information Fostering scenario involving two individuals working on the same/similar search task during the same time.

own history and that of the world, offers a strategy to think about finding something based on some significant events that happened in the world during her grandmother's lifetime. That triggers Alice to think about a post-WWII symbolism and how her grandma had talked about it. But Alice does not know enough about that era. Realizing this, the system then offers to connect her to someone who had either lived during that time or has studied it. Alice takes that suggestion and lets the system connect her to a WWII historian. He offers a couple of suggestions. Alice likes one of them – a brooch with an angel symbol to mark the time of peace. Now Alice can focus her searching and look for this particular item. She finally finds it at an antique storefront in an e-commerce site.

Here are the highlights from this scenario: (1) the Information Fostering system is being proactive in offering the suggestions without Alice specifically asking for them; (2) the system is able to use very little data from Alice's ongoing activities, thus addressing the cold start problem; and (3) the recommendations offered include people and strategies, in addition to objects.

A mock-up of this browser-based system is shown in Figure 3 and its various components are described below.

- *Queries*: this component lists potential queries a user may run at a given moment. The queries are found by not simply looking at relatedness as it is done in most query recommendation systems, but also by using task and topic knowledge.
- *Documents*: this component presents a few documents (Webpages) that the user may find useful. They are based on not just what others found to be useful for the same task, but also based on where a user is in the given search process. For instance, if the user is in the beginning stage of a search episode, it may be more appropriate to present documents with broad information. If the task is considered to be complex or obscure, this component will show documents deemed to be explanatory and easier to read (based on automatically computed readability scores and information content).
- *Strategy*: this component provides a combination of queries, documents, and potentially relevant segments. The system

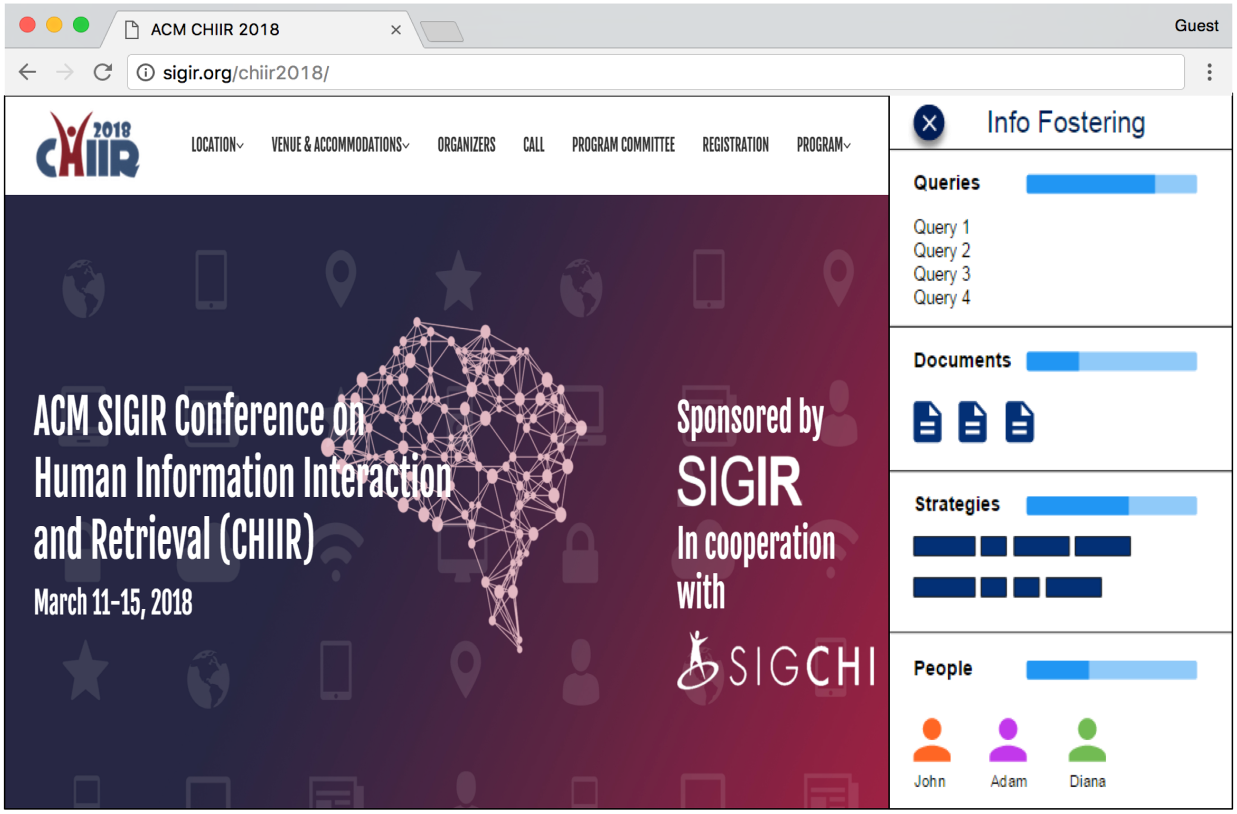

Figure 3: Mockup of an Information Fostering system, embedded in a Web browser.

infers the nature of the task as well as at which stage of the task the user is to guide the strategy recommendations.

- *People*: this component features potential people to collaborate at various stages of the search.

With each of the components, a confidence score is calculated. This score, ranging from 0% to 100%, is one of the outcomes for various techniques/algorithms that make up the Information Fostering system. These scores could be used by the user to decide if he/she wants to take any of the recommendations. Note that the design of this system uses the idea of peripheral awareness [17], which allows one to keep the attention on the ongoing task without being distracted while still having easy access to useful information on the periphery. One could also easily close that peripheral window if he/she wishes not to have such support at all.

Building such an Information Fostering system will require solving a number of problems. There are already several works, including those of the author, that address some of these problems. But all of these have been opportunistic at best. What it means is that these works identified a specific opportunity in information seeking situation where prediction of potential problems could lead to proactive recommendations, but it is not clear if or how any of these approaches could be generalized. To do so, we must approach the problem of Information Fostering top-down, focusing on systematically characterizing its components and thinking through implications in various scenarios. But first, it is important to understand what is already been done.

3 RELEVANT RESEARCH

3.1 Task characterization and classification

Individuals engage in information seeking activities to accomplish different goals in various contexts and situations. Characterizing goals and contexts of information seeking can inform the design of personalized IR systems that adapt to different types of information problems. One way of exploring goals and contexts is to consider the tasks that motivate information seeking. The relationship between aspects of tasks and information seeking behavior have been discussed and examined in many studies. Various task classifications have been developed to categorize tasks at each or all of the three levels, work tasks, information seeking tasks, and information search tasks.

Campbell [7] focused on objective task complexity and developed a typology of complex tasks using four task attributes: (1) multiple paths to arrive at a desired end-state; (2) multiple desired end-states; (3) conflicting interdependence among paths to multiple desired outcomes; and (4) uncertain or probabilistic linkages among paths and outcomes. Complex tasks can be classified by determining the degree to which a task integrates each attribute and the total number of attributes contained in a task. Bystr´om and J´arvelin

[6] approached task complexity differently by categorizing it based on users' perceptions. They discovered that as the complexity of a task increased, the number of information sources consulted increased, so did the needs for domain knowledge and problem-solving information. Meanwhile, the success of the information seeking process decreased.

Tasks have also been analyzed from the angle of information search tasks, which refer to the activities users perform in IR systems to support their work tasks [15]. Reid [28] classified search tasks into tasks that are internally generated and those that are externally generated. An internally generated task is conceived and executed by the same person, whereas an externally generated task is a task where the task performer is different from the task setter. She argued that a task-oriented test collection is a realistic method of evaluating IR systems because it recognizes the primary importance of the task in user motivation. Gwizdka and Spence [13] suggested that subjective task difficulty is affected by a few factors such as the unique Webpages visited, the dwell time on each page, and the linearity of the navigation paths. The relative importance of these factors in predicting subjective difficulty is also influenced by objective task difficulty.

Li and Belkin [15] addressed the limitation of previous task classifications (e.g., focusing only on one or a few facets of tasks) by developing a faceted classification that comprehensively reflects the characteristics of tasks, and this faceted classification can be used for classifying all levels of tasks (i.e. work, information seeking, and search). This faceted classification scheme captures both the external characteristics of a task (i.e. source, task doer, time, process, product, and goal) and the internal attributes of a task that include task features (i.e. objective complexity and interdependence) as well as users' perception of a task (i.e. salience, urgency, difficulty, complexity, topical knowledge, knowledge of task procedure). This scheme has been widely employed by researchers in designing and characterizing tasks (e.g., [9, 19]) in IR studies. Due to its versatility and comprehensiveness, it will be the classification scheme used for characterize search tasks when a proposal for moving forward is presented later in this paper.

3.2 Recommender systems

Recommender system is defined as an assistant tool that can automate or support a general recommendation process from the system side [38]. During the information seeking process, recommendations from systems can help users filter out part of irrelevant information, decrease their cognitive load, and offer them cues (e.g., queries, terms), answers, and objects (e.g. documents, Web pages) to address their information needs [27]. Therefore, to better support information seeking practices, recommender systems must obtain preferences from people concerning the relevant domain and understand the gaps in their knowledge base based on their information seeking and search behaviors [38]. To reach this goal, both the type and content of recommendations need to be tailored to different tasks, gaps, and goals of information seekers.

The previous studies of recommender systems can be classified into two categories: collaborative filtering (CF) and content-based filtering (CB) [29]. Specifically, CF uses information-filtering techniques in formulating recommendations based on users' previous interactions with systems (e.g., queries issued, page visited and bookmarked, and search results clicked), whereas CB analyzes a set of documents rated by an individual user and detects content-based similarities between items to infer a user's preferences and needs for recommendations. To facilitate the inference of user's needs and the connections between documents, researchers in this field have employed different data mining techniques (e.g., decision tree, neural network, regression, K-NN, clustering, link analysis) according to different focused features of contents, interactions, and individuals [27].

For both of these two approaches, however, recommendations from previous systems have always been limited to a set of object types, such as queries, documents, images, and videos. These types of objects may be enough for serving navigational search goals of users who have specific known Websites or items in mind. If an information seeker needs to get advice or explore open-ended questions, then the recommender system must gain a deeper understanding of the user's goal and task, go beyond the limits of online resources, and provide more personalized options for them, such as search strategies, people who can answer the questions, or an idea as a starting point for exploration [30]. Under this circumstance, people can acquire the help and strategies (rather than merely objects) offered by proactive recommender systems and better tackle their potential needs and knowledge gaps. This will be one of the important components of an Information Fostering system.

3.3 Proactive information seeking

According to Wilson [45](p.49), information seeking behavior is "the purposive seeking for information as a consequence of a need to satisfy some goal." Guided by different goals and tasks, people often seek and search information in different ways. For example, when seeking to satisfy a navigational goal, an information searcher may directly type in a URL or specific queries. However, for information searchers with undirected, exploratory goals, queries issued in a search engine often need to be broad and cover more potentially relevant topics [30]. In this sense, being proactive in information seeking is to go beyond the immediate, short-term relationship between specific queries and documents, and to try to understand the tasks, topics, and goals of information seekers for better predicting and serving of their potential information needs.

In the relevant literature, the goals and tasks of information seeking have been analyzed mainly in the context of online information search and retrieval. Rose and Levinson [30] investigated the queries issued by users in a search engine and proposed a hierarchical typology of user's search goals. Similarly, drawing on the ideas of the sense-making approach, Savolainen and Kari [32] revealed the discontinuous and dynamic nature of Web searching episodes and developed a conceptual framework of knowledge gaps faced by searchers and corresponding gap-bridging strategies. According to their findings, an understanding of search goals, user's knowledge gaps, and gap-bridging strategies can help tackle the larger problems of representing user goals in an IR system. This could be instrumental in supporting proactive IR.

Besides the goals and gaps, researchers in related fields have also explored the relations between task facets and information seeking and search behavior, aiming at understanding and automatically predicting people's task contexts based on their interactions with information and/or systems. For example, Bystrom and Jarvelin

[6] found that the complexity of task is closely associated with the task doer's information seeking behavior (i.e., types of information needed, information channels used). Liu, Liu, Gwizdka, and Belkin [18] indicate that information searcher's behaviors (i.e., documents dwell time and number of content pages viewed per query) can be used as indicators to predict search task difficulty. In addition to search behavior per se, Mostafa and Gwizdka [26] suggest that various neural signals (i.e., eye movement, EEG, fMRI) can also be employed as indicators to test the hypothesized relations between behavioral markers and search task facets. To support proactive information seeking and retrieval, future studies need to continue exploring the underlying connections between information search behavior, task facets, and other contextual information. And this is what an Information Fostering system should do – understand the relationships among the nature of a search task, various contextual factors during a search episode, and potential problems faced and help to be offered.

4 PAST AND PRELIMINARY WORK

To address the larger problem of creating an Information Fostering framework and a system, one needs to solve several sub-problems. This section provides some details on how several scholars, including the author, have already made reasonable to substantial advancements on these sub-problems.

4.1 Information seeking barriers, failures, and solutions

An information seeking failure is defined as a situation in which an individual could not satisfy his/her information needs at work, at school, or in his/her everyday life. To overcome the limitations of previous works in which only failures in Web searching were considered, the author investigated individuals' failures in information seeking both online and offline. Employing Amazon's Mechanical Turk (MTurk) as a recruiting and surveying platform, the author was able to obtain a diverse pool of 63 participants from various age groups and educational backgrounds. MTurk users are arguably more representative of the general population than academic participants [11].

Data collected from a qualitative survey that gathered 208 real-life examples of information seeking failures and 10 semi-structured interviews with 10 different participants were analyzed using various theoretical frameworks of tasks (e.g., [15]), strategies (e.g., [4]), and barriers (e.g., [8, 31, 37]). The findings indicated that a wide range of external and internal factors caused individuals' failures that were affected by multiple aspects of information seekers' tasks and strategies. These components of information seeking mutually influenced one another and together they led to individuals' information seeking outcomes. Their information needs were often too specific to be fulfilled by the general information available to them. Also, although individuals overwhelmingly chose to search online, the Web might not be the ideal place to look for information in some situations. The respondents often wished to directly seek help from humans after having unsuccessful experiences on the Web. Some of the findings concerning the support individuals wanted are reported in [40, 41]. These studies compiled a list of barriers that users may face in information seeking, and developed a classification of "help" needed by users when having difficulties in finding information, which can be used in the design and data analysis for further research.

4.2 Generative query recommendations

Current work in recommendation in search tasks predominantly focuses on recommending queries from a search log. The possible queries are assumed to be given in advance. User's input is given to the recommendation system; the output is one of these possible queries. However, to gauge the quality of these queries – which are typically user queries – Mitsui and Shah [23] conducted a Mechanical Turk task in which they asked users to guess the query that generated a set of results. The results provided to users represented the first 10 results (i.e., the first page) generated from a real query to Google. It was found that a simple algorithm that output the most frequent words as a query gave more accurate queries than users. This suggests that users may not be able to accurately represent their information needs in a query – and therefore approaches that suggest other users' queries may not yield the most effective recommendations.

Considering this, there are some recent works that employ a generative approach to query recommendation in which completely new queries are generated rather than extracted. In this approach, a typical method is to use the popular topic modeling approach – Latent Dirichlet Allocation (LDA [5]) – to model a user's context as a distribution of topics. This model is trained on an unstructured corpus of text documents and can infer the topic distribution of any given document or set of documents. Given a user's past queries, the author modeled their past queries as a set of topics and recommended the least explored topic that was still relevant to their search, according to a threshold criterion. He then generated multi-word query terms from this topic, using a skipgram model [20] that could create coherent phrases. This was then compared to an existing work that suggested previously issued queries [39]. While the author's method [22] recommended more diverse queries than users, it still strayed from the original topic more so than the competing method, suggesting that a different model incorporating further information (such as task information) should be used to control the query generation process.

4.3 Strategy recommendations

The author has also conducted research in recommending strategies (a sequence of steps involving queries, documents, and relevant information) for exploratory search tasks. These tasks comprise multiple queries and hence lend themselves easily to the problem of Information Fostering. Typical features of an exploratory search task are its open-endedness and the multi-faceted nature of the task, requiring multiple queries [43]. In previous work, the author extracted implicit features of the search process based on the literature, which measured the discovery, creativity, and exploration of the users' search processes. He used these features to recommend trails of search queries to users who were likely to underperform in the future [34]. The recommendations, while a sequence of queries, were "search paths" as defined in by White and Huang [42], which is a series of syntactically related queries. This work was able to accurately predict the performance of a user a few steps ahead within their search, and moreover could offer effective recommendations. The recommendations offered became more effective as the recommendations were given later in the search process. They moreover greatly enhanced user performance, according to quantitative simulations and qualitative judgments of the simulations and user performance. Figure 4 shows the probabilities of helping vs. hurting a searcher through these strategy recommendations at different times. As shown, for the most part the method is able to help the searcher in improving their retrieval effectiveness.

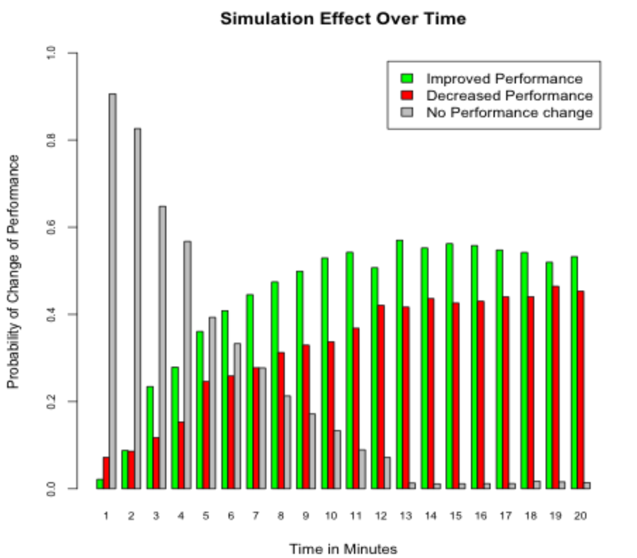

Figure 4: Simulation of help vs. hurt for recommending strategies to searchers.

4.4 Collaborator recommendation

As identified in the previous section, most recommender systems do not go beyond suggesting information objects. However, many studies have found that given a chance, people do want to find others who could work with them or at least help them in their information seeking tasks (e.g., [25]). The author investigated the problem of coming up with a recommendation for a collaborator during a search episode. Using 120 participants in 60 pairs working on an exploratory search task, the author built a model for finding suitable collaborators for an individual during different stages of his/her search. This model was then tested on real-life data obtained from Bing search logs and Internet Explorer browser logs. The data consisted of 8,969 search sessions by 8,051 users who were working on an exploratory search task during the overlapping time-frame. Using the collaborator recommendation model, the newly devised algorithm predicted potential collaborators for each of these searchers and simulated what would happen if they took these recommendations. Figure 5 shows the results of these simulations. As shown in the top graph, using the predicted recommendations for collaboration, the individual searchers would have achieved higher effectiveness (number of useful pages over total number of pages covered) and efficiency (amount of effectiveness per query). The bottom graph shows the likelihood of helping and hurting someone based on that recommendation at different times. This likelihood calculation can be used as the level of confidence while making a recommendation for a potential collaboration to someone. The details of the method, the model, and the results can be found in [12]. Continuing this thread of research, the author, with his colleagues, has also created new methods [35, 36] for detecting potential roles that collaborators could play in a collaborative search project to optimize the outcomes for their task, once again emphasizing on finding opportunities in an information seeking situation.

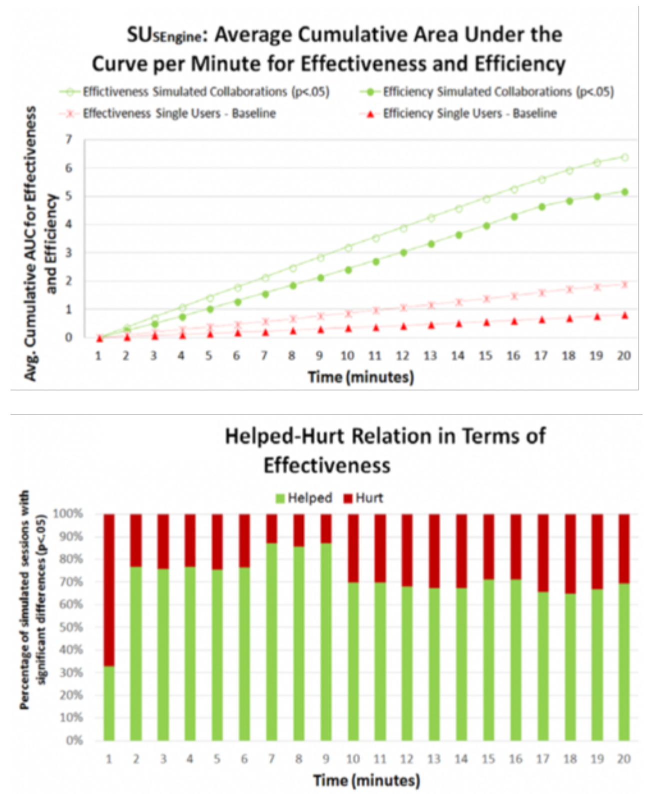

Figure 5: Simulations showing effectiveness and efficiency for recommended collaborations (top) and the likelihood of help vs. hurt (bottom).

5 METHODS FOR MOVING FORWARD

As described earlier, in order to create an Information Fostering or Proactive IR system, we need to understand a great deal about a person's ongoing processes as well as past behaviors. More specifically, we need to extract information about three crucial elements representing one's information interaction: topic, task, and intention/purpose. While there are several possible ways to pursue these goals, this section proposes two different methods – one based on collecting small but rich data to extract task, problems, and help information, and another based on using large-scale data with a more formalized approach from machine learning. The former could be more suitable for an academic environment, whereas the latter could be more appropriate in an industrial setting where large amounts of user data is available.

5.1 Extracting topic, task, problems, and help information

To explicate the topic of a user's information interaction activities, one could use a method described earlier in Section 4.2. In addition to the techniques described in that section, there have been several other approaches one could find in the literature for extracting topic information [1]. On the other hand, extracting task and intention information is considered hard. There have been a few recent efforts to identify an information seeker's intention in a search task [21, 24], but hardly any empirical work on automatically extracting the nature of a search task. Therefore, in this subsection, a short proposal is presented for learning about a person's task using behavioral data. In addition to extracting information about the task, this proposal also covers learning about what problems information seekers face and what solutions could be provided to them that are not just exclusive to the search system they are using. This work will involve building a Task Model as well as a Problem-Help Model.

An appropriate method for building these models is a user study with participants divided in a few (e.g., four) groups based on the kind of task they are assigned. These tasks could be constructed using the faceted task classification presented in Table 9 (p.1834-1835) of Li and Belkin [15]. While the authors in that article identified several facets of a task, many of them may either be not relevant or of no interest. Given the nature of the search scenarios considered here, it is imperative that the tasks will be done by individuals ('Task doer' facet), be unique ('Time → Frequency' facet), and have a single goal ('Goal → Quantity' facet). Also, given the nature of the study design (controlled lab experiments), the tasks will be externally assigned ('Source of task' facet), done in a short time ('Time → Length' facet), and have low interdependence ('Task characteristics → Interdependence' facet). The classification scheme also includes various factors based on user's perception of the task, and are not relevant for designing the tasks. This leaves out the aspects of 'Product', 'Goal → Quality', 'Time → Stage', and 'Task characteristics → Objective task complexity'.

During the lab session, the system will prompt the participant to take a brief questionnaire every time he/she goes to run a new query on a search engine. This can be achieved by modifying an open-source browser-based plug-in such as Coagmento [33].

For this questionnaire, the choices can be drawn from the literature. Existing studies have presented a variety of information seeking barriers/obstacles using varying terminology (e.g., [8, 31, 37]). Individuals frequently face difficulties when looking for information across various disciplines and professions. Barriers may originate from inside an individual (e.g., lack of subject knowledge, searching skills, or patience), be imposed on an individual from outside (e.g., time constraint, restricted access), or be interpersonal (e.g., insufficient support from other people) [31, 37]. Obstacles in finding information can also be profession-specific. For instance, Attfield and Dowell [2] discovered in their study of the field of journalism that product constraints – such as deadline and work-count – often brought challenges to journalists' information seeking. The problems could be deepened by the uncertain nature of writing for news articles. The author has synthesized the literature to generate a list of information seeking barriers [40, 41]. For this study, the list can be refined to include only problems associated with Web searching. For the "help" needed by individuals when encountering a gap, inspiration can be drawn from the work by Wang and Shah [41] that explored the user-reported remedies in information seeking. The original classification included eight types of support – such as experts and new Web features – that may assist in individuals' information seeking. Only the types that are applicable here should be retained. This results in the following questions and responses to be shown to the user before each query execution.

(1) What problem are you facing right now?
☐ Time constraint ☐ Too much information ☐ Information is too scattered ☐ Information is not up-to-date ☐ Poor quality display of text or graphics ☐ Information is unreliable ☐ Information is not available ☐ Topic is too unclear ☐ Unable to articulate my information needs ☐ Unaware of relevant information sources ☐ Unable to understand the information retrieved ☐ Other (please write)
(2) What will help you?
☐ Query recommendation ☐ Information source recommendation (e.g., documents, Websites) ☐ Step-by-step instructions to follow ☐ A peer/friend to talk to ☐ An expert to consult ☐ Other (please write)

The objective of this study is two-fold: using the behavioral data from one's search session, build a task model that could predict various aspects (nature) of the search task; and using the behavior data from one's search session, predict the potential problems as well as possible help at each query execution time (or for a given query segment). Figure 6 outlines these two objectives.

First, using all of the available behavioral data as features and considering task aspects as dependent variables, a prediction model, called Task Model, can be built by doing a multivariate ANOVA (MANOVA) and/or multiple regression analyses. For this analysis, first all the behavioral data from all the sessions (30 minutes/session) should be used. Various model-related parameters including R^2 and η (effect size) will inform about accuracy and robustness of the model. After that, a 10-fold cross-validation can be run to do training-testing on 70-30 data splits. This will provide prediction accuracy results.

Furthering this analysis, various subsets of the available data can be created:

(1) Based on task types (total 4);
(2) Type of behavioral data (three types – interaction logs, mouse movement, eye-tracking); and
(3) Amount of time (e.g., 5 minutes, 10 minutes, 15 minutes, 20 minutes, and 25 minutes).

A number of analyses (at least 4x3x5 = 60) can then be run to see how much and what kind of behavioral data contribute to detecting the nature of a search task. Appropriate exploratory factor analysis and PCA methods can also be used to derive grouping and reduction of variables.

Similar analyses can be performed for building a model that predicts potential problems and help (Problem-Help Model) with the following two differences: (1) Rather than creating subsets of behavior data according to time, they can be created by query segments; and (2) Prediction of "help" can be mediated by "problems" in addition to the behavioral data, which means "problems" will

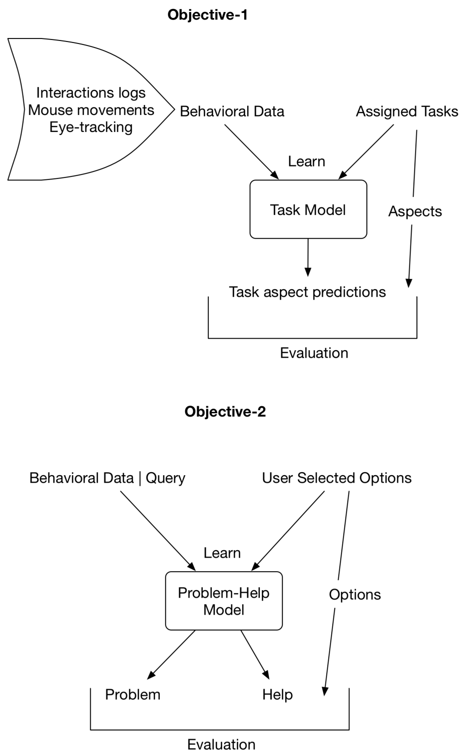

Figure 6: Outline of addressing two objectives for the user study.

possibly serve as an intervening variable, leading to some form of analysis of covariate (ANCOVA).

5.2 Building User, Background, and Activity models

Here is another path forward for building appropriate models that could help an Information Fostering system to project a user's trajectory and make suggestions based on potentially encountering any problems. In order to understand what a user is doing and where that path is heading, three kinds of information is needed: (1) the user's past behaviors; (2) the world knowledge; and (3) the user's current behavior. These three things will be materialized as User Model (M_U), Background Model (M_B), and Activity Model (M_A).

Both M_U and M_A, which are associated with a user's past or current activities, can be represented using the following attributes: Activity, Object, Content, Time-Start, Time-End, Prior, Posterior, and Assessment. Collectively these attributes for various activities/objects from the user's past or the present will capture the knowledge about what the user did before and/or what he/she is trying to do now. Background Model (M_B), on the other hand, can have the following attributes: Activity, Object, Content, Frequency, Duration, Prior, Posterior, and Assessment.

To build these models, data from past lab and field studies involving people's searching and browsing activities can be used. These datasets will be used to build M_U and M_B, whereas M_A can be built on-the-fly using various behavioral data, such as query issuing, clicks, page visitation, dwell time, etc.

Each of these models contain information for an action, its response, and an assessment. For instance, there could be a datapoint about a query (action), a result set (response), and assessment (whether the user clicked on a result or not). Using these attributes, the three models could be represented as the following:

$$M_A = \sum_i (\alpha_i a_i - r_i) \tag{1}$$

$$M_B = \sum_i (\beta_i a_i - r_i) \tag{2}$$

$$M_U = \sum_i (\gamma_i a_i - r_i) \tag{3}$$

Here, α, β, and γ are parameters associated with those models. In order to see if the user's current activity as represented using M_A will lead to something useful or not, we could look at the slope of its corresponding function, given an action a_i. For this, we need to take a partial derivative of M_A with respect to as following.

$$\begin{aligned}\frac{\partial}{\partial a_j} M_A &= \sum_i (\alpha_i a_i - r_i) \frac{\partial}{\partial a_j} (\alpha_0 a_0 + \ldots + \alpha_j a_j + \ldots + \alpha_n a_n - r_i) \\ &= \sum_i (\alpha_i a_i - r_i) \alpha_j \end{aligned} \tag{4}$$

This informs us about the slope of M_A along the action that we are considering (a_j). We can then use M_B and M_U to see how this particular action in the user's and/or the world's past was assessed. Based on that, we could predict the chances of the user to take that action and continue on the path. If we determine that there is a good likelihood of the user continuing on that trajectory, we could evaluate where that trajectory could go (next possible actions) and find out corresponding assessments using the same approach as described above. In other words, at a given time, we start looking at various possibilities based on a computed projection of the user's trajectory and assessments from the past behaviors of the user and of the world. If we detect a problem along the way, we could offer the user a different trajectory/strategy that we have calculated to show better promise.

In previous work, the author extracted implicit features of the search process based on the literature, which measured the discovery, creativity, and exploration of users' search processes. He used these features to recommend trails of search queries to users who were likely to underperform in the future [34]. The recommendations, while a sequence of queries, were *search paths* as defined in by White and Huang [42], which is a series of syntactically related queries. This work was able to accurately predict the performance of a user a few steps ahead within their search, and moreover could offer effective recommendations. But this work was limited to using the data about only the participants who did the same task in the controlled lab environment. The proposal presented here allows one to not only scale it up, but also relax the assumption about the nature of the task.

6 CONCLUSION

We lack enough understanding of addressing the problems that information seekers face due to their inability to express their information needs, recognizing a potential problem during a search episode, and identifying support needed that goes beyond what a typical search system could provide. Most recommender systems try to mitigate these problems by suggesting information objects (queries, documents), disregarding a deeper understanding of the task at hand or the possibility of recommendations that involve process/strategy, people, and other forms.

Information Fostering is an idea of being proactive in an information seeking situation by projecting ahead for potential problems and opportunities, and guiding the user to a path that could avoid those problems and/or capture those opportunities before it is too late. Such a path may include recommendations for not only information objects such as queries and documents, but also processes/strategies, and people. A true Information Fostering system may even recommend an information seeker to stop or reconsider a pursuit, like what a good teacher or a friend would do, instead of blindly giving him/her a suggestion for an information object.

In this Perspective Paper, a case was made for this new paradigm that looks at the other side of the information seeking coin – providing relevant information and help/guidance to a person without them explicitly asking or even realizing the need. This was done by reviewing existing systems and literature and outlining what we need next, as well as proposing a couple of methods for moving forward the research agenda. The author believes Information Fostering will provide us the next conceptual leap in the field of human-centered IR, and will become even more relevant as we move further toward conversation-based intelligent assistants for our informational objectives.

7 ACKNOWLEDGMENTS

This work is supported by the National Science Foundation (NSF) grant IIS-1717488.

REFERENCES

[1] J. Allan. Introduction to Topic Detection and Tracking. In *Topic Detection and Tracking: Event-based Information Organization*, volume 12, pages 1–16. 2002.
[2] S. Attfield and J. Dowell. Information seeking and use by newspaper journalists. *Journal of Documentation*, 59(2):187–204, apr 2003.
[3] N. J. Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5:133–143, 1980.
[4] N. J. Belkin, P. G. Marchetti, and C. Cool. BRAQUE: Design of an Interface to Support User Interaction in Information Retrieval. *IP&M*, 29(3):325–344, 1993.
[5] D. M. Blei, B. B. Edu, A. Y. Ng, A. S. Edu, M. I. Jordan, and J. B. Edu. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.
[6] K. Byström and K. Järvelin. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2):191–213, 1995.
[7] D. J. Campbell. Task Complexity: A Review and Analysis. *The Academy of Management Review*, 13(1):40–52, 1988.
[8] S. Chowdhury, F. Gibb, and M. Landoni. A model of uncertainty and its relation to information seeking and retrieval (IS&R). *JDoc*, 70(4):5, 2014.
[9] M. J. Cole, J. Gwizdka, C. Liu, R. Bierig, N. J. Belkin, and X. Zhang. Task and user effects on reading patterns in information search. *Interacting with Computers*, 23(4):346–362, 2011.
[10] B. Dervin. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2):36–46, 1998.
[11] J. Erickson, C. S. Mackenzie, V. H. Menec, and D. S. Bailis. The effect of time perspectives on mental health information processing and help-seeking attitudes and intentions in younger versus older adults. *Aging & Mental Health*, 21(3):259–271, 2017.
[12] R. González-Ibáñez, C. Shah, and R. W. White. Capturing collabportunities: A method to evaluate collaboration opportunities in information search using pseudo-collaboration. *JASIST*, 66(9):1897–1912, 2015.
[13] J. Gwizdka and I. Spence. What Can Searching Behavior Tell Us About the Difficulty of Information Tasks? A Study of Web Navigation. *Proceedings of the American Society for Information Science and Technology*, 43(1):1–22, 2007.
[14] P. Ingwersen. Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52:3–50, 1996.
[15] Y. Li and N. J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Information Processing and Management*, 44(6):1822–1837, 2008.
[16] Y. Li and Y. Guo. Be proactive for better decisions: Predicting information seeking in the context of earthquake risk. *International Journal of Disaster Risk Reduction*, 19:75–83, 2016.
[17] O. Liechti and Y. Sumi. Editorial: Awareness and the WWW. *International Journal of Human Computer Studies*, 56(1):1–5, 2002.
[18] C. Liu, F. Yang, Y. Zhao, Q. Jiang, and L. Zhang. What does time constraint mean to information searchers? In *Proceedings of the 5th Information Interaction in Context Symposium on - IIiX '14*, pages 227–230, 2014.
[19] J. Liu, C. Liu, J. Gwizdka, and N. J. Belkin. Can search systems detect users' task difficulty? *Proceeding of the ACM SIGIR Conference*, (August 2015):845, 2010.
[20] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, J. Dean, K. Chen, J. Dean, T. Mikolov, and K. Chen. Distributed Representations of Words and Phrases and their Compositionality. In *NIPS'14*, volume cs.CL, pages 3111–3119, 2013.
[21] M. Mitsui, J. Liu, N. J. Belkin, and C. Shah. Predicting Information Seeking Intentions from Search Behaviors. In *Proceeding of the ACM SIGIR Conference*, pages 1121–1124, New York, New York, USA, 2017. ACM Press.
[22] M. Mitsui and C. Shah. Multi-Word Generative Query Recommendation Using Topic Modeling. *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, pages 27–30, 2016.
[23] M. Mitsui and C. Shah. Query Generation as Result Aggregation for Knowledge Representation. In *Proceedings of the HICSS*, pages 4365–4374, 2017.
[24] M. Mitsui, C. Shah, and N. J. Belkin. Extracting Information Seeking Intentions for Web Search Sessions. In *Proceeding of the ACM SIGIR Conference*, pages 841–844, New York, New York, USA, 2016. ACM Press.
[25] M. R. Morris. A Survey of Collaborative Web Search Practices. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 1657–1660, Florence, Italy, 2008.
[26] J. Mostafa and J. Gwizdka. Deepening the Role of the User. In *Proceedings of ACM CHIIR Conference*, pages 63–70, 2016.
[27] D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim. A literature review and classification of recommender systems research, 2012.
[28] J. Reid. A Task-Oriented Non-Interactive Evaluation Methodology for Information Retrieval Systems. *Information Retrieval*, 2(1):115–129, 2000.
[29] F. Ricci, L. Rokach, and B. Shapira. Introduction to Recommender Systems Handbook. In *Recommender Systems Handbook*, pages 1–35. 2011.
[30] D. E. Rose and D. Levinson. Understanding user goals in web search. In *Proceedings of the 13th conference on World Wide Web - WWW '04*, page 13, 2004.
[31] R. Savolainen. Cognitive barriers to information seeking: A conceptual analysis. *Journal of Information Science*, 41(5), 2015.
[32] R. Savolainen and J. Kari. Facing and bridging gaps in Web searching. *Information Processing and Management*, 42(2):519–537, 2006.
[33] C. Shah. Coagmento- A Collaborative Information Seeking, Synthesis and Sense-Making Framework (an integrated demo). In *Proceedings of Computer Supported Cooperative Work (CSCW)*, Savannah, GA, February 2010.
[34] C. Shah, C. Hendahewa, and R. González-Ibáñez. Rain or shine? Forecasting search process performance in exploratory search tasks. *Journal of the Association for Information Science and Technology*, 67(7):1607–1623, jun 2016.
[35] L. Soulier, C. Shah, and L. Tamine. User-Driven System-Mediated Collaborative Information Retrieval. In *Proceedings of ACM SIGIR*, pages 485–494, Gold Coast, Australia, 2014.
[36] L. Soulier, L. Tamine, and C. Shah. MineRank: Leveraging users' latent roles for unsupervised collaborative information retrieval. *IP&M*, 52(6):1122–1141, 2016.
[37] M. Świgoń. Information limits: Definition, typology and types. *Aslib Proceedings*, 63(4):364–379, 2011.
[38] L. Terveen and W. Hill. Beyond recommender systems: Helping people help each other. *HCI in the New Millennium*, (1):487–509, 2001.
[39] H. Vahabi, M. Ackerman, D. Loker, R. Baeza-Yates, and A. Lopez-Ortiz. Orthogonal query recommendation. *Proceedings of ACM RecSys 2013*, pages 33–40, 2013.
[40] Y. Wang, S. Sarkar, and C. Shah. Investigating Information Seekers' Selection of Interpersonal and Impersonal Sources. *Proceedings of ACM CHIIR 2017*, pages 353–356, 2017.
[41] Y. Wang and C. Shah. Exploring support for the unconquerable barriers in information seeking. *Proceedings of the ASIST*, 53(1):1–5, jan 2016.
[42] R. W. White and J. Huang. Assessing the scenic route: measuring the value of search trails in web logs. *Proceeding of the ACM SIGIR*, pages 587–594, 2010.
[43] R. W. White and R. A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan & Claypool Publishers, 2009.
[44] T. Wilson. On user studies and information needs. *Journal of Documentation*, 37(1):3–15, dec 1981.
[45] T. D. Wilson. Human information behavior. *Informing Science*, 3(2):49–55, 2000.