Capturing collabportunities: A method to evaluate collaboration opportunities in information search using pseudocollaboration

Citation to Publisher Version: González-Ibáñez, Roberto, Shah, Chirag & White, Ryen W. (2014). Capturing collabportunities: A method to evaluate collaboration opportunities in information search using pseudocollaboration. *Journal of the Association for Information Science and Technology 66*(9), 1897-1912. https://dx.doi.org/10.1002/asi.23288

Citation to *this* Version: González-Ibáñez, Roberto, Shah, Chirag & White, Ryen W. (2014). Capturing collabportunities: A method to evaluate collaboration opportunities in information search using pseudocollaboration. *Journal of the Association for Information Science and Technology 66*(9), 1897-1912. Retrieved from http://dx.doi.org/doi:10.7282/T31R6SB8

Capturing *Collabportunities*: A Method to Evaluate Collaboration Opportunities in Information Search Using Pseudocollaboration

Roberto González-Ibáñez
Departamento de Ingeniería Informatica, Universidad de Santiago de Chile, Santiago 9170124, Chile. E-mail: roberto.gonzalez.i@usach.cl

Chirag Shah
School of Communication & Information (SC&I), Rutgers University, 4 Huntington Street, New Brunswick, NJ 08901-1071. E-mail: chirags@rutgers.edu

Ryen W. White
Microsoft Research, One Microsoft Way, Redmond, WA 98052. E-mail: ryenw@microsoft.com

In explicit collaborative search, two or more individuals coordinate their efforts toward a shared goal. Every day, Internet users with similar information needs have the potential to collaborate. However, online search is typically performed in solitude. Existing search systems do not promote explicit collaborations, and collaboration opportunities (collabportunities) are missed. In this article, we describe a method to evaluate the feasibility of transforming these collabportunities into recommendations for explicit collaboration. We developed a technique called pseudocollaboration to evaluate the benefits and costs of collabportunities through simulations. We evaluate the performance of our method using three data sets: (a) data from single users' search sessions, (b) data with collaborative search sessions between pairs of searchers, and (c) logs from a large-scale search engine with search sessions of thousands of searchers. Our results establish when and how collabportunities would significantly help or hinder the search process versus searches conducted individually. The method that we describe has implications for the design and implementation of recommendation systems for explicit collaboration. It also connects system-mediated and user-mediated collaborative search, whereby the system evaluates the likely benefits of collaborating for a search task and helps searchers make more informed decisions on initiating and executing such a collaboration.

Received July 16, 2013; revised February 25, 2014; accepted March 6, 2014

(wileyonlinelibrary.com). DOI: 10.1002/asi.23288

Introduction

Many Internet users search for information about similar or identical topics simultaneously with no awareness of each other's highly related search activity. As motivation for our research, we analyzed search engine logs containing traces of search behavior from consenting users of a browser toolbar distributed by the Microsoft Bing search engine on a particular exploratory search topic, in this case, the British Petroleum (BP) *Deepwater Horizon* oil spill (explained later). This analysis revealed that, around the time of the incident in 2010, there were always at least two people searching for related information within the same hour, and about 38% of the time there were at least two people searching for the topic within the same minute. Figure 1 illustrates this phenomenon for users searching on this topic in each hour of the day during October 6, 2010 (Figure 1a), and also in each minute from 12:00 to 1:00 PM of the same day (Figure 1b). Unless people previously coordinate and agree to collaborate during search—with or without the support of tools—collaboration opportunities (referred to in this article as *collabportunities*) are typically missed, but they could be captured. To address this issue, search engines could provide direct support for connecting searchers with candidate collaborators who are pursuing similar search tasks simultaneously. Searchers could then decide whether to pursue the suggested collaboration.

In information retrieval (IR) research, many methods have been developed to help people search more effectively,

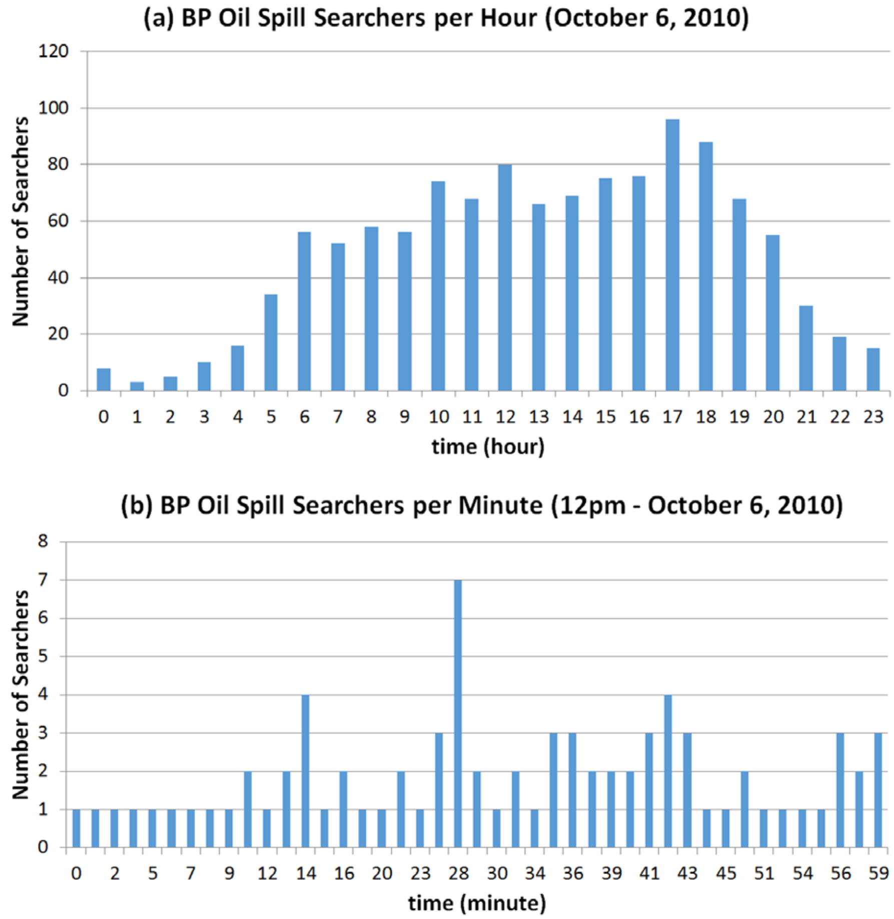

FIG. 1. Number of users searching for information about the British Petroleum (BP) oil spill per hour on October 6, 2010 (**a**), and per minute within a single hour on October 6, 2010 (**b**). Log sample obtained from the Microsoft Bing web search engine. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

including query suggestions (Anick & Kantamneni, 2008), result recommendations (Resnick & Varian, 1997), information filtering (Adomavicius & Tuzhilin, 2005), and personalization (Belkin, 2000). In addition to supporting an individual's search activity, researchers have argued that collaboration could facilitate more enhanced information exploration in part by exploiting the different skills and experiences of collaborators (Hyldegard, 2006, 2009; Shah, 2010b; Shah & González-Ibáñez, 2011; Twidale, Nichols, & Paice, 1997). For instance, a subject specialist (e.g., an attorney) and a search specialist (e.g., a legal librarian) could collaborate to retrieve better results and deeper understanding of the material and its sources. There are many other similar scenarios in which two or more individuals can combine their skills via collaboration to fulfill shared information needs and improve the quality of search outcomes. Unfortunately, in many search settings, such opportunities to collaborate are latent and often remain undiscovered. To the best of our knowledge, there are no current IR solutions that dynamically identify and evaluate collaboration opportunities to promote explicit and intentional collaboration among searchers.

Capitalizing on collabportunities could allow an information seeker to discover information that is difficult to find or synthesize by an individual (Pickens, Golovchinsky, Shah, Qvarfordt, & Back, 2008), or it could help boost novelty/diversity in addition to relevance (Shah, 2010a). There are two ways such collaborative information-seeking (CIS) activities are currently supported, system mediated and user mediated (Golovchinsky, Pickens, & Back, 2008). The former imposes roles that can enhance search performance at the expense of autonomy for searchers, whereas the latter gives more control to searchers with little or no intervention from the system. These approaches have their advantages and disadvantages, but a common disadvantage is that neither of them can provide a sense of relative benefit to the searcher from engaging in collaboration, which leads to lost collabportunities. This problem could be addressed by (a) identifying collabportunities and (b) informing searchers about potential benefits of collaborating during their search.

On the assumption that specific collabportunities exist at a given time, one challenge that we have to address is determining whether they should be transformed into recommendations for collaboration and offered to users. To perform this evaluation, it is necessary to measure the potential of a collabportunity, estimating how it could help or hinder the information search process. In other words, we have to measure the relative costs and benefits of collaborating on a search task at a given time. As stated by González-Ibáñez, Shah, and White (2012), collaboration is typically associated with more successful searching; however, in the same way that collaboration can help, it can also hurt the process and its products, for example, reducing efficiency by introducing the need to coordinate search activity or reducing effectiveness by directing searchers to irrelevant material. In the context of collabportunity evaluation, if the benefits of collaborating with another searcher exceed the associated costs, then such opportunities can be transformed into actual recommendations for collaboration and recommended to both searchers by the search system.

There are many aspects of collabportunities that we could investigate, including the ways in which search activity could be disrupted or enhanced through collaboration. To narrow our exploration in this area, we focus our study on a single exploratory topic deemed representative of the types of scenarios in which we believe that this approach could help. We answer the following research question.

- When are good opportunities for people to collaborate in an exploratory search? In other words, can we identify points along a user's search process at which it is advantageous for them to collaborate with other searchers performing a similar task?

To address this question, we propose a method to evaluate collabportunities using an adaptation of the pseudocollaboration technique described by González-Ibáñez et al. (2012). In our previous work, we devised a pseudocollaboration method of timely assistance to searchers. In the work reported in this article, we extend that technique to evaluate the feasibility of potential collaborations between searchers who share similar information needs based on the simulation of collaborative search processes between them. From these simulations, we evaluate at each time point (each minute in our case) the benefits (help) and costs (hurt) of collaboration opportunities in terms of two measures, search effectiveness and search efficiency.

As mentioned, we focus on a typical search scenario comprising exploratory information-seeking behavior. Such search tasks occur frequently, and we believe that they are sufficiently lengthy and complex (the research of Borlund and Ingwersen [1999] offers an empirical justification of this assumption) that assistance from another searcher may be valuable. As part of our investigation, we design and conduct an experiment with three data sets comprising logs of search behavior from different sources: (a) logs from a laboratory study with single users performing the exploratory search task, (b) logs from a laboratory study with collaborative pairs performing the exploratory search task, and (c) logs from the Microsoft Bing search engine covering thousands of users' search sessions on the same topic as was used for (a) and (b). The search engine log data also provide us with an opportunity to assess the performance of our method in a large-scale naturalistic setting similar to that in which it would obtain in the wild.

The remainder of this article is organized as follows. The next section describes some relevant literature on collaborative IR, mediated collaboration, and other relevant areas such as mining on-task behaviors from historic search logs. The third section provides a detailed explanation and an example of collaboration opportunities. The fourth section describes our approach to evaluate collabportunities. The fifth section describes our experiment, data, and measures. The sixth section provides a detailed explanation of the results derived from our experiments and revisits our research question. Finally, we discuss implications for the design and implementation of search systems capable of realizing the benefits of collaboration where appropriate and present future directions of our work.

Background

Previous research on collaborative IR (CIR) and collaborative information seeking has focused on implicit/unintentional and explicit/intentional forms of collaboration (Fidel et al., 2000; Fidel, Pejtersen, Cleal, & Bruce, 2004; Golovchinsky et al., 2008; Morris, 2008; Shah, 2009, 2010a). Research on implicit/unintentional collaboration, such as collaborative filtering, has targeted methods capable of exploiting individual behaviors in a population to aid users within it.

Instead of diving into definitional details, we provide Figure 2, which shows collabportunity evaluation with

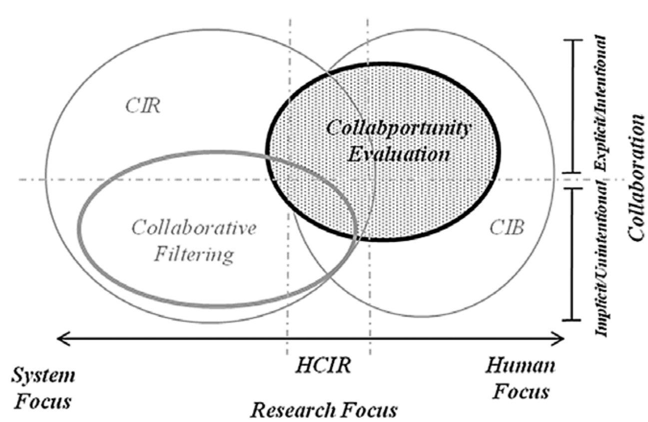

FIG. 2. Collaboration opportunities (collabportunities) evaluation with respect to collaborative information retrieval (CIR), collaborative information behavior (CIB), and collaborative filtering.

DOI: 10.1002/asi

respect to related research topics on the human–system and explicit–implicit dimensions.[1] The research focus of our study lies at the intersection of human–computer interaction, information retrieval, and human information behavior. Moreover, as we explain in the following section, collabportunity evaluation has as an objective the identification, assessment, and transformation of opportunities to collaborate in information search that are typically found in non-collaborative situations and also in scenarios in which implicit/unintentional forms of collaboration are developed. We also note that our approach is different from others such as collaborative filtering (Resnick & Varian, 1997) in that collabportunity evaluation focuses on search processes (queries and result selections) rather than products (e.g., task outcomes).

Search logs collected by search engines have also been analyzed to extract the on-task behavior of searchers. The matching with the current search situation has been performed at the query level (Agichtein, Brill, & Dumais, 2006; Joachims, 2002) using sequences of related queries (Radlinski & Joachims, 2005) and, most recently, more sophisticated models of on-task behavior (White et al., 2013). These approaches leverage the historic behavior of many searchers captured in search logs to identify resources of potential interest to the current searcher given their ongoing task. Long-term personalization of search behavior (e.g., Bennett et al., 2012) is also relevant in this context, focusing on prior task-relevant behavior from a single searcher only. None of this research on mining historic search behavior supports the initiation of synchronous collaborations between searchers as we propose in this article.

Research on explicit/intentional collaboration in information search has focused primarily on the study of behaviors and collaborative practices that can be supported with technology. Work on CIR is often characterized by the nature of collaboration, as outlined by Pickens et al. (2008):

- System or algorithmically mediated. The system acts as an active agent and provides mediation among the collaborators to enhance their productivity and experience. Example systems include Cerchiamo (Golovchinsky, Adcock, Pickens, Qvarfordt, & Back, 2008; Golovchinsky et al., 2008) and Querium (Golovchinsky, Dunnigan, & Diriye, 2012).
- User or interface mediated. The control lies with the collaborators, with the system being a passive component. The users drive the collaboration, and the system primarily provides various functions at the interface level. Examples include Ariadne (Twidale & Nichols, 1996), SearchTogether (Morris & Horvitz, 2007), and Coagmento (Shah, 2010a).

In both cases (system and user mediated), it is assumed that collaboration over search happens as a part of a larger, ongoing collaborative project among participants. Golovchinsky et al. (2012) describe a tool called Querium that supports and mediates explicit collaboration among users. It does so by providing specialized tools that enable them to perform actions such as information sharing and communication. In addition, the tool incorporates algorithmic mediation, which is system driven and does not require explicit user intervention. In systems such as SearchTogether (Morris & Horvitz, 2007) and Coagmento (Shah, 2010a), users have complete control over a set of features that support the collaborative search process. However, unlike Querium, in these systems, collaboration depends exclusively on explicit actions from users. Synchronous social question-answering systems, such as Aardvark (Horowitz & Kamvar, 2010) and IBM Community Tools (Weisz, Erickson, & Kellogg, 2006), let people pose questions to others via synchronous communication channels such as instant messaging. They assume that the participants in the dialog have clearly defined roles (an asker with a question and an answerer with the subject-matter expertise to furnish a possible answer), rather than the searcher roles that we target with our method, in which both parties assume similar roles and have similar information-seeking objectives simultaneously.

Our approach is oriented toward the detection and evaluation of latent opportunities to collaborate to determine their viability to be transformed into explicit and intentional collaboration. In this sense, we consider the identification and evaluation of collabportunities as a specialized form of system-mediated collaboration.

Collabportunities

The preceding sections have briefly introduced the concept of collabportunity in information search. Collabportunity refers to latent opportunities for collaboration between two or more individuals with common information needs at a similar time. As illustrated in Figure 1, there are often multiple users searching for the same topic at or around the same point in time (at least in the case of the exploratory topic that we targeted in the investigation). If these searchers could be connected, then they could benefit from shared experiences, including sharing resources that they have encountered during their searches so far, queries that have been particularly useful, or perspectives on the topic itself.

To illustrate this idea, consider the following example depicted in Figure 3: Users A and B have a common information need; however, they are not aware of this situation. On a given day, both users are seeking and gathering information about the same search topic simultaneously, namely, the *Deepwater Horizon* oil spill, an environmental disaster that occurred in 2010 in the Gulf of Mexico. They are seeking information about causes, implications, and reactions pertaining to this event. Despite the collaboration potential between users A and B (the collabportunity)

[1]See Appendix for definitions of relevant terms.

DOI: 10.1002/asi

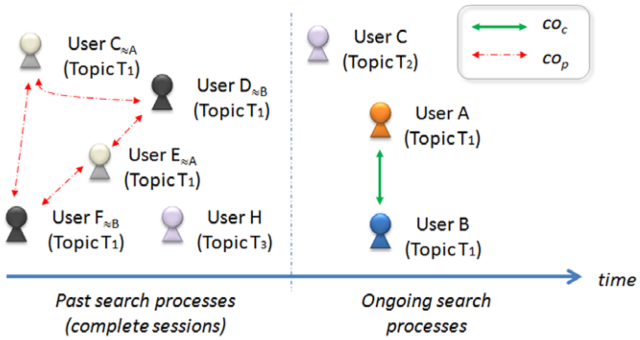

FIG. 3. View of collaboration opportunities (collabportunities) over time. co_c, current collabportunities; co_p, past collabportunities based on complete search sessions. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

while working on a search task using current search engine technology, this opportunity vanishes because of the lack of a reliable method for identifying, assessing, and promoting it as an explicit and intentional form of collaboration.

Figure 3 depicts two types of search processes, ongoing search processes (reflecting current search activity) and past search processes (reflecting historic search activity). In past search processes, people conduct search sessions on a variety of topics, and multiple people search on the same topic (e.g., users $C \approx_A$, $D \approx_B$, $E \approx_A$, $F \approx_B$). We use the symbol ≈ to denote similarity along one or more dimensions between users (e.g., $C \approx_A$ means that user C is similar to user A). Missing collabportunities like those represented between users $C \approx_A$ and $D \approx_B$ in Figure 3, or even actual collaboration episodes, could serve as valuable sources of knowledge to simulate and predict how viable current collabportunities are. This idea is further developed in the following section.

On every occasion when people share common information needs, as in the case of users A and B in Figure 3, there are latent collabportunities. In this sense, all collabportunities are potential candidates to become recommendations for actual collaborations. However, not every collabportunity guarantees a successful collaboration. A collabportunity indicates only that there exists the potential for actual collaborations between people. For instance, in our example, if users A and B were connected to allow collaboration, the process and its results might end up being better or worse than if they were to continue working in isolation. Thus, a fundamental problem to solve is the estimation of costs and benefits of collabportunities before actual collaborations take place or recommendations are made. Once collabportunities are properly evaluated, if their benefits significantly outweigh the associated costs, then collabportunities can be transformed into recommendations for collaborations and offered to the user. At that point, the decision resides with users on whether they accept the invitation and pursue the collaboration, although the system could provide an estimate of the likely benefit to help inform the decision. The next section describes a method to evaluate collabportunities in determine whether they should be promoted to recommendations for collaboration.

Methods

The evaluation of collabportunities in information search requires procedures and measures to evaluate properly the collaboration process and its products. Unfortunately, at any given time, collabportunities are just opportunities; they are neither complete processes nor products to be evaluated. To overcome this problem, we propose an evaluation approach of simulations of collaboration processes and estimations of their potential results. This estimation is performed using an adapted version of pseudocollaboration (González-Ibáñez et al., 2012), which supports searchers in their information-seeking processes by estimating an opportune moment to recommend products (e.g., pages found, queries generated) derived from past sessions that help users to find useful information and minimize the required effort.

Pseudocollaboration works through simulated collaborations using search logs of users with complete search sessions (see Figure 3). The simulation of collaborations follows six steps.

1. Given a user A with a search session in progress, locate the set of users {≈A} in the session log search patterns similar to those displayed by A during the first few minutes of the session.
2. Perform all possible combinations of users in {≈A} and their respective search sessions to simulate collaborative search processes.
3. For each simulated collaboration, *sc*, evaluate its performance per minute.
4. Determine when such simulated collaborations show significant benefits with respect to the individual search sessions of users in {≈A}.
5. Select products such as pages or queries from *sc* that display the highest improvement.
6. With the selected products, provide product recommendations to A when predictions display the best relation of cost to benefit (help–hurt).

As explained by González-Ibáñez et al. (2012), the term *pseudocollaboration* is used to indicate a “lack of explicitness and intentionality in which convenient and impersonal combinations of users’ search sessions take place” (p. 2). In this article, however, we use pseudocollaboration to predict benefits and costs of collabportunities, which later could become explicit and intentional forms of collaboration. Let us explore this further by using the following scenario. Two independent users, A and B, are looking for digital cameras in a popular online store at about the same time. Based on their queries and the types of cameras

DOI: 10.1002/asi

they have explored so far, it seems that they are looking for similar specifications. With this information, it is possible to find in the search logs of the online store that many other past users in {≈A} and {≈B} have been looking for similar cameras, leading them to purchase different models. Our method would perform all combinations of the completed search processes of users in {≈A} and {≈B} to simulate how the collaboration between A and B would play out. Successful combinations of users in {≈A} and {≈B} could show an increasing information exposure to useful content that helps users to make informed decisions. Multiple combinations displaying significant improvements compared with individual performance could serve as an indicator of the viability of the collabportunity between A and B.

Adapting the pseudocollaboration method to evaluate collabportunities involves the following steps.

1. Having a collabportunity, co_c, between users A and B (note that the approach can be applied to more than two users), find pairs of users in {≈A} and {≈B} with search patterns similar to those of A and B, respectively, during the first few minutes of the search session. This is illustrated in Figure 3.
2. For each user pair (α,β) with α ∈ {≈A} and β ∈ {≈B}, align and combine the individual search sessions of α and β to simulate collaborative search processes.
3. For each simulated collaboration, *sc*, evaluate its performance in each minute.
4. Identify when simulated collaborations *sc* show significant improvements with respect to the baseline performance of users in {≈A} and {≈B}.
5. Determine the help–hurt relation for each *sc* to estimate the goodness of an eventual collaboration between A and B.
6. Finally, if the ratio between the number of *sc* that helps and the total number of simulated collaboration is greater than a given threshold, then co_c is captured, otherwise it is discarded.

Algorithm 1 presents a generalized form of this procedure to evaluate a collabportunity with *N* users over time. As detailed in Algorithm 1, the procedure *EvaluateCollabportunity(co_c)* relies on different functions and subevaluations, briefly explained below.

Algorithm 1. General procedure to evaluate a collabportunity and determine whether it should be captured or discarded.

```
Proc EvaluateCollabportunity(co_c: collabportunity,threshold)
output capture
begin
    /* find collabportunities co_p in logs with N users similar to those in co_c */
    CO_p←getPastCollabportunities(co_c)
    for each co_p ∈ CO_p do
      sc←simulateCollaboration(co_p)
      for each user in co_p do
        for t←t_0 to getMaxTime(sc) do
          psc←getPerformance(sc,t)
          pu←getPerformance(user,t)
          if (psc>pu) or (psc<pu) at p=0.05 then
              //save performance difference at time t
              Δperformance[user,t]←psc-pu
          end if
        end for
      end for
      /* help: (psc-pu)>0 and hurt: (psc-pu) <0 */
      if HelpHurtRatio(Δperformance)>1 ∀ user in co_p then
         help←help+1
      end if
    end for
    if (help/size(CO_p)>threshold) then
        capture←true
    else
      capture←false
    end if
end
```


DOI: 10.1002/asi

First, the function *getPastCollabportunities* finds and retrieves past collabportunities co_p similar to co_c. Unlike current collabportunity co_c, past collabportunities co_p are based on search sessions that are complete and available through logs. Therefore, for each co_p a simulated collaborative search process *sc* is generated through the function *simulateCollaboration* in line 6. The function *getMaxTime* in line 8 returns the calculated time in which simulated collaboration *sc* is complete. Then, the function *getPerformance* in lines 9 and 10 can be any performance measure such as number of useful pages visited, *search effectiveness*, and *search efficiency*. The function *HelpHurtRatio* in line 18 computes the relative cost–benefit of simulated collaboration *sc* with respect to each user by using results contained in the array *Δperformance*. As specified in line 18, an *sc* will be considered as positive if the benefits exceed the costs for each user in the corresponding collabportunity, co_p; in a such case, a collabportunity is considered to be helpful (line 19). Finally, after all past collabportunities co_p are evaluated, the decision on whether the input collabportunity co_c is captured depends on the numbers of co_p in $\mathrm{CO_p}$ that succeed versus the total number of co_p in $\mathrm{CO_p}$ evaluated, which is obtained with the function *size* in line 22. This result is then compared with a given threshold, which can be tuned for precision and recall as needed in the application scenario.

Given the above general procedure to evaluate collabportunities, the next section presents an evaluation approach to test our method with a particular focus on the research question stated earlier. In other words, we seek to determine the location of good opportunities for individuals to collaborate in an exploratory search task.

Evaluation

The data that we used were gathered from laboratory settings and through logs of the Microsoft Bing Web search engine. The use of a range of data types allowed us to evaluate the performance of our method from different perspectives, improving the robustness and generalizability of the conclusions that we can draw from our investigation. Below we describe the experiment, the performance measures, and data that we used in this study.

Experiment

To test our approach, we designed an experiment using three data sets. The experiment focuses on evaluating whether a given collabportunity should be captured and considers the following initial condition. Given a non-collaborative situation, let co_c be a collabportunity between users A and B while working on an exploratory search task about the *Deepwater Horizon* oil spill. The objective of this experiment is to evaluate past collabportunities found in the different datasets—with searchers performing the same task—to estimate the feasibility of co_c.

With each data set, we created all possible pairs of users where each pair represents a past collabportunity co_p equivalent to co_c. The datasets contain complete search sessions, so simulated collaborations were generated for each co_p by aligning and combining the search sessions, which include queries, search result pages (SERPs), and content pages.

Performance Measures

Search sessions from each data set constitute baselines with which to compare the performance with the simulated collaborative processes. For the performance measures required by Algorithm 1, we use search effectiveness and search efficiency, and a procedure to perform comparisons over time in terms of the area under the curve (AUC) of the search effectiveness and search efficiency plots.

Search effectiveness (Equation 1) corresponds to the precision in terms of the fraction of information found that is useful (useful coverage) until time *t*. The purpose of this measure is to quantify the extent to which searchers are reaching useful content that helps with task completion.

$$\mathit{Search\ Effectiveness}(u,t)=\frac{\mathit{UsefulCoverage}(u,t)}{\mathit{OverallCoverage}(u,t)} \quad (1)$$

In the definition of search effectiveness presented in Equation 1, useful coverage comprises all content pages (reached through a SERP click or through post-click navigation) on which the searcher spends 30 seconds or more. A 30-second time out has been used in previous work (Fox, Karnawat, Mydland, Dumais, & White, 2005; White & Huang, 2010) to identify cases in which searchers derive utility from examined web pages. With respect to the overall coverage that appears in the denominator of Equation 1, this refers to all distinct pages visited by the user until time *t*. In an ideal scenario, this should be done with respect to coverage of relevant pages (widely used measures in IR such as precision and recall), but our approach is concerned with measuring performance from the perspective of searchers. In this sense, search effectiveness is based on implicit measures of usefulness, which, as pointed out by González-Ibáñez et al. (2012), show high overlap with explicit relevant coverage.

Although search effectiveness provides evidence of searchers' ability to find useful pages, it does not consider the effort required in that process. We consider that effort can be expressed in terms of query formulations, which can be related to underlying cognitive processes involved in the transformation or expression of internal information needs into a set of terms that a search engine can process. We formalize this relation as search efficiency (Equation 2), which indicates the relationship between search effectiveness and the effort required from the searcher to achieve that usefulness. Effort in this case is expressed in terms of the number of queries issued until time *t*. Search efficiency increases if the precision to find useful pages also increases and the number of queries used in the process remains low.

$$\mathit{Search\ Efficiency}(u,t)=\frac{\mathit{Search\ Effectiveness}(u,t)}{\mathit{NumQueries}(u,t)} \quad (2)$$

DOI: 10.1002/asi

Our goal is to identify good opportunities to collaborate, so we must compare sessions to estimate the costs and benefits of collaborating at different times. For discrete search sessions, we use the trapezoidal rule for nonuniform grids as a numerical method of integration (Equation 3) to perform comparisons in terms of the area under the curve (AUC) in a window of time.

$$\int_{t_0}^{t_c} f(t)\,dt \approx \frac{1}{2}\sum_{k=t_0}^{t_c}(t_{k+1}-t_k)\left(f(t_{k+1})+f(t_k)\right) \qquad (3)$$

AUC in this case represents how effective and efficient searchers are until time t. For a given searcher or team, the larger the AUCs for each of the two measures, the more effective or efficient the user is in finding useful pages.

Data

To perform the experiment described above, we used three data sets, (a) SU_{Lab}, logs from a laboratory study with single users' search sessions; (b) CP_{Lab}, logs from a laboratory study with collaborative pairs' search sessions; and (c) $SU_{SEngine}$, logs from the Microsoft Bing search engine with search sessions of single users. Each data set contains complete search sessions about the *Deepwater Horizon* oil spill, we were sure of the existence of past collabportunities co_p between searchers that can be used to evaluate a collabportunity co_c according to the procedure described in Algorithm 1. The selection of this topic is based on its relevance and popularity at the time when the laboratory studies were conducted (early autumn 2010). As reported by González-Ibáñez, Haseki, and Shah (2013), preliminary studies on the topic and pilot runs conducted before the laboratory studies showed considerable amounts of information about the oil spill. In the particular case of SU_{Lab} and CP_{Lab}, the participants in the study were instructed to address the following task, which was designed to be a realistic work task (Borlund, 2000):

> A leading newspaper has hired your team to create a comprehensive report on the causes, effects, and consequences of the recent Gulf oil spill. As a part of your contract, you are required to collect all the relevant information from any available online sources that you can find. To prepare this report, search and visit any website that you want and look for specific aspects as given in the guideline below. As you find useful information, highlight and save relevant snippets. Make sure you also rate a snippet to help you in ranking them based on their quality and usefulness. Later, you can use these snippets to compile your report, no longer than 200 lines, as instructed. Your report on this topic should address the following issues: description of how the oil spill took place, reactions by BP as well as various government and other agencies, impact on economy and life (people and animals) in the Gulf, attempts to fix the leaking well and to clean the waters, long-term implications, and lessons learned.

We used the three data sets independently as follows. First, the SU_{Lab} data set was used to evaluate co_c with our method using a small sample of single users' search sessions collected in a controlled study. CP_{Lab}, on the other hand, was used to evaluate co_c with a larger data set and also to investigate the potential of collabportunities compared with actual collaborative search sessions. Finally, $SU_{SEngine}$ was used to test the external validity of the approach with behavioral data collected in an uncontrolled setting (the Microsoft Bing search engine). The remainder of this section describes each of the data sets in more detail.

Single users' lab study data set: SU_{Lab}. The single users' data set comprises 11 search sessions, one session per participant, lasting for approximately 20 minutes each. Participants in this study were undergraduate students from different fields at Rutgers University. Six participants were women, and five men. Their ages ranged between 18 and 24 years (mean 20.50, *SD* 1.71). Recruitment was conducted through public announcements (e.g., e-mail lists, flyers) and a web form to register recruits' information and schedule their sessions. During the sessions, the participants were instructed to collect relevant information about different aspects of the *Deepwater Horizon* oil spill such as reactions, causes, and effects (see task description above for specifics). The logs collected during this study comprised time-stamped data containing the list of identifiers for each user, queries issued, SERPs, content pages, relevance judgments assigned to the pages encountered by the user, and pages' dwell times. Recruitment was conducted through announcements posted on campus and also delivered through e-mail distribution lists. For this study, participants were compensated with $10 for a 1-hour session, and participants also competed for additional financial rewards based on their performance in the task.

Collaborative pairs lab study data set: CP_{Lab}. The collaborative data set comprises search logs of 60 pairs or 120 participants, again with approximately 20 minutes of active searching. Likewise, in the single user lab study, participants were undergraduate students from different fields at Rutgers University. Seventy-one participants were women, 49 men. Their ages ranged between 17 and 30 years (mean 21.03, *SD* 2.44). Recruitment was conducted through public announcements (e.g., *listservs,* announcements posted on bus stops and Rutgers facilities) and a web form to register recruits' information and schedule their sessions. Pairs in this study were also instructed to collect relevant information about the *Deepwater Horizon* oil spill (see task description above). Each session log consists of time-stamped data containing unique identifiers for each team, unique identifiers for each user, queries issued, SERPs, content pages, relevance judgments, and pages' dwell time. Recruitment was conducted through announcements posted on campus and also delivered through e-mail distribution lists. A key requirement of this study was that participants enroll for the study in pairs along with someone with whom they'd had experience working together. As with

SU_{Lab}, participants were compensated with $10 for a 1-hour session. They also competed as teams for financial rewards based on task performance.

Large-scale search engine data set: $SU_{SEngine}$. We had access to search logs obtained from a large-scale search engine. Logs were gathered from Microsoft Bing for a 2-week period (from October 1, 2010, to October 14, 2010) when there was still significant interest in the *Deepwater Horizon* oil spill among web searchers. We used anonymized logs of URLs visited by users who explicitly consented to provide data through a widely distributed browser toolbar. At the time of toolbar installation, or upon first use if the toolbar was preinstalled, users were presented with a dialog requesting their consent for data sharing. Log entries included a unique user identifier, a timestamp for each page view, and the URL of the page visited. We excluded intranet and secure (https) URL visits at the source. To remove variability caused by cultural and linguistic variation in search behavior, we included only log entries generated by users in the English-speaking United States. From these logs, we extracted search sessions on Google, Yahoo!, and Bing via a methodology similar to White and Drucker (2007). Search sessions comprised queries, clicks on search results, and pages visited during navigation once users had left the search engine. Search sessions ended after a period of user inactivity exceeding 30 minutes. Logs were analyzed on premises at Microsoft by one of the authors (R. White).

We intersected the text of queries generated by the laboratory participants (described in the previous section) with those of search engine users to find sessions related to the oil spill task. Note that these sessions were not the actual sessions from the laboratory searchers, they shared queries in common only with the laboratory study. In total, 149 distinct queries (21.3%; e.g., attempts to fix the oil spill cleaning up the oil spill how fix bp oil leak bp oil spill impact on economy) matched the logged Bing queries, yielding a total of 14,934 search sessions from 12,173 users. We excluded common queries that may be unrelated to oil spill searching (e.g., crude oil greenpeace), giving us a final set of 8,969 search sessions from 8,051 users with which to assess the performance of our method.

Results

As indicated earlier, all experiments comprise an evaluation of a collabportunity co_c between two users while independently working on an exploratory search task about the *Deepwater Horizon* oil spill.

Experiment on SULAB

Logs in this data set correspond to users performing the same task, we generated all possible pairs with the 11

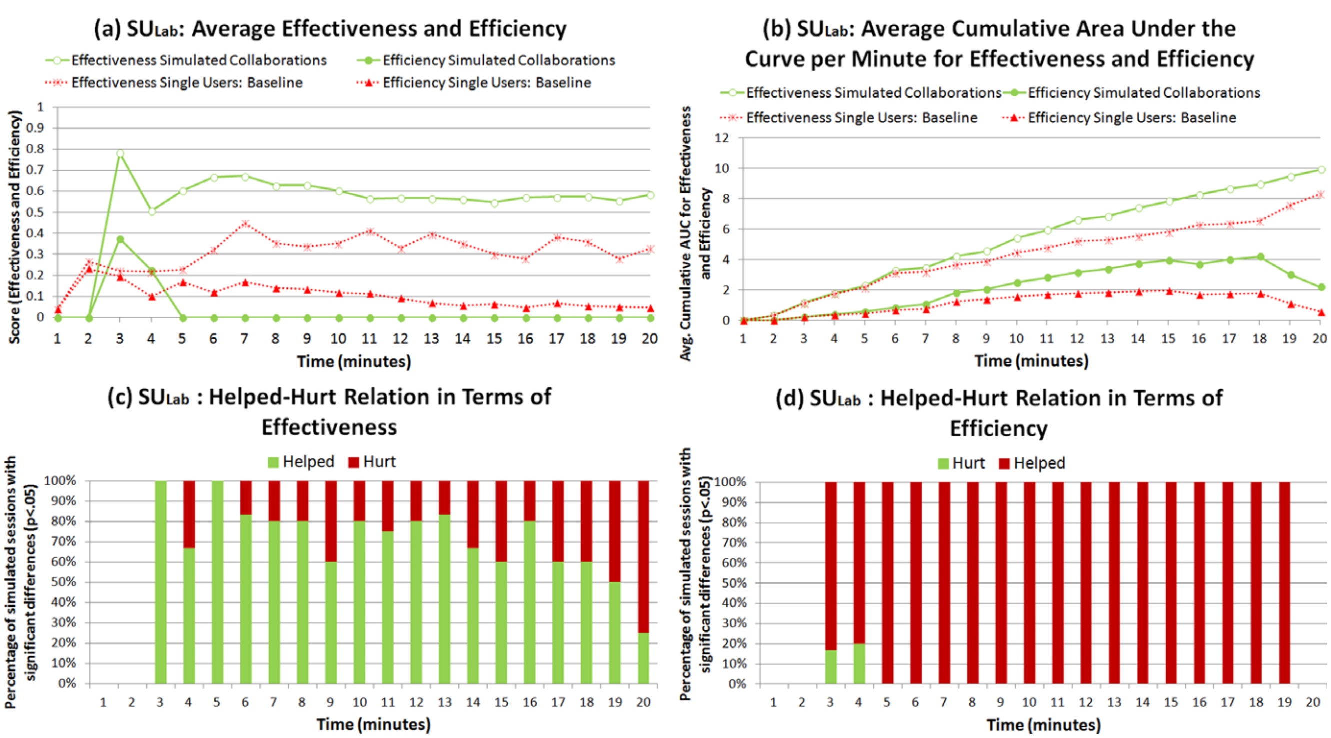

FIG. 4. Aggregated results for search effectiveness and search efficiency in each minute for the data set SU_{Lab}. **a:** Search effectiveness and search efficiency in each minute. **b:** Average of AUC for search efficiency and search effectiveness at each minute. **c:** Help–hurt relation in term of search effectiveness in each minute. **d:** Help–hurt relation in terms of search efficiency in each minute. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

DOI: 10.1002/asi

participants in the data set. Each pair (α,β) with $\alpha \in \{\approx A\}$ and $\beta \in \{\approx B\}$ represents a collabportunity co_p between users in α and β that will be employed to estimate the feasibility of co_c. Overall, we produced 55 collabportunities co_p and the corresponding simulated collaborative searches *sc*. We then computed the performance (using the measures explained above) in each minute for all co_p and compared that performance with the individual performance of the users. Figure 4a shows how simulated collaborations outperform individual search effectiveness at different minutes. The average search effectiveness of simulated pairs with reported significant differences is higher than the baseline (users in co_p working individually) most of the time. A comparison in terms of search efficiency shows that fewer than 20% of the simulated pairs obtain higher search efficiency during minutes 3 and 4. The cost of collaborating according to this data set exceeds the benefits (all of the simulated pairs showed search efficiencies significantly lower than the baseline). For individual cases, this means that a single searcher was able to cover more useful pages with the use of fewer queries (search efficiency) than in the simulated collaborations with the other 10 searchers in the data set. For instance, at the end of the session, participant P129 covered 10 useful pages with a search effectiveness of 0.63. To find such pages, P129 issued seven queries that led him to a search efficiency of 0.09. However, in a simulated collaboration with participant P141, the pair covered 14 useful pages with a search effectiveness of 0.53. This was achieved through 16 queries that led to a search efficiency of 0.03. When considering cumulative scores (Figure 4b), both search effectiveness and search efficiency significantly outperform the baselines starting in the sixth minute. Figure 4c and 4d shows the percentage of simulated sessions with significant differences (positive [helped] and negative [hurt]) at $p < 0.05$ (using z-score) for both performance measures. No bars are displayed during the first 2 minutes because no significant differences were found. For the case of search effectiveness (Figure 4c)—starting in minute 3—it was observed that more than 60% of the simulated collaborations, *sc*, displayed significant benefits. The highest peaks are found in minutes 3 and 5 (100%), only after minute 18 dropping below 50%. Conversely, search efficiency (Figure 4d) shows significant benefits for all simulated collaborations from minute 5 to minute 19. At early times (minutes 3 and 4), more than 80% of the simulated sessions display significant benefits with respect to the baseline. Overall, these results suggest that co_c is likely to improve search effectiveness and search efficiency. Pairs achieve higher exposure to useful pages than their individual members. Moreover, they do so requiring less effort expressed in the number of queries formulated.

Experiment on $\mathrm{CP}_{\mathrm{LAB}}$

In a similar manner, we ran the experiment on the CP_{Lab} data set. This data set contains logs from *pairs* of users performing a search task collaboratively, so we divided each collaborative search session and produced a total of 120 individual sessions. We then generated all possible combinations of pairs, discarding those that matched with actual pairs in the original data set. Each generated pair represents a collabportunity, co_p. Overall, we produced 7,080 co_p and the corresponding simulated collaborative searches, which provides us with a larger set of samples of co_p for estimating the feasibility of co_c. Next, we calculated search effectiveness and search efficiency in each minute for all co_p and compared them with the individual performance of their users and also with that of the original pairs in the data set. Figure 5 provides a more detailed

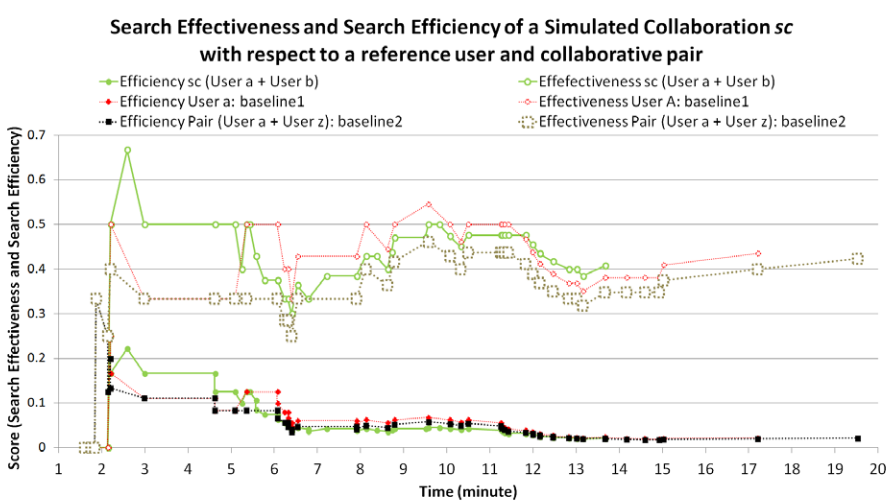

FIG. 5. Performance comparison in terms of search effectiveness and search efficiency of a simulated collaboration, *sc*, against two baselines. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

DOI: 10.1002/asi

view of the performance comparison between one simulated user pair (α_1,β_1) with $\alpha_1 \in \{\approx A\}$ and $\beta_1 \in \{\approx B\}$ and two baselines, (a) individual performance of user α_1 and (b) collaborative performance of user α_1 working with his actual collaborator as indicated in the data set (here referred to as $\beta_2 \in \{\approx B\}$). Results for both performance measures show that the simulated collaboration outperforms the two baselines approximately between minutes 2 and 5 ($p < 0.05$ using z-score). After minute 5, this gap is reduced, and slight differences are observed in terms of search effectiveness. Unlike the example provided in the previous section for a single user in SU_{Lab}, here our approach showed the potential benefits of a collabportunity in which a simulated pair (α_1,β_1) is able to reach search effectiveness. In addition, the ratio between search effectiveness and the required effort expressed in terms of the number of queries issued by the simulated pair (α_1,β_1) is lower than that for the individual user (α_1) or if working with an actual partner (β_2) during the study.

A summary of a complete evaluation and comparison with all the simulated pairs *sc* is provided in Figure 6. Results in terms of averages of search effectiveness and search efficiency in each minute show significantly higher levels of search effectiveness and search efficiency at $p < 0.05$ (using z-score) along the session, with the exception of minutes 2 and 3 (Figure 6a). Upon examining the cumulative scores in Figure 6b, we can see that the simulated collaborations have higher cumulative AUCs for both measures during the entire session. In terms of costs and benefits, Figure 6c indicates that most of the simulated pairs (80 to 100%) show significant benefit at $p < 0.05$ (again using z-score) in terms of search effectiveness until minute 9. After this point, the probability of achieving higher search effectiveness in *co* drops to 70%, although the collaboration is still much more likely to help rather than hurt. A similar trend is observed for search efficiency; significant benefits are found mainly at early stages of the search process (Figure 6d). Previous research has shown that search queries can become more specific as the session proceeds (Aula, Khan, & Guan, 2010; Radlinski & Joachims, 2005; Xiang et al., 2010), meaning that the benefit from collaboration with another searcher, who might have his own specific needs, could diminish.

It is interesting to note that simulated collaborations also outperform the search effectiveness and search efficiency of actual pairs during the entire session. This finding suggests that there may be other searchers in a given population who could help team members achieve better results by adding them to the team. In addition, by incrementing the set of *sc*, our evaluation procedure tests a wide variety of collaborative alternatives to estimate more precisely the impact of a co_c.

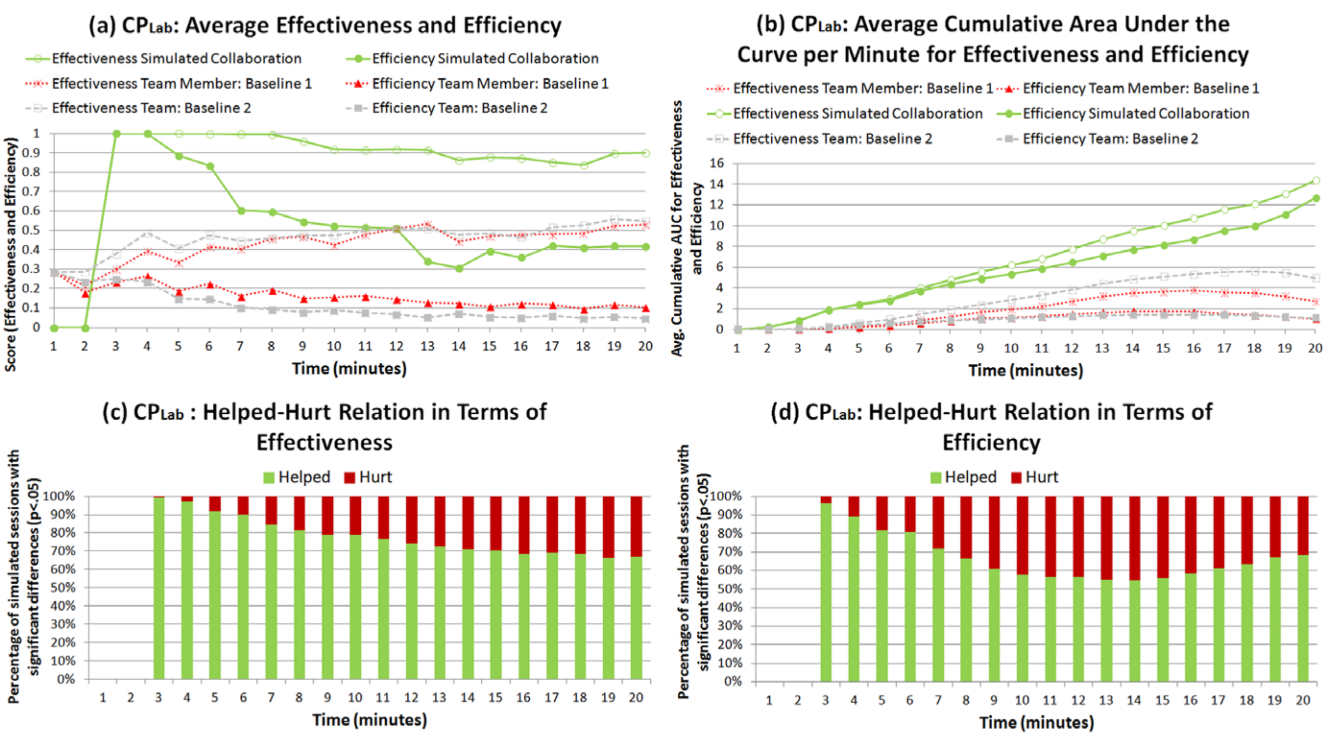

FIG. 6. Aggregated results for search effectiveness and search efficiency in each minute for the data set CP_{Lab}, **a:** Effectiveness and efficiency in each minute. **b:** Average of AUC for search efficiency and search effectiveness at each minute. **c:** Help–hurt relation in term of search effectiveness in each minute. **d:** Help–hurt relation in terms of search efficiency in each minute. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

DOI: 10.1002/asi

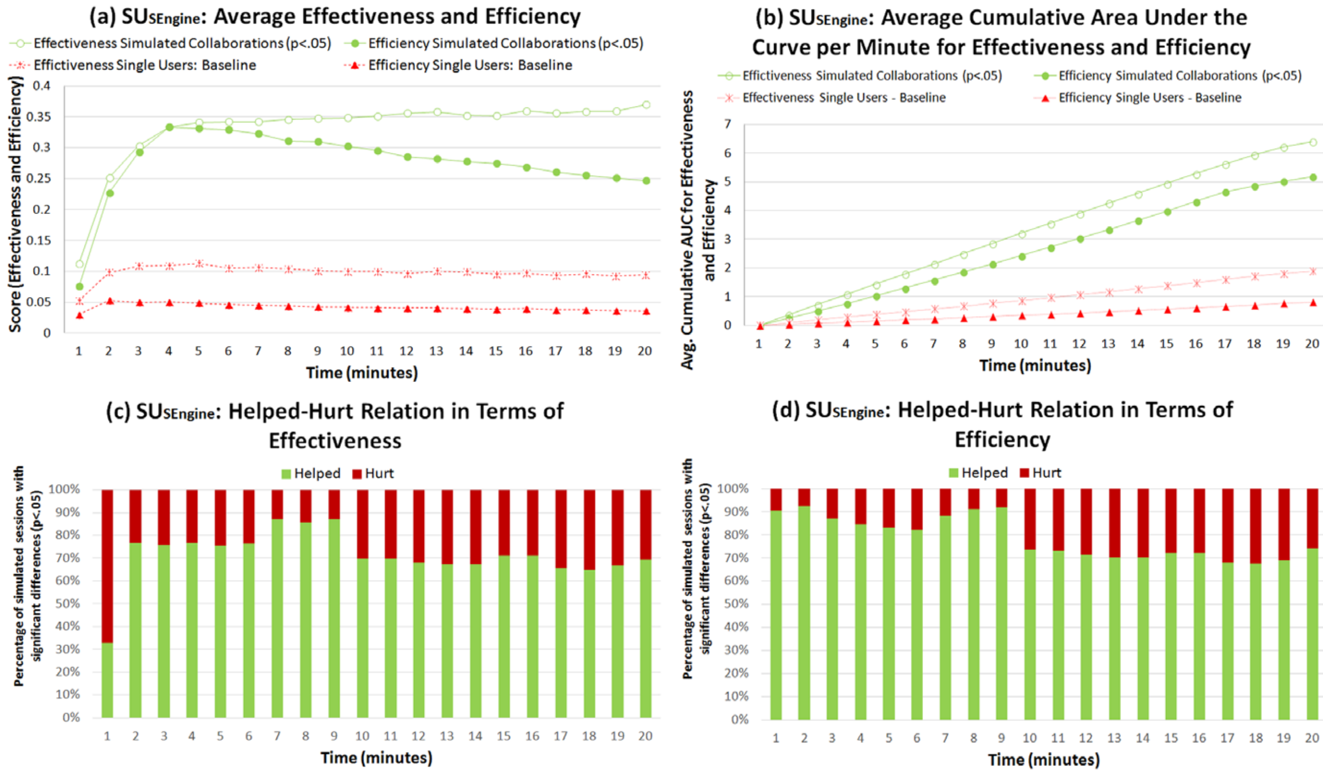

FIG. 7. Aggregated results for search effectiveness and search efficiency in each minute for the data set $SU_{SEngine}$. **a:** Effectiveness and efficiency in each minute. **b:** Average of AUC for search efficiency and search effectiveness at each minute. **c:** Help–hurt relation in term of search effectiveness in each minute. **d:** Help–hurt relation in terms of search efficiency in each minute. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Experiment on SUSENGINE

Our last evaluation consisted of running the experiment on the $SU_{SEngine}$ data set. Using this data set, we constructed teams of paired users following the same procedure used with the laboratory study data set. We could not guarantee that users in the log data were actually engaged in the same task (even though there was substantial topical similarity), so we estimated the association automatically from the log data. To do this, we matched users' sessions based on the first 2 minutes of their searching. Specifically, given a search session, *s*, we looked in our log data for other sessions {≈S} from other users with at least one query in common (exact match following trimming and lowercasing) during the first 2 minutes of that session. Each user for each session in {≈S} that met this criterion was regarded as a teammate to produce co_p and their corresponding *sc* sessions using the method described earlier. Each user was included only once in the set of candidate matches for a given reference session. If there were multiple candidate sessions from a given user, one was randomly selected. Using this method, we identified on average 1386.4 matching sessions for each session of interest (12,387,851 in total). Because the laboratory tasks were 20 minutes in duration, we focused on users' logged interactions during the first 20 minutes of each search session.

Unlike our results with the previous data sets (SU_{Lab} and CP_{Lab}), simulated pairs always outperformed the baseline in terms of search effectiveness and search efficiency (Figure 7a), with steady gains over the course of the search session (Figure 7b). Possible explanations for these differences include differences in the size and constitution of the population from which other searchers are drawn or differences in the search tasks that people are attempting, even though they are topically related. The fact that search-engine log data are inherently noisier also might have some effect on the findings. More investigation into the nature of the observed differences is required, but overall the trends in the log-based results mirror what was observed in the other two studies.

In terms of search effectiveness, our results indicate that, more than 80% of the time, simulated pairs offer significant benefits at $p < 0.05$ (using *z*-score) between minutes 7 and 8 (Figure 7c). The same evaluation indicates that the worst time to collaborate would be the first minute, for which our results indicate that approximately 30% of the simulated pairs showed more benefits than costs with respect to the baseline. For search efficiency, Figure 7d illustrates that, before minute 10, collaborating could help to increase the likelihood of finding more useful information with less effort. This situation changes from minute 10 to minute 20,

DOI: 10.1002/asi

when the number of sessions that outperform the baseline falls between 60% and 70%. As highlighted in the previous section, this variation could be attributed to searchers becoming more specific in their searches, which makes collaboration less likely to enhance individual search effectiveness and search efficiency in finding useful information for the specific search task.

Summary and Discussion

We have presented results of an experiment aimed at understanding the performance of our method using three different data sets. The primary objective of this experiment was to determine whether a collabportunity, co_c, in a noncollaborative situation was beneficial. We used different data sets to test our evaluation procedure and demonstrate potential benefits of co_c at different times. This assumes that users A and B in co_c will have search trends similar to those found in past collabportunities, co_p, in each data set.

Evaluation results obtained for each data set can be used to determine whether co_c is captured or discarded, a decision that depends on a predetermined threshold (see line 22 in Algorithm 1). For example, if the threshold is defined as 70% (which can be adjusted depending on the task and requirements), then results from the experiment with the three data sets suggest that collabportunity between users A and B is feasible between the second and the ninth minutes. Therefore, co_c should be promoted to a recommendation for collaboration during that window of time.

Our results, although different, were consistent in indicating the potential benefits of co_c in enhancing search effectiveness. On the other hand, we found that search efficiency is sensitive to the variability of the data available to simulate collaborative behavior. As explained in the previous section, both SU_{Lab} and CP_{Lab} used search sessions of a small group of users with specific characteristics (i.e., students from Rutgers University), whereas $SU_{SEngine}$ contains search sessions from thousands of users, which are assumed to have different demographic characteristics. Moreover, although we are not aware of the nature and goals of the search task performed by users in $SU_{SEngine}$, it is likely that their searches took place in natural settings and possibly with motivations beyond the scope of motivators provided in the studies (i.e., financial rewards). In this sense, not only the size of the data available but also similarities and, more important, differences among users could be key factors in the identification of collabportunities with a greater potential to become actual collaboration.

Among the results derived from this study, the experiment on CP_{Lab} was found to be especially interesting, in that it showed that our approach could outperform a baseline of real teams, in which users collaborated intentionally and explicitly. Although we observed these gains, differences and similarities between searchers might have a significant effect on the utility of this approach when employed in practice. Those in the real teams shared common ground (i.e., they were friends, couples, relatives, etc.), suggesting similarities on a range of different levels, including personality, interests, and skills. Conversely, the simulated pairs involve combinations of users who may lack common ground and may possess different characteristics. The evaluation of simulated pairs does not consider the underlying processes of explicit/intentional collaboration such as communication and coordination, which in the case of users without common ground could result in poor collaborations (Clark and Brennan, 1991; McCarthy, Miles, & Monk, 1991).

Although the method that we propose does not guarantee successful collaboration, it provides an estimation of the costs and benefits of a collaboration opportunity. Such an estimation could be utilized by a search system and is the first step in the selection of collabportunities with the potential to become an actual collaboration. Despite the promise of our method, it may be necessary to incorporate additional layers of evaluation that consider similarities and differences among searchers, aiming to maximize the likelihood of transforming collabportunities into successful collaborations.

We have shown that our method is capable of indicating when in the search process would be a good moment to collaborate. Appropriate timing in this context refers to specific periods when the benefits exceed the costs. It is important to mention that these results are specific for the task and topic evaluated. The method may report similar or different time periods depending on the task, topic, and information available about past search sessions.

We explored two performance measures that can be integrated with our method. Search effectiveness shows when a collabportunity could lead users to find information more precisely. Search efficiency quantifies expected effort at a given time to achieve a particular search effectiveness level. These measures are complementary and help to estimate benefits and costs from different perspectives. Although we limited our study to these performance measures, we note that other performance measures and features could be integrated with our approach.

Finally, we highlight the experiment on the search engine dataset $SU_{SEngine}$, which demonstrates the validity of our method when applied to data obtained from noncontrolled settings. This helps us understand the potential practical utility of our approach and highlights the potential of collabportunities in large-scale search engines.

Conclusions and Future Work

Beyond retrieving and ranking, many solutions and approaches have been developed to support the information search processes of users. Solutions based on query suggestions, results recommendation, information filtering, and personalization are some examples found in the literature. Although some of these cases, such as information filtering, consist of implicit and unintentional forms of collaboration, some researchers have indicated that collaboration is another such possibility that could enhance

DOI: 10.1002/asi

information exploration (Hyldegard, 2006, 2009; Shah, 2010b; Twidale et al., 1997). It is often the case, however, that information search is performed in solitude even when many others may be performing the same or similar tasks around the same time. Having many users (potentially many thousands, depending on the topic and the search environment) searching for information about the same topic at the same time implies latent opportunities for collaborations between searchers. These opportunities are usually missed because of the lack of methods to identify them.

We have explored these collabportunities, in particular, a way to estimate their feasibility. This raises the possibility of opportunities for collaboration being captured rather than lost over time. We have proposed a method to perform this evaluation with the assumption that underlying IR systems are capable of accessing data about past search sessions. Our approach is based on simulated collaborative search processes and performance measures that are used to determine the potential benefits and costs of collaborating with another individual. In addition, this method also indicates when good opportunities for searchers to collaborate exist.

We evaluated our method with an experimental design for three different data sets, which included data from laboratory studies and also logs from a large-scale search engine. We relied on two performance measures, namely: effectiveness and efficiency. Results derived from this study were consistent in demonstrating the potential of a given collabportunity to enhance individual performance.

The method is presented here in a generic form. We focus on an implementation in the back end of systems capable of keeping track of search sessions from multiple users over time. For example, implementations at the level of IR systems could help to identify and evaluate collabportunities in real time, by connecting different searchers with similar queries. Integration with other collaboration approaches such as algorithmic mediation for collaboration in information search (Pickens et al., 2008) is possible.

Implementing mechanisms to detect and evaluate collaboration opportunities to promote explicit and intentional collaboration between users poses several challenges beyond the technical domain. One such challenge is protecting the privacy of users. As explained, our approach relies on active monitoring of user search activity to identify potential collaborations. Moreover, it requires methods to mine past users' logs to evaluate the feasibility of current collabportunities. There are also sensitivities around connecting searchers with strangers simply based on queries issued, and user consent may need to be sought before they are entered into a pool of potential collaborators and their ongoing search activity shared with others. These aspects are beyond the scope of this study; however, we acknowledge their importance and the need for in-depth investigations that help researchers to understand the cultural feasibility, social implications, and legal limitations of implementing this kind of technology as well as the privacy-preserving mechanisms required.

Once collabportunities are properly evaluated and promoted to recommendations for collaboration, other questions must be addressed. For example, if users accept collaboration, how are they going to coordinate their actions? Will a communication channel be enabled while the collaboration takes place? Or will collaboration be based solely on information sharing? How best can these recommendations be communicated to users in an unobtrusive yet informative way? These are fundamental questions that must be answered in developing systems to lessen the number of missed collabportunities and to help searchers help each other. Responding to these questions is beyond the scope of this article. Further explorations are required to investigate the experience of users and other human factors as well as social factors when offered the possibility to collaborate with someone else.

We have evaluated our method in the context of a particular exploratory search task and topic, namely, the *Deepwater Horizon* oil spill. Although further work with more topics is needed, we believe that exploratory search scenarios of this nature are especially likely to benefit from the additional support that we have described, primarily because they may extend over many queries and involve examining resources from multiple locations. Marchionini (2006) considers broad search activities such as *learn* and *investigate* as essential parts of exploratory search, which include specific actions of discovery, knowledge acquisition, synthesis, and evaluation, among others. In this sense, our future work targets two important objectives: (a) the definition and evaluation of a method capable of identifying collabportunities in real time and (b) the refinement and testing of our current evaluation method with different tasks and scenarios (i.e., controlled and semicontrolled) to validate our findings and address questions such as those stated above.

Finally, we note that, although the experiments reported here were conducted using laboratory study data sets and web search logs, the approach itself is independent of such data. One could take the algorithm we presented and run similar analyses on a different data set to achieve similar effects. This paves the way for research and development on performing selective collaboration and bridging system-mediated and user-mediated collaborative search.

Acknowledgments

The work described in this article was in part supported by the Institute of Museum and Library Services (IMLS) Early Career Development grant # RE-04-12-0105-12.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734–749.

Agichtein, E., Brill, E., & Dumais, S. (2006). Improving web search ranking by incorporating user behavior information, In Proceedings of SIGIR'06 (pp. 19–26).

Anick, P., & Kantamneni, R.G. (2008). A longitudinal study of real-time search assistance adoption. In Proceedings of ACM SIGIR'08 (pp. 701–702).

Aula, A., Khan, R.M., & Guan, Z. (2010). How does search behavior change as search becomes more difficult? In Proceedings of SIGCHI'10 (pp. 34–44).

Belkin, N.J. (2000). Helping people find what they don't know. Communications of the ACM, 43(8), 58–61.

Bennett, P., White, R.W., Chu, W., Dumais, S., Bailey, P., Borisyuk, F., & Cui, X. (2012). Modeling the impact of short and long-term behavior on search personalization. In Proceedings of SIGIR'12 (pp. 185–194). Portland, OR: ACM Digital Library.

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. Journal of Documentation, 56(1), 71–90.

Borlund, P., & Ingwersen, P. (1999). The application of work tasks in connection with the evaluation of interactive information retrieval systems: Empirical results. In Proceedings of MIRA: Electronic Workshops in Computing (pp. 29–46). ACM Digital Library.

Clark, H.H., & Brennan, S.E. (1991). Grounding in communication. In L.B. Resnick, R.M. Levine, & S.D. Teasley (Eds.), Perspectives on socially shared cognition (pp. 127–149). The American Psychological Association.

Fidel, R., Bruce, H., Pejtersen, A.M., Dumais, S.T., Grudin, J., & Poltrock, S. (2000). Collaborative Information Retrieval (CIR). The New Review of Information Behaviour Research, 1(1), 235–247.

Fidel, R., Pejtersen, A.M., Cleal, B., & Bruce, H. (2004). A multidimensional approach to the study of human-information interaction: A case study of collaborative information retrieval. Journal of the American Society for Information Science and Technology, 55(11), 939–953. doi: 10.1002/asi.20041

Fisher, K.E., Erdelez, S., & McKechnie, L. (2005). Theories of Information Behavior. Medford, NJ: Information Today.

Fox, S., Karnawat, K., Mydland, M., Dumais, S., & White, T. (2005). Evaluating implicit measures to improve web search. ACM Transactions on Information Systems (TOIS), 23(2), 147–168.

Golovchinsky, G., Pickens, J., & Back, M. (2008). A taxonomy of collaboration in online information seeking. In Proceedings of JCDL 2008 Workshop on Collaborative Exploratory Search (pp. 1–3). Pittsburgh, PA.

Golovchinsky, G., Adcock, J., Pickens, J., Qvarfordt, P., & Back, M. (2008). Cerchiamo: A collaborative exploratory search tool. In Proceedings of CSCW'08 (pp. 1–2). San Diego.

Golovchinsky, G., Dunnigan, A., & Diriye, A. (2012). Designing a tool for exploratory information seeking. In Proceedings of ACM SIGCHI'12 (pp. 1799–1804). ACM Digital Library.

González-Ibáñez, R., Shah, C., & White, R. (2012). Pseudo-collaboration as a method to perform selective algorithmic mediation in collaborative IR systems. In Proceedings of ASIS&T'12 (pp. 1–4).

González-Ibáñez, R., Haseki, M., & Shah, C. (2013). Let's search together, but not too close! An analysis of communication and performance in collaborative information seeking. Information Processing and Management, 49(5), 1165–1179.

Horowitz, D., & Kamvar, S.D. (2010). The anatomy of a large-scale social search engine. In Proceedings of WWW'10 (pp. 431–440). ACM Digital Library.

Hyldegard, J. (2006). Collaborative information behaviour—exploring Kuhlthaus Information Search Process model in a group-based educational setting. Information Processing and Management, 42(1), 276–298.

Hyldegard, J. (2009). Beyond the search process—exploring group members information behavior in context. Information Processing and Management, 45(1), 142–158.

Joachims, T. (2002). Optimizing search engines using clickthrough data. In Proceedings of SIGKDD'02 (pp. 133–142).

Marchionini, G. (2006). Exploratory search. Communications of the ACM, 49(4), 41–46.

McCarthy, J.C., Miles, V.C., & Monk, A.F. (1991). An experimental study of common ground in text-based communication. In Proceedings of ACM SIGCHI'91 (pp. 209–215). ACM Digital Library.

Morris, M.R. (2008). A survey of collaborative web search practices. In Proceedings of ACM SIGCHI'08 (pp. 1657–1660).

Morris, M.R., & Horvitz, E. (2007). SearchTogether: An interface for collaborative Web search. In Proceedings of ACM UIST'07 (pp. 3–12). ACM Digital Library.

Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., & Back, M. (2008). Algorithmic mediation for collaborative exploratory search. In Proceedings of ACM SIGIR'08 (pp. 315–322). ACM Digital Library.

Radlinski, F., & Joachims, T. (2005). Query chains: learning to rank from implicit feedback. In Proceedings of SIGKDD'05 (pp. 239–248). ACM Digital Library.

Resnick, P., & Varian, H.R. (1997). Recommender Systems. Communications of the ACM, 40(3), 56–58.

Shah, C. (2009). Lessons and challenges for Collaborative Information Seeking (CIS) systems developers. In Proceedings of GROUP 2009 Workshop on Collaborative Information Behavior (pp. 1–4).

Shah, C. (2010a). Coagmento—A collaborative information seeking, synthesis and sense-making framework. In Integrated Demo at CSCW 2010 (pp. 1–4).

Shah, C. (2010b). Working in collaboration—What, why, and how? In Proceedings of Second workshop on Collaborative Information Seeking at CSCW 2010 (pp. 1–6).

Shah, C. (2012). Collaborative information seeking: The art and science of making the whole greater than the sum of all. The Information Retrieval Series, Vol. 34. Springer.

Shah, C., & González-Ibáñez, R. (2011). Evaluating the synergic effect of collaboration in information seeking. In Proceedings of ACM SIGIR'11 (pp. 913–922). ACM Digital Library.

Twidale, M.B., & Nichols, D.M. (1996). Collaborative browsing and visualisation of the search process. In Proceedings of ELVIRA 3, Aslib 48(7–8), 177–182.

Twidale, M.B.T., Nichols, D.M.N., & Paice, C.D. (1997). Browsing is a collaborative process. Information Processing and Management, 33(6), 761–783.

Weisz, J.D., Erickson, T., & Kellogg, W.A. (2006). Synchronous broadcast messaging: The use of ICT. In Proceedings of SIGCHI '06 (pp. 1293–1302). ACM Digital Library.

White, R.W., & Drucker, S.M. (2007). Investigating behavioral variability in web search. In Proceedings of WWW'07 (pp. 21–30). ACM Digital Library.

White, R.W., & Huang, J. (2010). Assessing the scenic route: Measuring the value of search trails in web logs. In Proceedings of ACM SIGIR'10 (pp. 587–594). ACM Digital Library.

White, R.W., Chu, W., Hassan, A., He, X., Song, Y., & Wang, H. (2013). Enhancing personalized search by mining and modeling task behavior. In Proceedings of WWW'13 (pp. 1411–1420). ACM Digital Library.

Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E., & Li, H. (2010). Context-aware ranking in Web search. In Proceedings of ACM SIGIR'10 (pp. 451–458). ACM Digital Library.

Appendix: Definitions (Based on Shah, 2012)

- **Information retrieval (IR)** is the area of study concerned with searching for documents, for information within documents, and for metadata about documents as well as searching structured storage, relational databases, and the web.
- **Information seeking (IS)** is the process or activity of attempting to obtain information in both human and technological contexts. IS, in this work, is seen as incorporating IR.
- **Information behavior/human information behavior** is the study of the interaction among people, information, and the situations (contexts) in which they interact (Fisher, Erdelez, & McKechnie, 2005).
- **Coordination** is a process of connecting different agents together for a harmonious action. This often involves bringing people or systems under an umbrella at the same time and place.

During this process, the involved agents may share resources, responsibilities, and goals.

- **Cooperation** is a relationship in which different agents with similar interests take part in planning activities, negotiating roles, and sharing resources to achieve joint goals. In addition to coordination, cooperation involves all the agents following some rules of interaction.
- **Collaboration** is a process involving various agents who may see different aspects of a problem. They engage in a process through which they can go beyond their own individual expertise and vision by constructively exploring their differences and searching for common solutions. In contrast to cooperation, collaboration involves creating a solution that is more than merely the sum of each party's contribution. The authority in such a process is vested in the collaborative rather than in an individual entity.
- **Explicit collaboration** occurs when various aspects of collaboration are clearly stated and understood. For instance, a group of students working on a science project together knows that (a) they are collaborating and (b) who is responsible for doing what.
- **Implicit collaboration** occurs when collaboration happens without explicit specifications. For instance, visitors to Amazon.com receive recommendations based on other people's searching and buying behavior without knowing those people.
- **Active collaboration** is similar to explicit collaboration, the key difference being the willingness and awareness of the user. For instance, when a user of Netflix rates a movie, he is actively playing a part in collaborating with other users. However, because he did not explicitly agree to collaborate with others, he might not even know those users.
- **Passive collaboration** is similar to implicit collaboration, the key difference being the willingness and awareness of the user. For instance, when a user visits a video on YouTube, he passively contributes to the popularity of that video, affecting the ranking and popularity of that video for others. The key difference between active and passive collaboration is user's willingness and control over the actions. In the case of active collaboration, user agrees to do it (rating, comments), whereas, in case of passive collaboration, the user has very little control (click-through, browsing patterns).

DOI: 10.1002/asi