On Domain Expertise-based Roles in Collaborative Information Retrieval

Laure Soulier, Lynda Tamine and Wahiba Bahsoun

IRIT - Paul Sabatier University
118 route de Narbonne, 31062 Toulouse Cedex 9

Email address: `soulier, tamine, wbahsoun@irit.fr` (Laure Soulier, Lynda Tamine and Wahiba Bahsoun)

On Domain Expertise-based Roles in Collaborative Information Retrieval

Laure Soulier, Lynda Tamine and Wahiba Bahsoun

IRIT - Paul Sabatier University
118 route de Narbonne, 31062 Toulouse Cedex 9

Abstract

Collaborative information retrieval involves retrieval settings in which a group of users collaborates to satisfy the same underlying need. One core issue of collaborative IR models involves either supporting collaboration with adapted tools or developing IR models for a multiple-user context and providing a ranked list of documents adapted for each collaborator. In this paper, we introduce the first document-ranking model supporting collaboration between two users characterized by roles relying on different domain expertise levels. Specifically, we propose a two-step ranking model: we first compute a document-relevance score, taking into consideration domain expertise-based roles. We introduce specificity and novelty factors into language-model smoothing, and then we assign, via an Expectation-Maximization algorithm, documents to the best-suited collaborator. Our experiments employ a simulation-based framework of collaborative information retrieval and show the significant effectiveness of our model at different search levels.

Keywords: Collaborative Information Retrieval, Domain Expertise, Ranking Model, Learning-Method

1. Introduction

An information retrieval (IR) task is generally perceived as an individual process (Catledge and Pitkow, 1995). However, the analysis of user intent

within an IR task has underlined the increasing need of collaboration to answer multifaceted (Kashyap et al., 2010) and multitopical queries (Castells et al., 2011) where result diversity is expected. Indeed, for such queries and more generally, for complex and exploratory ones (Denning and Yaholkovsky, 2008), collaboration between searchers favors the synergic effect towards the coverage of different aspects of the search results (Shah, 2012b).

Shah (2012b) defines collaboration as *"a process involving various agents that may see different aspects of a problem [...] (and) can go beyond their own individual expertise."* Either synchronous or asynchronous, CIR relies on the awareness, the division of labor and the sharing of knowledge assumptions (Foley and Smeaton, 2010; Morris and Teevan, 2009) for satisfying a *"mutual beneficial goal"* (Shah, 2012b) of different collaborators within an IR task. In this context, previous research also highlighted the large number of application domains for a collaborative search task, such as medical (Morris and Morris, 2011), academic (Moraveji et al., 2011) or political (Mascaro and Goggins, 2010) domains.

The issue of collaboration has given rise to the need of revisiting search interfaces, IR techniques and IR models that emphasize consensual document rankings (Joho et al., 2009; Shah et al., 2012). A role taxonomy (Golovchinsky et al., 2009) has been proposed assuming that users can be assigned to different tasks or goals in order to solve a shared information need. This taxonomy states different pairs of roles, such as *peers*, *domain expert/domain novice* or *prospector/minor*, where the latter has already been considered in previous CIR work (Pickens et al., 2008). More particularly, the pair of roles of *domain expert* and *domain novice*, which is addressed in this paper, is based on the assumption that collaborators have different domain expertise levels. Examples of this pair of roles can be found in four application domains: 1) the medical domain (McMullan, 2006; ECDPC, 2011) in which patients and physicians collaborate in order to find and analyze medical information using the web considering that patients dispose of much more time and motivation and can leverage from physicians' domain expertise in order to distinguish, for instance, similar symptoms; 2) the e-Discovery domain (Attfield et al., 2010; Privault et al., 2010) in which the assessment of privileged documents is performed by experts from different trades, namely lawyers, reviewers and lead counsel; 3) the librarian domain (Rudd and Rudd, 1986; Twidale et al., 1997) in which users and librarians collaborate for satisfying users' bibliographic information; and 4) the question-answering domain (Horowitz and Kamvar, 2010; White and Richardson, 2012) in which users

collaborate with the asker for solving his/her own information need. More-over, previous work surrounding user search behavior domain (Allen, 1991; Hembrooke et al., 2005; White et al., 2009) found that users with these two types of roles based on domain expertise level act differently within a search session in terms of submitted queries, used vocabulary or search success. These differences in search behavior raise the question whether a single re-trieval model is adequate in a CIR setting. Therefore, we assume that an adapted retrieval setting towards this pair of users may allow to enhance the users' knowledge throughout the collaborative search session.

With this in mind, we aim to address in this paper the issue of designing a system-mediated CIR framework that considers the difference of domain expertise levels of the collaborators. This difference runs in a wide spectrum including extremums which are namely *domain experts* and *domain novices*, described in the role taxonomy (Golovchinsky et al., 2009). To the best of our knowledge, this is the first attempt for the design of a CIR ranking model built upon domain expertise-based roles differentiated by a vertical distinction highlighting a hierarchy between domain expertise levels from the most experienced user to the less experienced one. More precisely, we propose a CIR ranking model for supporting a synchronous collaborative search between the symmetric domain expertise-based roles with respect to a shared information need, i.e., novice and expert roles. The goal of the user with the highest domain expertise level towards the query topic is to refine his/her knowledge about the query topic by focusing on specific documents, while the goal of the user with the lowest domain expertise level towards the query topic is to get a better understanding of the query topic by exploring documents with a generic vocabulary. The collaborative retrieval task is an iterative and interactive process which, at each time a collaborator selects a document, ranks a list of those documents that have not been previously selected. The document ranking takes into account the division of labor principle, as detailed in section 2.1.1, by avoiding redundancy between si-multaneous displayed document lists to both users, and the characteristics of his/her role, assuming that the most experienced user towards the query topic would assess documents as relevant if they satisfy two constraints: 1) if they are specific and 2) if they offer a novelty gain with respect to his/her domain knowledge.

A two-step collaborative document-ranking model is proposed for ranking documents according to the domain expertise-based roles. It includes a document relevance scoring and a document allocation to user roles. The

4

first step integrates document specificity and novelty according to user roles within a language model smoothing. Then, the Expectation-Maximization (EM) learning method (Dempster et al., 1977) applied on the document relevance scoring, assigns documents to the most likely suited user. Finally, we ensure that currently displayed document lists do not include the same documents. In order to evaluate our proposed model, we carry out a thorough experimental evaluation for measuring the retrieval effectiveness of our model and analyzing the impact of user roles on the ranking effectiveness.

More particularly, the underlying research questions are:

- How to adapt language model smoothing to realize user domain expertise based document relevance scores? As discussed before, search behaviors of experts are different from novices' ones, particularly in term of vocabulary technicality. Therefore, in addition to fitting the query topic, the relevance of a document also depends on the used vocabulary regarding the domain expertise level of the user. For this purpose, we propose to integrate the user domain expertise level within the smoothing parameter of a language model-based document scoring.

- How to utilize user expertise-based document relevance scores in a document ranking framework for collaborative search? Considering a CIR search session, the challenge remains on how to optimize the collaboration and satisfying users, both at an individual and collective level, in so far as they can find relevant documents with respect to the shared information need and their own domain expertise and interest levels. Therefore, we focus on determining which user could be more satisfied by a document according to his/her knowledge expertise and the query topic.

In the following section, we review previous work surrounding CIR domain to put our work in context. Section 3 presents our CIR model aiming at supporting collaboration between users, characterized by different domain expertise levels. In section 4, we focus on experiments. Section 5 discusses the proposed model, concludes the paper and introduces an overview of our future work.

2. Related Work

2.1. Collaborative Information Retrieval

2.1.1. Definition and Basic Notions

CIR has been defined as a process involving multiple users which interact with each other in order to solve a specific information need (Hansen and Järvelin, 2005). A CIR setting is characterized by two main dimensions:

- The *human activity* dimension represents collaboration as a process that encapsulates four human behavior activities, from the highest to the smallest granularity level: cooperation, coordination, contribution and communication (Shah, 2012b).

- The *spatio-temporal* dimension is presented in (Golovchinsky et al., 2009; Shah, 2012b). On the spatial side, *collocated* and *remote* collaboration are distinguished considering the spatial closeness of the users. On the temporal side, the differentiation between *synchronous* and *asynchronous* collaboration depends upon whether or not user activities take place at the same time.

Moreover, CIR is surrounded by three principles, namely the division of labor, awareness and sharing of knowledge (Morris and Teevan, 2009; Foley and Smeaton, 2010). All of these enable avoiding undesired redundant work.

The *division of labor* aims at splitting up work among users. More particularly, two main lines of approaches can be highlighted:

- A task-based approach which assigns distinct search tasks among collaborators, such as looking for diversity or analysing more in-depth document relevance (Pickens et al., 2008; Shah et al., 2010).

- A document-based approach which 1) splits the search results in order to display to users distinct document lists (Morris et al., 2008; Shah et al., 2010; Soulier et al., 2013) and/or 2) removes from document result lists those documents currently seen by collaborators (Foley and Smeaton, 2009; Soulier et al., 2013).

The *awareness* alerts users of already seen documents or previous submitted queries. Collaborative interfaces may support the awareness principle by means of shared workspace (Shah, 2012a), enabling users to be informed on the selected documents by the other collaborators, or shared interactive

tabletops (Morris et al., 2006; Smeaton et al., 2006) enabling to synchronously see other users' actions.

The *sharing of knowledge* enables the information flow among users by means of shared workspaces including annotation or bookmark facilities or adapted tools favoring brainstorming among users, such as instant messaging (Shah, 2012a; Gianoutsos and Grundy, 1996).

2.1.2. Previous Work

Authors in (Joho et al., 2009; Shah et al., 2012) identified two main lines of research to support collaboration within IR tasks. The first one, user oriented, consists of designing novel appropriate interfaces. Several interfaces have been proposed for supporting different collaborative search tasks: finding documents (Golovchinsky and Diriye, 2011; Filho et al., 2010), video (Rodriguez Perez et al., 2011) or images (Morris et al., 2006). Most of the collaborative interfaces integrate tools such as interactive tables (Morris et al., 2006), visual techniques, such as user implication representation (Erickson, 2010), or IR techniques, such as query reformulation using an evolving search experience (Morris and Horvitz, 2007).

The second line of works, more system-mediated oriented and close to our contribution, suggests revisiting traditional single-IR ranking techniques and models in the light of collaborative IR involving a group of users searching together (Foley and Smeaton, 2009; Morris et al., 2008; Pickens et al., 2008; Shah et al., 2010; Soulier et al., 2013). Among these works, we mainly distinguish two main categories depending on whether or not user roles are considered. Within the first category of works, Foley et al. (Foley and Smeaton, 2009; Foley, 2008) have proposed the implementation of the two main collaborative search policies in an IR setting: division of labor and sharing of knowledge. While the former is based on the splitting of ranked lists provided to the collaborators, the latter is expressed through a collaborative relevance feedback process for document ranking and query expansion. In order to consider users' characteristics, the authors estimate document relevance by a relevance term-weighting formula which combines the relevance statistics of each user through users' authority expressing the users' expertise towards the search task. Experiments on simulated TREC based collaborative search scenarios show the value of the division of labor policy within a collaborative search and the impact of user's authority on the search effectiveness. Nevertheless, they underline the failing impact of considering relevance feedback as a diversity factor on retrieval effectiveness.

This work is the closest one to ours presented in this paper. The main line of difference relies on the fact that we predifine user roles and ensure both collaborative retrieval effectiveness and mutual benefits to users through personalized document rankings, in contrast to Foley and Smeaton (2009) who estimate the global relevance of documents by linearly combining users' authority. Morris et al. (2008) have enriched the interface SearchTogether (Morris and Horvitz, 2007) by integrating two ranking algorithms. The *smart-splitting* algorithm combines division of labor policy with a clustering algorithm for displaying document lists to each individual collaborator. The *groupization* algorithm provides a final ranked list of documents at the collaboration group level. Both of the underlying methods are based on personalized scores estimated using a profile BM25-based weighting model. As well as work in (Morris et al., 2008; Foley and Smeaton, 2009), we ensure the division of labor by displaying to users distinct document subsets. Moreover, we reinforce this principle by splitting results among users by means of the learning Expectation-Maximisation algorithm which assigns documents to the most likely suitable users by taking into account its role within the collaborative search process.

The second category of works (Pickens et al., 2008; Shah et al., 2010; Soulier et al., 2013) consider user roles detailed in a role taxonomy for CIR (Golovchinsky et al., 2009). Pickens et al. (2008) consider two asymmetric roles of users: the *prospector* explores new fields of the information space and the *minor* ensures the richness of the fields explored. For both roles, term-weighting and document-ranking functions have been proposed using relevance and freshness scores. Experiments using the TREC-vid dataset show the synergic effect of collaborative search in comparison with result merging of individual searches. Shah et al. (2010) consider two other user roles within a collaborative search: the *gatherer* and the *surveyor*. The task of the *gatherer* is to quickly select accurate documents, whereas the surveyor focuses on providing diversity through selected documents. The authors propose to merge selected document sets provided by both users and then to assign documents to user roles according to the k-means clustering method. The latter is applied on document scores estimated using a voting function. Experiments using TREC ad hoc dataset show promising results but emphasize the need of improving the *surveyor* role by enhancing search diversity. Unlike other work which consider roles within a two users-size group, Soulier et al. (2013) propose a collaborative ranking model devoted to solving a

multi-faceted information need by a group of multiple users, viewed as experts of at least one facet of the query topic. These facets are identified by the Latent Dirichlet Allocation method. Accordingly, both documents and experts are modeled by a multi-topical representation where each element represents a facet of the query. Collaborators perform a symmetric and interactive task in which they aim at identifying relevant documents within displayed document lists. The latter are built according to a learning algorithm which assigns documents to the most likely suitable expert with respect to its relevance towards the document query facets. Experiments on the TREC interactive dataset highlight promising results whatever the size of the collaboration group. In this previous work, users are identified according to a horizontal distinction characterized by expertise levels in different sub-topics. In contrast, our proposed model considers a vertical distinction among users with a hierarchy between expertise levels. More generally, we attempt to consider the difference in the domain expertise level of the collaborators, including expert/novice users, in order to enhance the collaborative document ranking. In addition, we consider here the particular context of collaboration between a *domain expert* and a *domain novice* for any types of information needs. Accordingly, the model has been adapted in two main aspects: 1) users and documents are modeled by a term-based representation within the proposed model and 2) the document scoring with respect to each user has been tuned with respect to the characteristics of users' roles.

To the best of our knowledge, we address in this paper a new pair of domain expertise-based roles, namely *domain expert* and *domain novice*. The pair of users, assuming *domain expert* and *domain novice* roles, have fundamentally different skill sets compared to pairs of roles considered in (Pickens et al., 2008; Shah et al., 2010) and have, accordingly, different approaches and preferences while exploring information.

2.2. Collaboration Between Domain Expert and Domain Novice

Generally speaking, two users, involved within a same search session, are characterized by a relative difference of domain expertise level towards the shared information need. Collaborative models (Pickens et al., 2008; Shah et al., 2010) assume that people are different and by collaborating, they can leverage other searchers' expertise and skills in order to solve a shared information need. Therefore, within a collaborative setting, one collaborator can have more knowledge toward the topic compared to the other ones. Moreover, the domain expertise level difference between users

can be explicitly defined within an application domain. Users are also labeled as *domain experts* or *domain novices* considering the extremums of the spectrum of their domain expertise levels.

More particularly, previous work surrounding search behaviour analysis (Allen, 1991; Hembrooke et al., 2005; Hsieh-yee, 1993; Kang and Fu, 2010; White et al., 2009) have highlighted differences between *domain experts* and *domain novices*. One main difference remains on the fact that *domain experts* are more familiar with technical vocabulary (Allen, 1991; Hölscher and Strube, 2000; Kang and Fu, 2010; White et al., 2009) unlike *domain novices* who need more term suggestion for getting a better insight of the domain (Hsieh-yee, 1993; Hölscher and Strube, 2000). Accordingly, the expertise may be estimated through term specificity (Kim, 2006) but other authors assimilate the expertise level to the document reading level (Kim et al., 2012) which they integrate in a personnalisation model relying either on a language model (Collins-Thompson and Callan, 2004) or a predictive one (Collins-Thompson et al., 2011). As well as query terms are more sophisticated for experts, their queries are longer and more keyword-oriented considering their facilities to formulate a well-know information need (White et al., 2009; Kang and Fu, 2010). Similarly, search strategies and success are different considering the knowledge expertise of users (White et al., 2009; Hembrooke et al., 2005). Indeed, the knowledge level of users impacts on the perception of the information need and, therefore, the search success considering that novices can be limited for linking semantic fields around a topic which may affects the relevance of retrieved documents (Hembrooke et al., 2005).

Considering their collaboration need and their search behaviour differences, Filho et al. (2010) have already introduced the pair of *domain expert* and *domain novice* within their experimental protocol. Their collaborative interface has been evaluated by a user-study involving Linux experts and Linux novices where the former help the latter to solve an information need. In addition, the pair of *domain expert* and *domain novice* roles can be found in two main categories of collaboration:

1. The dialogical collaboration (Toomela, 2007), also known as a mutual benefit-based collaboration (Shah, 2012b) and more close to our contribution, operates on the principle that users initiate a collaboration in

order to take advantage of its synergic effect and therefore, to get mutual benefits within the solving process of the information need. The whole set of users are active and complementary within the collaboration process. Below, we discuss two application domain examples of this type of collaboration.

- The medical domain in which the main challenge remains on finding relevant information about diseases or medication process (Morris and Morris, 2011). Wald et al. (2007) introduces the notion of "triangulation of patient-Web-physician" in order to characterize the relationships between this pair of users. Patients become active consumers of medical information (ECDPC, 2011) and prepare consultation (Attfield et al., 2006; Fox, 2008) whereas physicians are perceived as information searcher helpers. This statement is highlighted by a survey (Podichetty et al., 2006) in which 80% of healthcare professionals states that at least one of their patients during their career presents printed information extracted from the web at a consultation. Moreover, developing electronical interaction with patients is also a way to build trust-based relationships with new or prospective patients (Erdem and Harrison-Walker, 2006). In summary, McMullan (2006) asserts that patients and physicians collaborate by finding and analysing information from the web.

- The e-Discovery domain refers to the management of electronic data in order to use them in the case of a civil or criminal litigation and government inspection (Conrad, 2007). The complexities of data management, involving the review, the selection, the circulation of privileged materials requires such different responsibilities, skills and qualifications that a collaboration is needed (Attfield et al., 2010). Two collaboration scenarios can be considered:

  (a) The pair lead counsel-contract reviewers for which the main task of Lead counsel, viewed as experts, and contract reviewers, viewed as novices in the e-Discovery domain is to identify collaboratively privileged documents. The relevance judgement of these documents is not obvious and needs collaboration (Wang and Soergel, 2010).

  (b) The pair reviewers-lawyers, where the task is similar to the

first pair. Lawyers, viewed as novices, can benefit from reviewers' experience towards subtleties of keywords search tools (Privault et al., 2010).

2. The unidirectional collaboration (Toomela, 2007) operates on the principle that the information need is initiated by only one user which gets the benefits of other collaborators' knowledge in order to solve his/her information need. In most of the cases, users who ask for collaboration are information consumers, whereas the other collaborators are information producers, as shown in the two following examples of application domains.

- The librarian domain in which users, viewed as domain novices, are looking for bibliographic information and may ask for advice to information experts, namely librarians (Kuhlthau et al., 1992; Taylor, 1968; Twidale and Nichols, 1996; Wielhorski, 1994). In this context, some studies have shown that supporting collaboration between users and librarians by means of information systems avoids overload for both roles. On one hand, it enables users to refine their information need and reduce the amount of time spent for solving their information need. On the other hand, librarians, which take the ownership of users' information need, can leverage users' skills by delegating them some information search or analysis (Rudd and Rudd, 1986). Besides helping users to find relevant information, librarians enable users through their advice to develop their search skills in the bibliographic domain (Wielhorski, 1994). Moreover, Twidale and Nichols (1996) emphasized the technical vocabulary gap between these two collaborators and have proposed the *Adriane* collaborative interface devoted to supporting interactions between these two types of users and promote awareness within the search session.

- The question-answering domain in which the answerer helps the asker to solve his information need (White and Richardson, 2012). One of the main underlying challenges in this domain is the expert finding task which consists into identifying the most appropriate answerer, viewed as an expert, in order to give a high quality information for solving the question of another user, which can be viewed as a novice (Horowitz and Kamvar, 2010; White et al.,

2011; White and Richardson, 2012). For this purpose, adapted interfaces (Horowitz and Kamvar, 2010) or expert finding models (Smirnova and Balog, 2011; White et al., 2011) have been proposed in order to match the asker with the suited domain experts with respect to his information need. Another interesting aspect of the question-answering domain relies on communications between experts and novices. Isaacs and Clark (1987) state that these users involved within a collaborative search session need to adjust their understanding of the information need. In the same way, this clarification need among these users is shown by the fact that the amount of conversations is positively correlated with the domain expertise difference between both users (Isaacs and Clark, 1987; White and Richardson, 2012).

3. The Model

In this section, we describe our CIR model based on the two user roles of *domain expert* and *domain novice*. We first introduce the retrieval setting and then detail the ranking model.

3.1. The retrieval Setting

3.1.1. Framework

We focus, here, on the retrieval aspect aiming at supporting a synchronous collaborative search setting between users with symmetric roles, namely *domain expert* and *domain novice*. The aim is to provide different documents to each user in order that both of them get mutual benefits from the collaboration. As shown in Figure 1, the retrieval task involves a pair of users u_j and $u_{j'}$ who share the same information need, modeled by a query q. We assume that users have distinct domain expertise levels, and, for convenience, we characterize each user u_j by a distinct role $\mathcal{R}(u_j) \in \mathcal{R}$ with $\mathcal{R} = \{$domain expert, domain novice$\}$, representative of his/her knowledge expertise. The *search session* S is launched by the submission of a shared information need, modeled by a query q, and ends when the collaborators do not select any more documents. At the initialization step S_{init}, the model provides for each user $u_j \in U$ a different initial ranked list $l^{init}(u_j)$ of documents with respect to the query topic q and his/her role. Afterwards, search session S consists of *iterative search steps* S_{iter}: each search iteration $k \in S_{iter}$ is launched by the user's feedback through selection of document

d_i. Accordingly, a ranked list $l^k(u_j, \mathcal{D}_{ns}^k)$ of documents that were not previously selected $\mathcal{D}_{ns}^k$ at time-stamp k is determined and displayed for user u_j according to his/her role $\mathcal{R}(u_j)$ and the query topic q. With respect to the division of labor principle, intermediate ranked lists provided at the different search iterations are split, ensuring empty intersections.
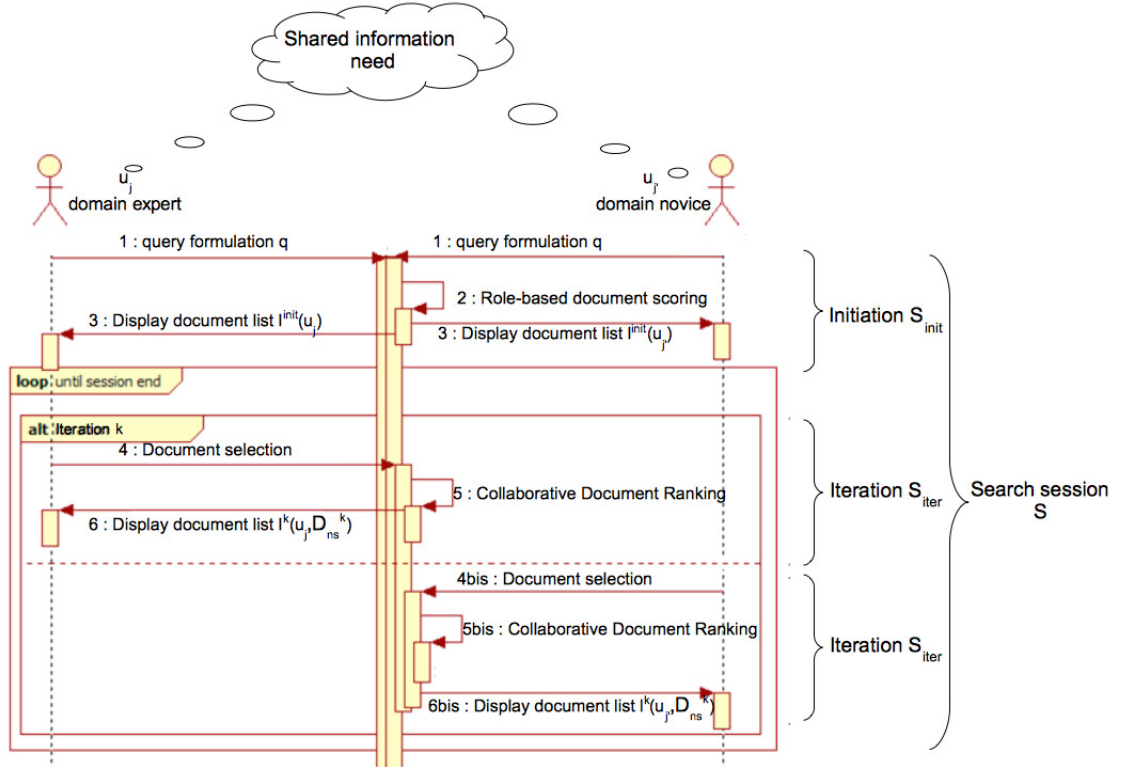

Figure 1: Our collaborative information retrieval framework

3.1.2. Division of Labor

Our model also ensures the division of labor principle among collaborators according to three main aspects:

(a) Documents are allocated according to an Expectation Maximization-based method which assigns documents to users according to the fact that one user is more likely suitable than the other for assessing the document relevance.

(b) The intersection of document lists currently displayed to the whole set of collaborators is empty.

(c) Documents already seen by at least one member of the collaboration group are not considered for the collaborative document ranking.

3.1.3. User Modeling

We formalize users by two main components, namely their roles and their profiles, respectively connected to their relative domain expertise level and their own domain expertise level towards the information need.

User role. We introduce and formalize two symmetric user roles, namely *domain expert* and *domain novice*, based on the relative knowledge expertise of users. One can define domain expertise as experience and skills that have been acquired by the user through past works. Characteristics of user roles are based on the following assumptions:

- Experts use a more specific vocabulary and an appropriate terminology during a search session (Hölscher and Strube, 2000; Vakkari et al., 2003; White et al., 2009).

- Previous work adopted search strategies applied on non-experts, namely novices, for educating them to detect relevant documents, such as query suggestion, and allowing to gain domain knowledge on the query topic (Hsieh-yee, 1993; Hölscher and Strube, 2000; White et al., 2009).

- One challenge of document relevance is to take into consideration its novelty in addition to it similarity regarding the query topic (Harman, 2002; Soboroff and Harman, 2005). The novelty need is even more important within an iterative search process in order to avoid retrieving documents similar to already selected ones.

Accordingly, *domain experts* represent problems at deep structural levels and are generally interested in discovering new associations among different aspects of items, or in delineating the advances in a research focus surrounding the query topic; this led to emphasize both novelty and specificity in the information space. In contrast, *domain novices* represent problems in terms of surface or superficial aspects and are generally interested in enhancing their learning about the general query topic; this led to emphasize mainly novelty.

Considering these statements, user roles are connected to two different measures: document novelty and document specificity, described in what follows.

- **Document novelty level** $\mathcal{L}_n(d_i, \mathcal{D}(u_j)^k)$ estimates the marginal gain of knowledge provided by a selected document d_i by user u_j according to the already selected document set $\mathcal{D}(u_j)^k$ by the same user at time-stamp k. We use the "Distance-based Item Novelty" (Castells et al., 2011) applied to the selected document set $\mathcal{D}(u_j)^k$ for user u_j. The novelty level $\mathcal{L}_n(d_i, \mathcal{D}(u_j)^k)$ of document d_i given a previously selected document set $\mathcal{D}(u_j)^k$ by user u_j at time-stamp k is estimated as follows:

$$\mathcal{L}_n(d_i, \mathcal{D}(u_j)^k) = min_{d_{i'} \in \mathcal{D}(u_j)^k} d(d_i, d_{i'}) \qquad (1)$$

  where distance d depends on the Jaccard similarity function $sim(d_i, d_{i'})$ between two documents d_i and $d_{i'}$ and is estimated by $d(d_i, d_{i'}) = 1 - sim(d_i, d_{i'})$.

- **Document specificity level** $\mathcal{L}_s(d_i)$ estimates the level of description of terms used in document d_i. We assume that a term is specific if its frequency in the collection C is low. For this purpose, the specificity level $\mathcal{L}_s(d_i)$ of document d_i is estimated using a specificity indicator (Kim, 2006) applied on unique terms. This latter, noted $Pspec$, is based on the inverse document frequency of term t in the whole collection, computed as follows:

$$\mathcal{L}_s(d_i) = avg_{t \in d_i} \; Pspec(t) = avg_{t \in d_i} \; (-log(\frac{df_t}{N})) \qquad (2)$$

  where df_t is the number of documents including term t, N is the collection size.

User profile. In this work, the user profile is associated with his/her expertise level modeled by taking into account the relevance of documents he/she selects while assuming his/her role within the search task. These profiles can then be used to rank and split the lists of candidate documents in response to the shared query. With this in mind, a user profile $\pi(u_j)^k$ is assigned to each user u_j depending on his/her domain expertise, based on term-frequency modeling. Considering our model setting, the knowledge profile $\pi(u_j)^k$ of user u_j is automatically updated within each iteration k. The domain expertise for user u_j is extracted from the collaborative query q and its selected

documents $\mathcal{D}(u_j)^k$ from the beginning of the search session until iteration search k. For a particular term t_v within user profile $\pi(u_j)^k$, the weight is equal to the average of its weight within query q and each selected document $d_{i'} \in \mathcal{D}(u_j)^k$. The user profile is modeled as follows:

$$\pi(u_j)^k = \{(t_1, w_{1j}^k), \ldots, (t_v, w_{vj}^k), \ldots, (t_z, w_{zj}^k)\} \tag{3}$$

where z is the total number of terms included in both query q and selected documents $\mathcal{D}(u_j)^k$ at time-stamp k by user u_j. The weight w_{vj}^k of term t_v for user profile $\pi(u_j)^k$ is estimated as follows:

$$w_{vj}^k = \frac{w_{vq} + \sum_{d_{i'} \in \mathcal{D}(u_j)^k} w_{vi'}}{1 + \sum_{d_{i'} \in \mathcal{D}(u_j)^k} |d_{i'}|} \tag{4}$$

where w_{vq} and $w_{vi'}$ denote respectively the weight of term t_v in query q and document $d_{i'}$. Term weights are estimated using the tf*idf scores. The number of terms included in document $d_{i'}$ is noted $|d_{i'}|$. For convenience, we also model the query q by vector $q = \{(t_1, w_{1q}), \ldots, (t_v, w_{vq}), \ldots, (t_z, w_{zq})\}$. We note that at the initialisation of the search session, only query q is considered as evidence source of the user profile.

3.2. Collaborative Document Ranking over an Iteration

The collaborative document-ranking model over search iterations includes two main steps, described in Figure 2:

- *Step 1:* A role-based document scoring in order to estimate the document relevance probability for both users given their respective roles.

- *Step 2:* A user document allocation in order to assign to each document the most likely suited collaborator according to his/her knowledge expertise modeled through his/her profile.

3.2.1. Role-based Document Scoring

Considering our CIR model framework, the document relevance depends upon the query topic and the users' characteristics. First, the document relevance probability $P^k(d_i|u_j)$ of document d_i given user u_j at iteration k is developed using conditional probabilities, as shown in Equation 5. Then, we estimate this probability assuming that the probability $P^k(u_j|q)$ is not
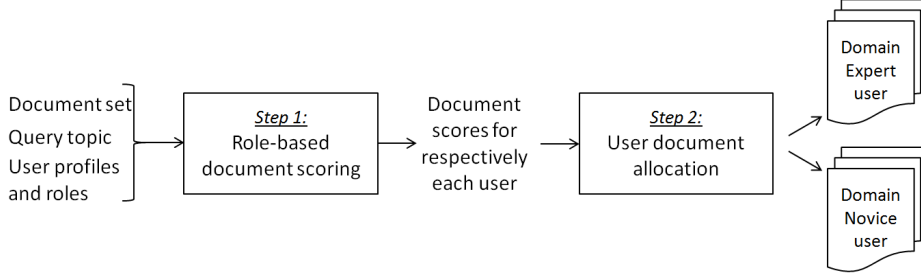
17

Figure 2: Steps of our CIR model based on *domain expert* and *domain novice*

discriminant for estimating the document relevance (Equation 6) and user u_j is independant to query q (Equation 7):

$$P^k(d_i|u_j, q) \quad = \frac{P^k(u_j|d_i,q) \cdot P^k(d_i|q)}{P^k(u_j|q)} \tag{5}$$

$$\propto P^k(u_j|d_i, q) \cdot P^k(d_i|q) \tag{6}$$

$$\propto P^k(u_j|d_i) \cdot P^k(d_i|q) \tag{7}$$

On the one hand, assuming that the document scoring with respect to the query topic is invariant regardless of the search iteration and the probability $P(d_i)$ is not discriminant for measuring the probability $P^k(d_i|q)$, the probability $P^k(d_i|q)$ can be rewritten as follows:

$$P^k(d_i|q) = P(q|d_i) \cdot P(d_i)$$
$$\propto P(q|d_i) \tag{8}$$

In order to estimate the probability $P(q|d_i)$, we assume that document d_i is represented by a multinomial distribution over terms θ_{d_i}. Therefore, the probability $P(q|d_i)$ corresponds to the probability $P(q|\theta_{d_i})$ and is estimated using the Jelinek-Mercer approach (Jelinek and Mercer, 1980):

$$P(q|\theta_{d_i}) = \prod_{(t_v, w_{vq}) \in q} [\lambda P(t_v|\theta_{d_i}) + (1 - \lambda)P(t_v|\theta_C)]^{w_{vq}} \tag{9}$$
$$with \ P(t_v|\theta_{d_i}) = \frac{tf(t_v, d_i)}{|d_i|}$$
$$and \ P(t_v|\theta_C) = \frac{tf(t_v, C)}{|C|}$$

18

where $tf(t_v, d_i)$ and $tf(t_v, C)$ are respectively the frequency of term t_v in document d_i and document collection C. The number of terms within document d_i and document collection C is noted $|d_i|$ and $|C|$ respectively. The two probabilities $P(t_v|\theta_{d_i})$ and $P(t_v|\theta_C)$ are combined using a tuning parameter λ.

On the other hand, the probability $P^k(u_j|d_i)$ is estimated using the language-based modelling by the probability $P(\pi(u_j)^k|\theta_{d_i})$ where user u_j is estimated by its term distribution over its profile $\pi(u_j^k)$ at iteration k. The latter, namely the probability of the user profile at iteration k according to the language model θ_{d_i} of document d_i, is computed using the Jelinek-Mercer approach (Jelinek and Mercer, 1980):

$$P^k(u_j|d_i) \propto P(\pi(u_j)^k|\theta_{d_i})$$
$$\propto \prod_{(t_v, w_{vj}^k) \in \pi(u_j)^k} [\lambda_{ij}^k P(t_v|\theta_{d_i}) + (1 - \lambda_{ij}^k) P(t_v|\theta_C)]^{w_{vj}^k} \quad (10)$$

where the tuning parameter λ in Equation 9 is replaced with λ_{ij}^k, a role-based smoothing parameter which depends on document d_i and user role $\mathcal{R}(u_j)$ at iteration k. It is estimated dynamically according to the role $\mathcal{R}(u_j)$ of user u_j and the assumptions described in section 3.1.3. Therefore, the tuning parameter λ_{ij}^k depends on the specificity of document d_i and its novelty level according to knowledge profile $\pi(u_j)^k$ of user u_j at iteration k.

For the *expert user* u_j and a given document d_i, the more specific and the newer the document d_i is according to its domain expertise, the larger λ_{ij}. Thus, we estimate λ_{ij} as follows:

$$\lambda_{ij}^k = \frac{\mathcal{L}_n(d_i, \mathcal{D}(u_j)^k) \cdot \mathcal{L}_s(d_i)}{\max_{d_{i'} \in \mathcal{D}} \mathcal{L}_n(d_{i'}, \mathcal{D}(u_j)^k) \cdot \mathcal{L}_s(d_{i'})} \quad (11)$$

where $\mathcal{D}$ expresses the document dataset.

For the *novice user* u_j, a given document d_i and an iteration k, we estimate λ_{ij}^k differently. The difference depends on the smoothing parameter λ_{ij}^k that integrates the fact that a novice user considers more relevant documents, those which have not a specific vocabulary. The less specific and the newer the document d_i is according to its domain expertise, the larger λ_{ij}^k.

Therefore, we propose to estimate λ_{ij}^k as follows:

$$\lambda_{ij}^k = \frac{\mathcal{L}_n(d_i, \mathcal{D}(u_j)^k) \cdot \mathcal{L}_s(d_i)^{-1}}{\max_{d_{i'} \in \mathcal{D}} \mathcal{L}_n(d_{i'}, \mathcal{D}(u_j)^k) \cdot \mathcal{L}_s(d_{i'})^{-1}} \quad (12)$$

3.2.2. Document Allocation to User Roles based on the Expectation-Maximization Algorithm

Here, we aim at optimizing the document relevance function over role-based scores, computed in the previous step using Equation 7. The optimized role-based document scores are used for ranking and classifying not already selected documents to the most likely suited collaborator considering his/her domain knowledge. For this purpose, we use the Expectation Maximization algorithm (Dempster et al., 1977; Marsland, 2009) which is an iterative one based on two steps. The first one, called *E-Step*, estimates probability $P(R_j = Rel|x_{ij}^k)$ of the relevance of document d_i given the score x_{ij}^k for iteration k. The latter corresponds to the document relevance probability according to the role of user u_j and the query q, presented in equation 7. The probability $P(R_j = Rel|x_{ij}^k)$ is estimated using a mixture model that considers, on the one hand, a gaussian probability law to model the relevance of the document for the user role and, on the other hand, an exponential probability law to model the non-relevance of the same document for the user role. The second step, called *M-Step*, updates the parameters of the mixture model by maximizing the likelihood of the probability. Notations are described in Table 1 and Algorithm 1 presents a general overview of our document allocation method. The latter is detailed in what follows.

1. Learning the document-user mapping.
We aim, here, at learning how users are likely to assess the relevance of a document through an EM-based algorithm, involving two steps:

- *The E-step.* The probability $P(R_j = Rel|x_{ij}^k)$ of the document relevance for user u_j considering the relevance of document d_i given user u_j, can be written differently using the Bayes' theorem.

$$
\begin{aligned}
P(R_j = Rel|x_{ij}^k) &= \frac{P(x_{ij}^k|R_j=Rel) \cdot P(R_j=Rel)}{P(x_{ij}^k)} \quad (13)\\[2mm]
&= \frac{P(x_{ij}^k|R_j=Rel) \cdot P(R_j=Rel)}{P(x_{ij}^k|R_j=Rel) \cdot P(R_j=Rel) + P(x_{ij}^k|R_j=NRel) \cdot P(R_j=NRel)}
\end{aligned}
$$

$X_j^k = \{x_{1j}^k, \ldots, x_{ij}^k, \ldots, x_{nsj}^k\}$	The score of non-selected documents $\mathcal{D}_{ns}$ where each element x_{ij} is estimated by equation 7 depending on user role $\mathcal{R}(u_j)$ and document d_i at iteration k.
X^k	The score matrix of non-selected documents $\mathcal{D}_{ns}$ according to user roles $\mathcal{R}$ at iteration k.
$R_j = \{Rel, NRel\}$	The hidden variable that refers to the relevance or irrelevance of a document for user u_j.
ϕ_j^k	The Gaussian probability density function of relevant documents respectively according to user u_j at iteration k.
ψ_j^k	The Exponential probability density function of irrelevant documents respectively according to user u_j at iteration k.
α_j^k	The weighting coefficient within the mixture model.
θ_j^k	The parameters for the score distribution at iteration k related to user u_j that correspond to the ϕ_j^k Gaussian one, μ_j^k and σ_j^k, and to the ψ_j^k Exponential one, λ_j^k.

Table 1: Notations used in Algorithm 1

Finally, the probability $P(R_j = Rel|x_{ij}^k)$ can be rewritten using the mixture model of Gaussian and Exponential ditributions as follows:

$$P(R_j = Rel|x_{ij}^k) \quad \propto \frac{\alpha_j^k \cdot \phi_j^k(x_{ij}^k)}{\alpha_j^k \cdot \phi_j^k(x_{ij}^k) + (1-\alpha_j^k) \cdot \psi_j^k(x_{ij}^k)} \tag{14}$$

- *The M-step.* This step updates the parameters θ_j^k and define the algorithm convergence. For this purpose, we define the Expected Complete Data Likelihood $\mathcal{L}(R_j = Rel|X_j^k, \theta_j^k)$ as follows:

$$\mathcal{L}(R_j = Rel|X_j^k, \theta_j^k) = \sum_{h=1}^{n} \sum_{j=1}^{2} log(P(x_{ij}^k, R_j = Rel|\theta_j^k))P(R_j = Rel|x_{ij}^k)$$

$$\tag{15}$$

where $P(x_{ij}^k, R_j = Rel|\theta_j^k) = \alpha_j^k \cdot \phi_j^k(x_{ij}^k)$.

Algorithm 1: EM-based collaborative document ranking

Data: $\mathcal{D}_{ns}^k$, U, u_j, X^k

Result: $l^k(u_j, \mathcal{D}_{ns}^k)$

begin

    /* Stage 1: Learning the document-users mapping   */

    while *nonconvergence* **do**

        $\delta_j^k = \varnothing$

        $\delta_{j'}^k = \varnothing$

        /* E-step   */

        forall the *documents* $d_i \in \mathcal{D}_{ns}$ **do**

            forall the *users* $u_j \in U$ **do**

$$P(R_j = Rel|x_{ij}^k) = \frac{\alpha_j^k \cdot \phi_j^k(x_{ij}^k)}{\alpha_j^k \cdot \phi_j^k(x_{ij}^k) + (1 - \alpha_j^k) \cdot \psi_j^k(x_{ij}^k)}$$

$$\delta_j^k = \delta_j^k \cup (d_i, P(R_j = Rel|x_{ij}^k))$$

        /* M-step   */

        forall the *users* $u_j \in U$ **do**

$$S_j^k = \sum_{h=1}^{n} P(R_j = Rel|x_{hj}^k)$$

$$\alpha_j^k = \frac{1}{n} S_j$$

$$\mu_j^k = \frac{1}{S_j} \sum_{h=1}^{n} P(R_j = Rel|x_{hj}^k) \cdot x_{hj}^k$$

$$\sigma_j^k = \frac{1}{S_j} \sum_{h=1}^{n} P(R_j = Rel|x_{hj}^k) \cdot (x_{hj}^k - \mu_j^k)^2$$

$$\lambda_j^k = \frac{\sum_{h=1}^{n}(1 - P(R_j = Rel|x_{hj}^k))}{\sum_{h=1}^{n}(1 - P(R_j = Rel|x_{hj}^k)) \cdot x_{hj}^k}$$

    /* Stage 2: Allocating documents to users   */

    forall the *documents* $d_i \in \mathcal{D}_{ns}$ **do**

        if $r_{jj'}(d_i, \delta_j^k, \delta_{j'}^k)^k = 1$ **then**

$$l^k(u_j, \mathcal{D}_{ns}^k) = l^k(u_j, \mathcal{D}_{ns}^k) \cup (d_i, P(R_j = Rel|x_{ij}^k))$$

    Return $l^k(u_j, \mathcal{D}_{ns}^k)$

We build for each user u_j a ranked list of documents δ_j^k according to the probability $P(R_j = Rel|x_{ij}^k)$.

2. Allocating documents to users

The objective, here, is to determine which user is the most likely to assess the relevance of a document. For this purpose, we use each user's document

list δ_j^k ranked according to the EM-based probabilities $P(R_j = Rel|x_{ij}^k)$, and more particularly, we focus on the rank of each document within both users' lists δ_j^k and $\delta_{j'}^k$. Intuitively, we assume that if the rank of a document is higher within document list δ_j^k of user u_j than within the other user $u_{j'}$ document list $\delta_{j'}^k$, this document should be more likely assessed by user u_j. Thus, we model the rank-based allocation function $r_{jj'}^k(d_i, \delta_j^k, \delta_{j'}^k)$ of document d_i as follows:

$$r_{jj'}^k(d_i, \delta_j^k, \delta_{j'}^k) = \begin{cases} 1 & if \ rank(d_i, \delta_j^k) < rank(d_i, \delta_{j'}^k) \\ 0 & otherwise \end{cases} \tag{16}$$

where $rank(d_i, \delta_j^k)$ and $rank(d_i, \delta_{j'}^k)$ represent the rank of document d_i respectively within ranked document list δ_j^k and δ_j^k. Therefore, if $r_{jj'}^k(d_i, \delta_j^k, \delta_{j'}^k) = 1$, document d_i is assigned to user u_j; otherwise, it is assigned to user $u_{j'}$. Finally, the model allows assigning the list $l^k(u_j, \mathcal{D}_{ns}^k)$ to the user u_j which has selected a document, launching iteration k.

Moreover, in order to reinforce the division of labor principle of a collaborative IR task (Foley and Smeaton, 2010), document d_i is included in the ranked list $l^k(u_j, \mathcal{D}_{ns}^k)$ of a user under the constraint it is not currently presented to the other collaborator $u_{j'}$ document list $l^k(u_{j'}, \mathcal{D}_{ns}^k)$. This avoids, according to the division of labor paradigm, the redundancy of the search effort task, ensuring that a document is seen/selected by only one user.

4. Experimental Evaluation

In CIR domain, official datasets with collaborative search log do not exist, except proprietary ones (Morris et al., 2006; Shah, 2012a). To tackle this lack, Foley and Smeaton (2009) have proposed an experimental framework which simulates collaboration through individual search logs. Simulations present several advantages compared to user studies, mainly different scenarios or parameter tuning can be performed with less costs and less time consuming (White et al., 2005), although at the cost of being able to generalize findings beyond the context of the simulation. As our goal is to develop a model that assigns user expertise-based relevance scores and not (yet) to assess its support in a naturalistic setting, we follow this framework in order to assess the validity of our model using a standard IR collection issued from

TREC[1] evaluation campaign. The objectives of our experimental evaluation are threefold:

1. Analyzing the retrieval effectiveness at the search session level.

2. Analyzing the retrieval effectiveness at the users' role level.

3. Analyzing the impact of the characteristics of the collaborator groups on the retrieval effectiveness.

In what follows, we describe the experimental setup and analyze the obtained results.

4.1. Experimental Setup

4.1.1. Dataset

In our experiments, we rely on the TREC 6-7-8 Interactive Track, previously used within the simulation-based framework proposed in (Foley and Smeaton, 2009). One of the goals of users who perform this track is to identify several instances, namely aspects, related to the information need (Over, 2001). The used document dataset is the TREC Financial Times of London 1991-1994 Collection, which is stored in Disk 4 of the TREC ad hoc collection. The collection includes 210,158 articles with an average number of terms of 316 per article. Table 2 introduces the statistics of the TREC 6-7-8 Interactive dataset. From the 20 TREC topics, 197 user queries have been extracted through runs of participants. After building pairs of collaborators according to four settings, as described in section 4.1.4, we obtain from 81 to 243 collaborative queries for testing the retrieval effectiveness of our model. TREC Interactive participants, namely universities, provide two types of output sets entitled "sparse format data" and "rich format data." The first

Number of documents	210 158
Number of TREC topics	20
Number of user queries	197
Number of collaborative queries	up to 243

Table 2: TREC Interactive dataset statistics

---

[1]http://trec.nist.gov

one encompasses a list of selected documents, whereas the second one proposes additional details such as the time-stamp, the submitted queries, and seen documents. For our experiments, we extracted individual meta-data for each individual user from all "rich format data." For this aim, we evaluate 7 participants, including the query, the list of documents selected by the user and their respective time-stamp label selection. The latter expresses how many seconds after the beginning of the search session this document has been selected. The evaluated participants are: Berkeley TREC 6 (bkl6), rmitMG TREC 6 (rmitMG6), rmitZ TREC 6 (rmitZ6), Berkeley TREC 7 (bkl7), Toronto A TREC 7 (torontoA7), Toronto B TREC 7 (torontoB7) and Berkeley TREC 8 (bkl8). Figure 3 illustrates meta-data described above included in the run of the user S1 of rmitMG6 participant for the TREC topic 326i. A set of meta-data is therefore assigned to each user.
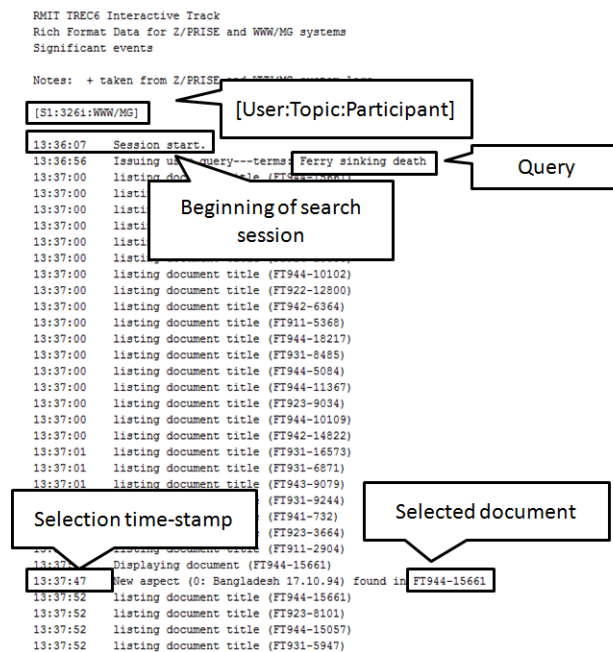

Figure 3: Example of run provided by user S1 belonging to participant rmitMG6 for the TREC topic 326i within the TREC-6 Interactive

4.1.2. Collaboration Simulation

Our experimental evaluation rests upon the simulation-based framework detailed in (Foley and Smeaton, 2009) adapted in two ways 1) user roles are identified through the computation of the domain-expertise levels of users involved in the simulation, based on the specificity of the documents they assessed as relevant; 2) evaluation metrics designed for estimating the effectiveness of collaborative search considering the iterative IR process on one hand, and the user roles on the other hand. We review in what follows the main components of the evaluation setting.

Collaborators. Collaboration involves a pair of users u_j and $u_{j'}$ sharing the same information need. The pairs of collaborators $c(u_j, u_{j'})$ are identified in the TREC interactive dataset and more particularly through runs illustrated in Figure 3 and provided by the set of participant groups P. We propose two ways of building pairs of collaborators depending on the level of domain expertise difference between users:

1. *Exhaustive pooling*: we consider, here, that the difference in the domain expertise level is relative within the pair of users. For this purpose, we note, within each participant group $\rho \in P$ and for a given TREC topic τ, all the combinations of pairs $c(u_j, u_{j'})$ of users u_j and $u_{j'}$ who have selected documents for the same information need.

2. *Selective pooling*: we consider, here, that the difference in the domain expertise level is absolute and determined in comparison with the expertise level of the overall participants. With this in mind, we built groups by combining domain experts and domain novices previously identified using a k-means classification. More particularly, for each TREC topic τ, we, first, computed a 2-means classification based on the expertise level of each user u_j involved within an interactive search session addressing topic τ. Then, within each participant group $\rho \in P$ and for a given TREC topic τ, we built all the combinations of pairs $c(u_j, u_{j'})$ of domain expert u_j and domain novice $u_{j'}$ with respect to the classification result.

User Roles. As our model considers user roles, we need to assign roles to users with respect to their domain expertise level, in compliance with roles' assumptions detailed in section 3.1.3. The expertise level $Expertise(u_j, \tau)$ of user u_j regarding topic τ is estimated according to two different methods:

1. *The specificity-based expertise $Expertise_S(u_j, \tau)$ depends on the average of the specificity level $\mathcal{L}_s(d_i)$, detailed in Equation 2, of selected documents $d_i \in \mathcal{D}^{\delta_\tau}(u_j)$ within a search session δ_τ belonging to topic τ.*

$$Expertise_S(u_j, \tau) = \frac{\sum_{d_i \in \mathcal{D}^{\delta_\tau}(u_j)} \mathcal{L}_s(d_i)}{|\mathcal{D}^{\delta_\tau}(u_j)|} \qquad (17)$$

2. *The authority-based expertise $Expertise_A(u_j, \tau)$ relies on the authority score detailed in (Foley and Smeaton, 2009). The authority-based expertise $Expertise_A(u_j, \tau)$ for user u_j and topic τ depends on terms involved within relevant judgments identified for TREC topic τ and documents $d_i \in \mathcal{D}^{\delta_\tau}(u_j)$ selected by user u_j within a search session δ_τ belonging to topic τ. The authority-based expertise $Expertise_A(u_j, \tau)$ is estimated by the correlation between the average term weights within these two categories of documents.*

Finally, a *domain expert* role is assigned to the collaborator with the highest specificity-based or authority-based domain expertise level; inversely, a *domain novice* role is assigned to the collaborator with the lowest specificity-based or authority-based domain expertise level.

Shared information need. A CIR setting is characterized by collaborators that share the same information need (Foley and Smeaton, 2010). We therefore consider the assumption expressed by Foley (2008) that collaborators formulate collaboratively the initial query for a particular shared information need. Moreover, this statement is strengthened by the fact that, within the TREC Interactive dataset, queries include generally 3 terms and 94% have at least one word in common (Foley, 2008). We exploit the meta-data of TREC participants to extract the query q_{u_j} submitted by each user u_j. Evaluating a pair of collaborators $c(u_j, u_{j'})$, the collaborative query q is obtained by merging the queries q_{u_j} and $q_{u_{j'}}$ submitted by the two collaborators u_j and $u_{j'}$ during their own individual interactive search. For instance, for TREC topic 326i, let us consider that the one user has submitted the query *"ferry sinking death"* and the other one, the query *"ferry sink people"*. The collaborative query for this user pair is *"ferry sink death people"*. Therefore, for a given TREC topic τ and a pair of collaborators $c(u_j, u_{j'})$, only one collaborative query q is generated. We notice that several queries $q \in Q_\tau$ are assigned to a TREC topic τ, depending on the participants' information needs formulations.

Search Session. Given a pair of collaborators $c(u_j, u_{j'})$, we need to build the timeline $L^S(u_j, u_{j'})$ of selected documents $\mathcal{D}^S(u_j, u_{j'})$ by the two collaborators u_j and $u_{j'}$ involved within a search session S. Using the time-stamp criteria of document selection provided in TREC metadata of participants P, we built the synchronous search session S, as shown in Figure 3. Thus, the synchronization of the two lists of documents $l^S(u_j)$ and $l^S(u_{j'})$ is based on the time-stamp selection criteria. Therefore, we obtain the timeline $L^S(u_j, u_{j'})$ of selected documents $\mathcal{D}^S(u_j, u_{j'})$. The overall methodology of search session simulation is illustrated in Figure 4.
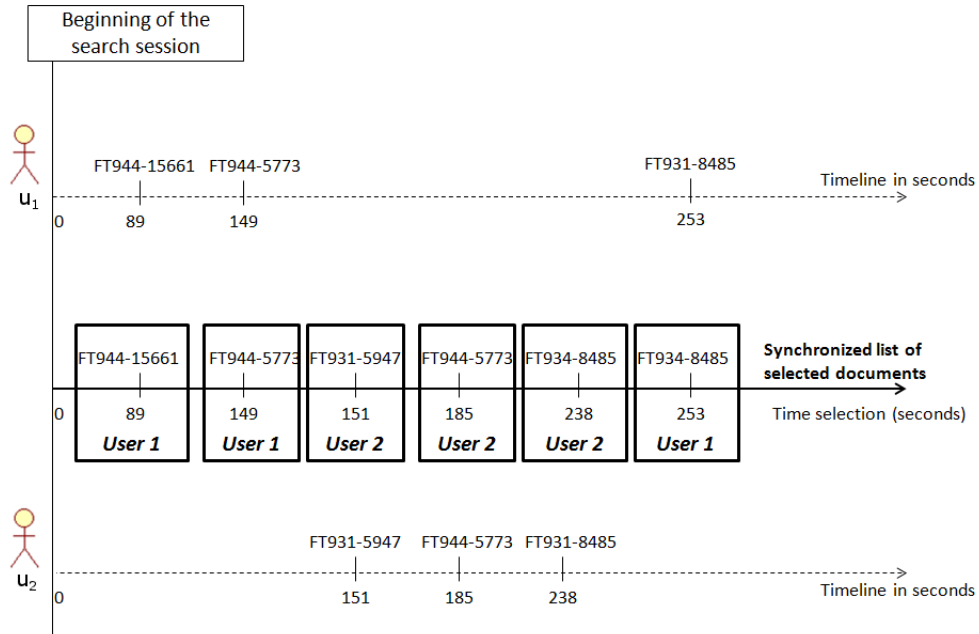

Figure 4: Search session simulation for rmitMG6 participants for TREC topic 326i

As done in (Foley and Smeaton, 2009), we assume that the maximum level of relevance recall is 30 since that users may examine around 30 documents in a list. Accordingly, we only used relevance feedback expressed at the top 30 to provide the successive document rankings. During the search session S, the document list $L^S(u_j, u_{j'})$ is considered dynamically according to the ranked lists $l^k(u_j, \mathcal{D}_{ns}^k)$ displayed to each user u_j at iteration k. Even though a document d_i is included in the synchronized document list $L^S(u_j, u_{j'})$, two conditions must be fulfilled for enabling a user u_j to select document d_i within

an iteration step k: 1) document d_i must be included in the user's currently displayed list $l^k(u_j, \mathcal{D}_{ns}^k)$ of 30 documents and 2) document d_i must not have been previously selected by the other collaborator $u_{j'}$ in order to ensure the division of labor principle.

After checking these conditions, the fact that user u_j selects a document launches an iteration k in order to display to user u_j a ranked list $l^k(u_j, \mathcal{D}_{ns}^k)$ of not already selected documents $\mathcal{D}_{ns}^k$. We emphasize that a session S includes one pair of collaborators $c(u_j, u_{j'})$, one TREC topic τ and one collaborative query q. Consequently, for each TREC topic τ, we can simulate several collaborative sessions δ_τ.

Let's introduce a small example illustrated on Figure 4. User u_1 has selected three documents, namely FT944-15661, FT944-5773 and FT931-8485 at respectively timestamps 89, 149 and 253. On another individual search log, we identify that user u_2 has also selected three documents, namely FT931-5947, FT944-5773 and FT931-8485 at timestamps 151, 185 and 238. We, therefore built the following time-line: $L^S(u_1, u_2) = \{(\text{FT944-15661}, u_1), (\text{FT944-5773}, u_1), (\text{FT931-5947}, u_2),$ $(\text{FT944-5773}, u_2), (\text{FT931-8485}, u_2), (\text{FT931-8485}, u_1)\}$. Therefore, the first iteration is launched only if the document FT944-15661 is currently displayed within the document list of user u_1. Later, the iteration 4 is launched if the document FT944-5773 is displayed within the document list of user u_2 and if the iteration 2 has not been launched, considering that in this case, this document would already has been selected by user u_1.

4.1.3. Metrics

Considering the collaborative search setting aiming at retrieving relevant documents and avoiding useless effort through the sharing of knowledge and the division of labor principles, we introduce two categories of metrics:

1. The precision-based measures aim at measuring the retrieval effectiveness of the collaborative search sessions. It is important to point out here that the length of the sessions in terms of the number of search iterations is irregular over the course of the whole search sessions. Indeed, the number of iterations varies between 3 and 12 according to collaborative search sessions. Highlighted irregularities of session lengths may induce a bias in basic precision measures by favoring long search sessions; to avoid this, we propose an evaluation measure for aggregating precision at the search-session level, namely the *micro-averaging precision* (Sebastiani, 2002), rather than

at the iteration search step level. With this in mind, we define three levels of analysis:

- The micro-averaging precision at the session level P_{micro} computes the precision of search sessions S abstracting the different intermediate-ranked lists of documents displayed within the search iterations. This measure is estimated as follows:

$$P_{micro} = \frac{1}{|\theta|} \sum_{\tau \in \theta} \frac{1}{|\delta_\tau|} \sum_{S \in \delta_\tau} \frac{\sum_{l \in \{L_S\}} Drel_l}{\sum_{l \in \{L_S\}} Dsel_l} \tag{18}$$

where θ is the set of TREC topic and δ_τ is the set of search sessions among the whole set of search session S related to TREC topic τ. L_S denotes the set of ranked lists displayed to collaborators within a search session S. $Drel_l$ represents the number of relevant documents in the list and $Dsel_l$ the number of selected, namely retrieved, documents in list l.

- The micro-precision at the role level $P_{micro}^{r,S}$ computes the precision of search session S considering ranked lists displayed for the user having the role r, estimated as below:

$$P_{micro}^{r,S} = \frac{1}{|\theta|} \sum_{\tau \in \theta} \frac{1}{|\delta_\tau|} \sum_{S \in \delta_\tau} \frac{\sum_{l \in \{L_{S,r}\}} Drel_l}{\sum_{l \in \{L_{S,r}\}} Dsel_l} \tag{19}$$

where $L_{S,r}$ denotes the set of ranked lists displayed to collaborators characterized by role $r \in \mathcal{R}$ within a search session S.

- The micro-precision at the iteration level P_{micro}^k computed for document lists displayed to both users at iteration k over all search session S is estimated as below:

$$P_{micro}^k = \frac{1}{|\theta|} \sum_{\tau \in \theta} \frac{1}{|\delta_\tau|} \sum_{S \in \delta_\tau} \frac{\sum_{l \in \{L_{S,r,k}\}} Drel_l}{\sum_{l \in \{L_{S,r,k}\}} Dsel_l} \tag{20}$$

where where $L_{S,k}$ denotes the set of ranked lists displayed to collaborators within a search session S at iteration k.

2. The collaboration-based measures aim at estimating the collaboration optimization through the diversification results over displayed document lists throughout the whole search session (Shah, 2012b):

- The coverage-based ratio Cov analyses the diversity of document lists displayed during the whole search session S. This ratio is estimated as follows:

$$Cov = \frac{1}{|\theta|} \sum_{\tau \in \theta} \frac{1}{|\delta_\tau|} \sum_{S \in \delta_\tau} \frac{Coverage(L_S)}{\sum_{l \in L_S} |l|} \tag{21}$$

where θ is the set of TREC topics and L_S is the set of displayed lists during search session S. $Coverage(L_S)$ corresponds to the number of distinct documents displayed during the whole search session S. The total number of documents displayed throughout the same session is noted $|l|$.

- The relevant coverage-based ratio $RCov$ analyses the relevance of distinct documents displayed during the whole search session S. We estimate the relevant coverage-based ratio as follows:

$$RCov = \frac{1}{|\theta|} \sum_{\tau \in \theta} \frac{1}{|\delta_\tau|} \sum_{S \in \delta_\tau} \frac{RelevantCoverage(L_S)}{\sum_{l \in L_S} |l|} \tag{22}$$

where $RelevantCoverage(L_S)$ corresponds to the number of distinct relevant documents displayed during the search session S.

For evaluating the retrieval effectiveness of our model, we use the evaluation measure at rank 30, this measure fits the assumptions of the maximum relevance recall measure and allows us to compare our results to Foley and Smeaton's model (Foley and Smeaton, 2009). Therefore, metrics are respectively noted $P@30$, $P@30^r$, $P@30^k$, $Cov@30$ and $RCov@30$.

4.1.4. Evaluation Scenarios

We defined four settings considering the different ways of building pairs of *domain expert* and *domain novice*:

- Setting $S_{Exh-Spec}$ in which collaborator groups are built using an exhaustive pooling and the users' roles are assigned according to the specificity-based expertise $Expertise_S(u_j, \tau)$.

31

- Setting $S_{Exh-Auth}$ in which collaborator groups are built using an exhaustive pooling and the users' roles are assigned according to the authority-based expertise $Expertise_A(u_j, \tau)$.

- Setting $S_{Sel-Spec}$ in which collaborator groups are built using a selective pooling and the users' roles are assigned according to the specificity-based expertise $Expertise_S(u_j, \tau)$.

- Setting $S_{Sel-Auth}$ in which collaborator groups are built using a selective pooling and the users' roles are assigned according to the authority-based expertise $Expertise_A(u_j, \tau)$.

Table 3 introduces the number of collaborative queries (and groups) within each considered search scenarios.

Setting	Number of collaborative queries/groups
$S_{Exh-Spec}$	243
$S_{Exh-Auth}$	243
$S_{Sel-Spec}$	95
$S_{Sel-Auth}$	81

Table 3: Number of queries/groups within each collaborative search setting.

4.1.5. Baselines

We compare our model to one individual-based baseline and three collaborative-based baseline scenarios. Below are all the scenarios tested in our experiments.

w/oDoL. This scenario includes all the components of our model, detailed in section 3.2.1 and 3.2.2 except the second aspect (b) of division of labor principle detailed in section 3.1.2. This scenario is a collaborative-based search setting considering that the document allocation step based on the EM algorithm, and detailed in section 3.2.2, assigns a user-oriented document score with respect to the domain expertise level of the whole set of users involved in the collaborative query.

w/oEM. This scenario includes only the first component of our model which estimates a user-oriented document score with respect to the domain expertise of one user, as presented in section 3.2.1. Moreover, we ensure the

division of labor principle by removing from the displayed lists documents displayed to other users, which leads to consider this scenario as a collaborative-based search setting.

w/oEMDoL. This scenario includes only the first component of our model which estimates a user-oriented document score with respect to the domain expertise of one user, as presented in section 3.2.1 and in scenario *w/oEM*. However, the second aspect of division of labor is not ensured and the ranked list of documents provided after the role-based document-scoring step is directly displayed to the user regardless of the rest of collaborators. Accordingly, this scenario leads to an individual-based search setting.

Foley and Smeaton's model, denoted FS. This scenario corresponds to the collaborative ranking model detailed in (Foley and Smeaton, 2009), and designed for a relevance feedback process. Authors propose both a document relevance scoring extending a probabilistic model (Robertson and Sparck Jones, 1976) and a query expansion method, adapted to a collaborative search setting. The authors propose a unique ranking expressing the global relevance of documents by taking into account user expertise-based factors, namely users' authority. The latter is estimated through the relevance of terms included in documents selected by the user. The similarity $sim(d_i, q)$ between document d_i and query q is expressed through a term-based partial user relevance weight $purw$, as follows:

$$
\begin{aligned}
sim(d_i, q) = & \ \sum_{t \in Q,D} purw(t) = \sum_{t \in Q,D} \frac{p(1-q)}{q(1-p)} \\
where\ p = & \ \sum_{u=1}^{2} \alpha_u \frac{r_{ut}+0.5}{R_u+1} \\
and\ q = & \ \sum_{u=1}^{2} \alpha_u \frac{n_t-r_{ut}+0.5}{N-R_u+1}
\end{aligned}
\tag{23}
$$

r_{ut}, respectively n_t, corresponds to the number of relevant documents identified by user u, respectively the number of documents, which contain term t. The number of documents identified by user u is noted R_u and the number of documents in the collection is N. The parameter α_u expresses the user's authority.

The query is expanded with top terms according to a term-weight, namely partial-user offer weight $puow$, based on the partial user relevance weight $purw$, computed as follows:

$$
puow(t) = \left(\sum_{u=1}^{2} \alpha_u \right) * purw(t)
\tag{24}
$$

To make this scenario comparable to our model, we assume that the users' authority expressed in (Foley and Smeaton, 2009), is similar to the expertise measure estimated through the selected document specificity, used for mining user roles in our experiments, as detailed in section 4.1.2. (Foley and Smeaton, 2009) propose two versions of the authority measure: the static one assigns the same value for a user throughout the search session, whereas the dynamic one considers that the user's value can evolve throughout the search session and that one user can be viewed as an expert at one iteration, and a novice for the following one. As our model considers predefined roles identified explicitly at the beginning of the session, we consider the static authority weight which is tuned for each setting detailed in section 4.1.4. Similarly to (Foley and Smeaton, 2009), we first tuned the authority measure, and obtained, for the four settings, an optimal value for the static-based authority α_u equals to 0.6 for the most authoritative user, viewed as the *domain expert*, and 0.4 for the other one, viewed as the *domain novice*.

4.2. Results and Discussion

4.2.1. Analyzing the Ranking Effectiveness at the Search Session Level

In this section we analyze the retrieval effectiveness of our model at the search-session level. First, we focus on settings $S_{Exh-Spec}$ and $S_{Exh-Auth}$ in which collaborator groups have been built using an exhaustive pooling. Then, we examine the case study of domain expert and domain novice groups created from a selective pooling. This second analysis concerns settings $S_{Sel-Spec}$ and $S_{Sel-Auth}$.

Exhaustive Pooling-based Collaborative Groups. Here, we analyze results for exhaustive pooling-based collaborative groups $S_{Exh-Spec}$ and $S_{Exh-Auth}$, respectively built upon an expertise level estimated using the document specificity measure or the authority one, as expressed in section 4.1.2. Table 4 presents the results of our model, noted *OurModel*, and the four baselines, described in section 4.1.5. Results show that our model generally overpasses the baselines according to both the precision-based measures and the collaborative-based ones. We do not observe distinct different statements between both settings which highlight the robustness of both indicators for identifying users' domain expertise. More particularly, our model reaches significant improvements, namely between 17% and 49% regardless of the metrics and the settings, over the collaborative model FS. This result shows the impact of considering users' roles within the

	Models	P@30	$\%Ch$	Cov@30	$\%Ch$	$RCov$@30	$\%Ch$
$S_{Exh-Spec}$	w/oDoL	0.260	+2.28%	0.383	+35.81% ∗∗	0.086	+42.56% ∗∗∗
	w/oEM	0.250	+6.34% ∗∗∗	0.364	+42.87% ∗∗∗	0.080	+52.36% ∗∗∗
	w/oEMDoL	0.283	−6.28%	0.279	+86.73% ∗∗∗	0.058	+112.92% ∗∗∗
	FS	0,221	+17.64% ∗∗∗	0.442	+17.07% ∗∗∗	0.086	+41.91% ∗∗∗
	OurModel	**0.266**	-	**0.520**	-	**0.123**	-
$S_{Exh-Auth}$	w/oDoL	0.259	+1.86%	0.414	+37.88% ∗∗∗	0.089	+45.17% ∗∗∗
	w/oEM	0.252	+4.98% ∗	0.396	+44.22% ∗∗∗	0.082	+56.46% ∗∗∗
	w/oEMDoL	0.284	−6.82%	0.273	+108.83% ∗∗∗	0.053	+142.18% ∗∗∗
	FS	0,216	+19.91% ∗∗∗	0.442	+28.54% ∗∗∗	0.086	+49.59% ∗∗∗
	OurModel	**0.265**	-	**0.571**	-	**0.129**	-

Table 4: Retrieval effectiveness of our collaborative model considering groups created from an exhaustive pooling - % Ch: our model improvement. Student test significance *: $0.01 < t < 0.05$; **: $0.001 < t < 0.01$; ***: $t < 0.001$

document scoring, and more particularly the domain expertise towards a document with respect to a topic, estimated by the role-based document scoring step. Another interesting aspect underlined from the results is that the division of labor and the EM-based document allocation step enable to enhance the retrieval effectiveness in contrast to model variants with only one of these components, namely w/oEM and $w/oDoL$. This suggests that 1) the document allocation step is effective in the document score learning for assigning a document to the most likely suitable user and 2) the integration of the division of labor allows enhancing the relevant coverage of displayed document subsets. However, we observe that the scenario $w/oEMDoL$ without any of these two components is slightly and not significantly more effective in terms of precision which suggests that an individual-based setting should be more effective. However, this statement is counterbalanced by the collaboration-based metrics. Indeed, the coverage-based retrieval effectiveness of our model is significantly better than the scenario $w/oEMDoL$ with a raise from 85% to 142%. We conclude that, even if the scenario $w/oEMDoL$ provides more effective document lists, it seems that documents are redundant over the whole set of displayed document lists, and therefore, the ratio of distinct relevant documents over the search session is lower. Thus, our model is more particularly oriented to residual precision.

We continue the results analysis by focusing on the retrieval effectiveness of our model at the iteration level. For this purpose, we analyzed the evolving micro-precision measure at each iteration k, described in Equation 20, throughout the whole session performed by the collaborative groups based on an exhaustive pooling, as detailed in Figure 5. We notice the increasing general trend of the curves with a noticeably improvement between 115% and 119% respectively for each setting $S_{Exh-Spec}$ and $S_{Exh-Auth}$, from the beginning to the end of the search session. This statement can be explained by the fact that document scores depend on both users' roles and relevance judgement-based user profiles enable to refine the shared information need, and, therefore, get a better insight of what relevant documents might be. However, we underline that the dramatic increase at the 11th iteration might not be considered as consistant considering the low number of groups reaching this iteration, namely 4 over 243 for both settings $S_{Exh-Spec}$ and $S_{Exh-Auth}$.
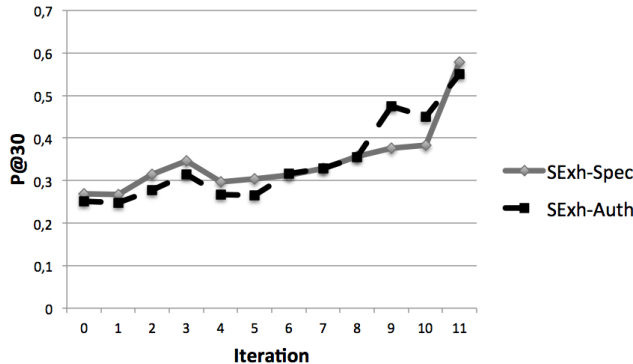

Figure 5: Analysis of the precision measure $Prec@30$ at the iteration level for collaborative groups based on an exhaustive pooling.

Selective Pooling-based Collaborative Groups. Here, we consider among the collaborators groups only those including domain experts and domain novices, explicitly defined through a selective pooling process, as expressed in section 4.1.2. Table 5 presents the obtained results using our model and the four baselines. Results highlight similar trend as those previously obtained for the collaboration groups built upon an exhaustive pooling presented in Table 4. More particularly, we observe a global improvement in terms of precision of our model in contrast to scenarios *w/oDoL, w/oEM* and *FS* and in terms of coverage-based ratios between 29% and 123% for the whole

	Models	$P@30$	$\%Ch$	$Cov@30$	$\%Ch$	$RCov@30$	$\%Ch$
$S_{Sel-Spec}$	w/oDoL	0.275	+4.09%*	0.362	+31.73%***	0.080	+29.63%***
	w/oEM	0.268	+7.01%*	0.335	+42.46%***	0.072	+43.99%***
	w/oEMDoL	0.303	−5.26%	0.258	+84.73%***	0.050	+105.88%***
	FS	0,208	+32.21%***	0.429	+10.95%*	0.075	+37.99%***
	OurModel	**0.287**	-	**0.477**	-	**0.103**	-
$S_{Sel-Auth}$	w/oDoL	0.251	+0.86%	0.400	+36.44%***	0.081	+35.52%***
	w/oEM	0.239	+5.87%	0.362	+50.11%***	0.070	+56.17%***
	w/oEMDoL	0.279	−9.29%	0.254	+114.48%***	0.048	+125.96%***
	FS	0,166	+51.20%***	0.429	+26.71%***	0.081	+34.22%***
	OurModel	**0.253**	-	**0.544**	-	**0.110**	-

Table 5: Retrieval effectiveness of our collaborative model considering groups based on a selective pooling - % Ch: our model improvement. Student test significance *: $0.01 < t < 0.05$; **: $0.001 < t < 0.01$; ***: $t < 0.001$

set of scenarios. We also notice that the expertise level criteria, based either on specificity or authority, does not impact the retrieval effectiveness of our model. Moreover, the obtained results for the selective pooling process highlight that our model is also effective for the particular study case of the collaboration between *domain expert* and *domain novice*, identified through the selective pooling methodology.

We also analyze the evolving micro-precision measure of our model throughout the whole session at the iteration level. Figure 6 illustrates such results for collaborative groups based on a selective pooling process. We can see that the results are similar to those illustrated in Figure 5, except a drop at iteration 8 for setting $S_{Sel-Spec}$ which does not seem consistant considering that only one collaborative group among the 95 ones got more than 8 iterations.

These results lead to different conclusions. First, our model allows retrieving more distinct and relevant documents than the other scenarios throughout the whole search sessions. Second, the significant improvement of our model with respect to scenario *w/oEM* emphasizes the importance of the document allocation step using the EM-algorithm. Indeed, this learning method enables to estimate the user preference for documents with an expertise level as close as his own domain expertise. Third, considering scenarios *w/oDoL* and *w/oEMDoL*, we suspect that applying the division of labor principle may lead as a prior state to likely slightly reduce the retrieval effectiveness of col-
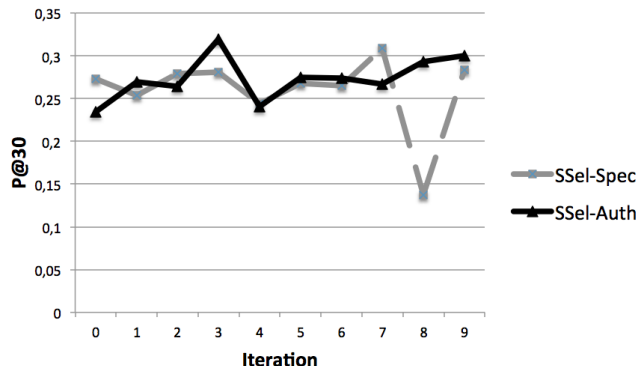
Figure 6: Analysis of the precision measure $Prec@30$ at the iteration level for collaborative groups based on a selective pooling.

laborative ranking models. However, coverage and relevant coverage-based ratios counterbalance this statement by highlighting that our model provides more distinct and relevant documents throughout the whole search sessions in contrast to models without division of labor which are likely to display redundant documents. Thus, we advocate that collaboration may improve the retrieval effectiveness in terms of diversity and relevance in contrast to an individual setting, instanciated by the baseline $w/oEMDoL$, which only provides to users relevant documents but do not avoid redundancy between displayed document lists. Fourth, our model retrieval effectiveness significantly improves over the collaborative document ranking FS' one. It can be explained by the fact that our model takes into account users' domain expertise for assigning documents to the most suitable users, and, therefore better fits with their knowledge, whereas the scenario FS estimates a global relevance based on the whole users' relevance judgments. Finally, our model can be applied in different cases. Collaborative groups can rely 1) on a relative domain expertise difference level, namely an exhaustive pooling, in which we consider that a collaborator is more willing to know the domain than the other one or 2) on a selective pooling which clearly identifies expert and novice users of the domain. Moreover, the expertise level of users can be estimated in different ways and results show the robustness of our model regardless of the difference in the domain expertise levels between the collaborators.

| | | $S_{Exh-Spec}$ | | $S_{Exh-Auth}$ | | $S_{Sel-Spec}$ | | $S_{Sel-Auth}$ | |
	Models	$P@30^r$	$\%Ch$	$P@30^r$	$\%Ch$	$P@30^r$	$\%Ch$	$P@30^r$	$\%Ch$
Domain expert	w/oDoL	0.253	+5.43%	0.264	+5.67%	0.268	+9.12%∗	0.285	+2.01%∗
	w/oEM	0.246	+8.44%∗	0.259	+7.70%∗	0.270	+8.22%	0.264	+9.78%
	w/oEMDoL	0.268	−0.59%	0.285	−2.30%	0.295	−0.94%	0.315	−7.87%
	FS	0.219	+22.11%∗∗	0.233	+19.10%∗	0.244	+19.74%∗	0.234	+24.08%∗
	ourModel	0.267		0.279		0.292		0.201	
Domain novice	w/oDoL	0.233	+0.60%	0.238	+1.67%	0.270	+0.47%	0.250	+4.11%
	w/oEM	0.217	+7.70%∗	0.227	+6.51%∗	0.242	+11.68%∗∗∗	0.238	+8.97%∗∗∗
	w/oEMDoL	0.253	−7.30%∗∗∗	0.253	−4.52%∗	0.287	−5.69%	0.262	−1.05.%
	FS	0.222	+5.63%	0.233	+3.86%	0.237	+14.17%	0.209	+23.81%
	ourModel	0.235		0.241		0.271		0.260	

Table 6: Retrieval effectiveness of our collaborative model according to the two user roles - % Ch: our model improvement. Student test significance *: $0.01 < t < 0.05$; **: $0.001 < t < 0.01$; ***: $t < 0.001$

4.2.2. Analyzing the Ranking Effectiveness at the User Role Level

Here, we refine our analysis by focusing our experiments at the user-role level. In contrast to the previous analysis which focuses on the retrieval effectiveness regardless of users' roles, this analysis consists in highlighting the global retrieval effectiveness of our model towards each role throughout the whole set of search sessions.

Table 6 gives a comparison between the retrieval effectiveness of different used ranking models regarding our four scenarios according to user roles. Within a search session, micro-precisions at rank 30 are estimated using ranked lists displayed to each user role as detailed in Equation 19. We can see that our model generally overpasses the baselines for both user roles but significant improvements are noticeable only in around 50% of the cases. Statements are similar to those obtained for the analysis at the search session level. On one hand, we note that our model provides better results, with significant improvements around 8%, than those obtained for the baseline *w/oEM*, highlighting the importance of this step in the result personalization. Indeed, the EM-based document allocation enables to assign a document to the most suitable user, and, therefore, to provide a better-adapted ranked list to the users' roles characteristics. On the other hand, we can see that the baseline *w/oEMDoL* provides better results, generally not significant, than our model for both user roles. One

possible reason of this statement could be due to the fact that the scenario $w/oEMDoL$ does not apply any type of division of labor and allows displaying to the a user the same top-ranked documents throughout the whole search session. In contrast, our model is more residual relevance-oriented considering that it applies a double division of labor based on the EM-based document allocation step in addition to the fact of removing currently seen documents from displayed lists. We underline that the effectiveness decreases are only significant in the case of the domain novice role within the exhaustive pooling setting. This observation can be explained by the fact that personalization depends heavily on the users' predefined roles rather than their expertise levels in the search session. Indeed, considering that the set of pairs of collaborators generated by the selective method S_{Sel-*} is included in the set of pairs built upon the exhaustive one S_{Exh-*}, we expect that the difference in the significance results comes from the non personalized rankings towards users having close expertise levels. This highlights another reason that probably explains these results: poor relevance feedback collected upon non personalized document rankings, particularly for novices, might impact negatively the retrieval effectiveness. Moreover, our model obtains significant improvements ranged between 19% and 24% with respect to the baseline FS for the domain expert users, but our model enhancement, between 3% and 23%, is not significant for the domain novice user. This contrast may be explained by the assumptions of the collaborative baseline FS. Indeed, this collaborative ranking model does not consider user roles and more particularly any personalization within the document ranking. Even if the document relevance is estimated through users' relevance judgments, the model applies a smoothing of documents rankings leading to an outcome that fits the users' expertise at the average level rather than the personal one. Thus, the users are treated finally as equal, and in the light of the results obtained according to the role-based analysis, we can assume that they are considered as novices or in other terms FS' search model is slightly as effective as our model is for novices: we note not significant improvements, from 3.86% to 23.81%, for our model.

Moreover, Table 6 highlights that the retrieval effectiveness of *domain experts* for the four considered scenarios is higher than the *domain novices'* one. In order to analyze whether the differences of retrieval effectiveness between *domain expert* and *domain novice* roles are significant, we propose a mean comparison of precisions between *domain expert* and *domain novice*

roles using Student t-test. As we can see from Table 7, the retrieval effectiveness differences between *domain novice* and *domain expert* roles are not statistically significant (p-value>0.5). Thus, we cannot conclude if our model favors or not one particular user' role in terms of displayed document ranking effectiveness.

Scenarios	$S_{Exh-Spec}$	$S_{Exh-Auth}$	$S_{Sel-Spec}$	$S_{Sel-Auth}$
Domain novice	0.234	0.241	0.271	0.259
Domain expert	**0.267**	**0.279**	**0.292**	**0.291**
%*Ch*	+13.76%	+15.19%	+7.78%	+11.91%
p-value	0.20	0.16	0.58	0.38

Table 7: Retrieval effectiveness statistics according the *domain expert* and *domain novice* roles - % Ch: *Domain expert* improvement. - p-value: student test p-value

Our last focus aims at analyzing whether our model enables to improve the search experience of users. In the same spirit, White et al. (2009) introduce the search challenge for novices to reinforce their search behaviors by, for instance, using appropriate vocabulary towards the information need. For this purpose, we aim, here, at analyzing whether domain novice users improve their knowledge about the shared information need, and more specifically, whether they leverage displayed document lists for enhancing the specificity of selected documents. To avoid the bias underlying the specificity-based domain expertise measure used for building collaborative groups, we only performed this analysis for settings relying on the authority-based domain expertise measure. Figure 7 illustrates the evolving average specificity of domain novices users' displayed documents in comparison to selected documents at each iteration.

Iteration 0 refers to the initialization step S_{init}, and the following ones represent iterations $k \in S_{iter}$. First, we focus on the average specificity of displayed document lists over search sessions for evaluating to what extent our model supports the domain expertise development for domain novices. For both exhaustive and selective pooling-based collaborative groups, we highlight that the average specificity of displayed documents drops at iteration 1 which corresponds to the first iteration which begins to personalize document lists by integrating document selection within user profiles. Iteration 0 considers only the query terms as evidence source of the user profile, therefore, document lists displayed at this level are not personalized for each user's role. From iteration 1, we notice an increasing trend for both curves representing
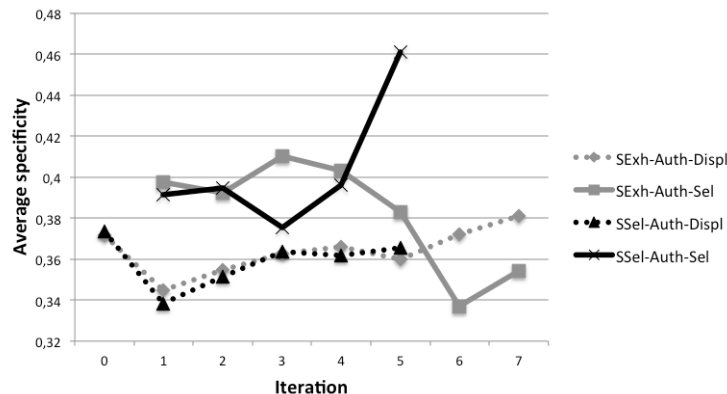
Figure 7: Analysis of domain novice knowledge improvement over collaborative search sessions.

the displayed documents, namely $S_{Exh-Auth-Displ}$, $S_{Sel-Auth-Displ}$. This trend highlights that our model gives the opportunity to the domain novice users to improve their search skills throughout the search session in so far as that our model displays document lists characterized by an increasing specificity level. Second, we aim at highlighting whether domain novice users' knowledge is improved throughout the search session. In this aim, we focus on the average specificity of their selected documents at each iteration rather than the displayed ones. We can see from Figure 7 that the average specificity of selected documents by novice users decreases over iterations for collaborative groups built upon an exhaustive pooling method while this measure increases for selective pooling-based collaborative groups. This highlights that collaborators previously identified as domain novices, using the selective pooling methodology, are more likely to enhance their domain expertise in contrast to users identified as novices only on the consideration of the relative domain expertise difference level among collaborators. Indeed, the exhaustive pooling-based collaborative groups might involve two domain experts with slightly different domain expertise levels.

5. Discussion and Concluding Remarks

In this paper, we presented a novel collaborative ranking model based on roles taking into account users' domain expertise level. Our model includes a document scoring based on user roles and an Expectation Maximization-based learning method for document allocation to user roles. Collaboration

is generally supported by a system-mediated approach in which the division of labor principle avoids redundancy among users' actions, and assigns documents to the most likely suitable user. Our evaluation drawn on a TREC-based simulation framework shows that a collaborative search built on our model is effective at both session and role levels. More specifically, we showed that retrieval effectiveness in terms of precision and diversity improvements are significant. By comparing our model to a state-of-the-art collaborative ranking model, we also underlined that the proposed model is more adapted for collaboratively ranking documents within a CIR task which assumes that users have different domain expertise levels.

Another interesting contribution of our work presented in this paper is that we highlight a generic approach for collaborative ranking models for different users' roles. Indeed, the first step of our model, namely the role-based document scoring, estimates the document relevance for a user by a role-based smoothing involving the specificity features underlying domain expert and novice characteristics. For another pair of roles, this step might be adapted by considering other roles' features within the document relevance smoothing toward the collaborator with respect to the shared information need.

However, our collaborative ranking approach has some limitations:

- We chose a particular instantiation of CIR setting remaining on a synchronous iterative process which ranks documents after each user's document selection in order to display new document lists to the same user which better match his evolving domain knowledge. We note, however that the system-mediated CIR ranking task can be viewed differently. For instance, the collaborative search session may be asynchronous, and in this context, interactions between users should be considered in order to analyze the information flow. Also, document lists could be updated only if users consider that these lists do not include any more relevant documents or after a query reformulation. Moreover, our model does not integrate any type of user-driven collaboration, such as interactions among users which can enable to have a better understanding of their current search strategies or their shared information need.

- Our model only proposes a collaborative document ranking model at the iteration level. Unfortunatly, it does not generate a final list of doc-

uments highlighting the collective relevance of both users throughout the whole search session.

- Our experimental setup is a simulation of collaboration based upon individual interactive search logs. We highlight that several previous work have performed simulation-based evaluation within a relevance feedback process (Keskustalo et al., 2006; White et al., 2005). To make our scenario as natural as possible, we used realistic assumptions: 1) we simulated difference levels of expertise using two indicators, namely specificity and authority, highlighting the expertise level of users (Foley and Smeaton, 2009; White et al., 2005) and 2) we relied on clickthrough data for inferring the relevance of documents, similarly to user-studies performed in (Pickens et al., 2008). However, this can lead to different weaknesses. Mainly, the fact that collaboration is extracted from individual logs raise the concern if our experimental setup really reflects users' behaviors within a collaborative search session. For instance and as mentioned above, the simulated search session does not include all the aspects surrounding a collaborative search session, namely interactions, communications and search strategies among users.

Accordingly, we plan to enhance our model by:

- Giving the possibility for users to refine their information need throughout the search session. Our model should allow users to reformulate the collaborative query during the task. Accordingly, users' profiles could be inferred from their search history including feedback and subtopics of interest issued from the submitted queries. Moreover, we plan to enhance the retrieval model with a final step aiming to provide an aggregated list of relevant documents.

- Extending the model for supporting multiple users involved within a collaborative search session. This should allow to optimize collaboration from larger user groups and also leverage collaboration from an increasing difference in the levels of users' domain expertise.

- Validating our simulation methodology by focusing more in-depth on two main aspects: 1) comparing our division of labor appliance with other forms of this principle, for instance by ensuring the topical diversity among document lists in addition to display distinct document

44

subsets and 2) comparing our relevance feedback process, which proposes an immediate personalization ranking after each document selection, with another version which updates document lists only after a query reformulation.

- Evaluating our model through a large-scale user study involving collaborators with an expertise difference level identified along the spectrum of the domain expertise. Moreover, we plan to involve dyads and triads in order to analyze the impact of the group size on the retrieval effectiveness of our collaborative ranking model.

Another direction for our future work is to mine dynamically users' profiles and roles from a running search session using users' interactions and behaviors. This contribution would aim at implicitly detecting roles which allows to leverage users' skills and search strategies in which they are the most effective. This process would upstream role-based collaborative ranking models, which, as of now, explicitly assign users' roles without any consideration of their skills and search strategies.

References

Allen, B., 1991. Topic knowledge and online catalog search formulation. The Library Quarterly , 188–213.

Attfield, S., Adams, A., Blandford, A., 2006. Patient information needs: pre- and post-consultation. Health Informatics Journal 12, 165–177.

Attfield, S., Blandford, A., Makri, S., 2010. Social and interactional practices for disseminating current awareness information in an organisational setting. Inf. Process. Manage. 46, 632–645.

Castells, P., Vargas, S., Wang, J., 2011. Novelty and diversity metrics for recommender systems: Choice, discovery and relevance, in: ECIR Conference Proceedings.

Catledge, L.D., Pitkow, J.E., 1995. Characterizing browsing strategies in the world-wide web, in: WWW Conference Proceedings, pp. 1065–1073.

Collins-Thompson, K., Bennett, P.N., White, R.W., de la Chica, S., Sontag, D., 2011. Personalizing web search results by reading level, in: CIKM Conference Proceedings, pp. 403–412.

Collins-Thompson, K., Callan, J., 2004. A language modeling approach to predicting reading difficulty, in: HLT/NAACL Conference Proceedings.

Conrad, J.G., 2007. E-discovery revisited: A broader perspective for ir researchers, in: ICAIL Conference Proceedings, pp. 321–345.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society, Series B 39, 1–38.

Denning, P.J., Yaholkovsky, P., 2008. Getting to "we". Commun. ACM 51, 19–24.

ECDPC, 2011. A literature review on health information-seeking behaviour on the web: a health consumer and health professional perspective. Technical Report. European Centre for Disease Prevention and Control.

Erdem, S., Harrison-Walker, L., 2006. The role of the internet in physician-patient relationships: the issue of trust. Business Horizons 49, 7–393.

Erickson, T., 2010. A social proxy for collective search, in: CSCW Conference Proceedings.

Filho, F.F., Olson, G.M., de Geus, P.L., 2010. Kolline: a task-oriented system for collaborative information seeking, in: ICDC Conference Proceedings, pp. 89–94.

Foley, C., 2008. Division of labour and sharing of knowledge for synchronous collaborative information retrieval. Ph.D. thesis. Dublin city University.

Foley, C., Smeaton, A.F., 2009. Synchronous collaborative information retrieval: Techniques and evaluation, in: ECIR Conference Proceedings, pp. 42–53.

Foley, C., Smeaton, A.F., 2010. Division of labour and sharing of knowledge for synchronous collaborative information retrieval. Inf. Process. Manage. 46, 762–772.

Fox, S., 2008. The engaged E-patient population, Pew internet & American Life Project 2008. Technical Report. URL: http://www.pewinternet.org/Reports/2008/The-Engaged-Epatient-Population.aspx.

Gianoutsos, S., Grundy, J., 1996. Collaborative work with the world wide web: Adding cscw support to a web browser, in: CSCW Conference Proceedings, pp. 14–21.

Golovchinsky, G., Diriye, A.M., 2011. Session-based search with querium, in: HCIR Conference Proceedings.

Golovchinsky, G., Qvarfordt, O., Pickens, J., 2009. Collaborative information seeking. IEEE Computer 42, 47–51.

Hansen, P., Järvelin, K., 2005. Collaborative information retrieval in an information-intensive domain. Inf. Process. Manage. 41, 1101–1119.

Harman, D., 2002. Overview of the trec 2002 novelty track, in: TREC 2002 Conference Proceedings, pp. 46–55.

Hembrooke, H.A., Granka, L.A., Gay, G.K., Liddy, E.D., 2005. The effects of expertise and feedback on search term selection and subsequent learning: Research articles. J. Am. Soc. Inf. Sci. Technol. 56, 861–871.

Hölscher, C., Strube, G., 2000. Web search behavior of internet experts and newbies. Comput. Netw. 33, 337–346.

Horowitz, D., Kamvar, S.D., 2010. The anatomy of a large-scale social search engine, in: WWW Conference Proceedings, pp. 431–440.

Hsieh-yee, L., 1993. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science 44, 161–174.

Isaacs, E.A., Clark, H.H., 1987. References in conversation between experts and novices. Journal of Experimental Psychology General 116, 26–37.

Jelinek, F., Mercer, R.L., 1980. Interpolated estimation of markov source parameters from sparse data, in: Pattern Recognition in Practice Conference Proceedings, pp. 381–397.

Joho, H., Hannah, D., Jose, J.M., 2009. Revisiting ir techniques for collaborative search strategies, in: ECIR Conference Proceedings, pp. 66–77.

Kang, R., Fu, W.T., 2010. Exploratory information search by domain experts and novices, in: IUI Conference Proceedings, pp. 329–332.

Kashyap, A., Hristidis, V., Petropoulos, M., 2010. Facetor: cost-driven exploration of faceted query results, in: CIKM Conference Proceedings, pp. 719–728.

Keskustalo, H., Järvelin, K., Pirkola, A., 2006. The effects of relevance feedback quality and quantity in interactive relevance feedback: A simulation based on user modeling, in: ECIR Conference Proceedings, pp. 191–204.

Kim, G., 2006. Relationship between index term specificity and relevance judgment. Inf. Process. Manage. 42, 1218–1229.

Kim, J.Y., Collins-Thompson, K., Bennett, P.N., Dumais, S.T., 2012. Characterizing web content, user interests, and search behavior by reading level and topic, in: WSDM Conference Proceedings, pp. 213–222.

Kuhlthau, C., Spink, A., Cool, C., 1992. Exploration into stages in the information search process in online information retrieval: communication between users and intermediaries, in: ASIS Conference Proceedings, pp. 67–71.

Marsland, S., 2009. Machine Learning: An Algorithmic Perspective. CRC Press.

Mascaro, C.M., Goggins, S., 2010. Collaborative information seeking in an online political group environment, in: International Workshop on CIS of CSCW Conference.

McMullan, M., 2006. Patients using the internet to obtain health information: How this affects the patient-health professional relationship. Patient Education and Counseling 63, 24 – 28.

Moraveji, N., Morris, M., Morris, D., Czerwinski, M., Henry Riche, N., 2011. Classsearch: facilitating the development of web search skills through social learning, in: CHI Conference Proceedings, pp. 1797–1806.

Morris, M.R., Horvitz, E., 2007. Searchtogether: an interface for collaborative web search, in: UIST Conference Proceedings, pp. 3–12.

Morris, M.R., Morris, D., 2011. Understanding the potential for collaborative search technologies in clinical settings, in: International Workshop on CIR of CIKM Conference, pp. 11–14.

Morris, M.R., Paepcke, A., Winograd, T., 2006. Teamsearch: Comparing techniques for co-present collaborative search of digital media, in: International Workshop on Horizontal Interactive Human-Computer Systems of IEEE Conference, pp. 97–104.

Morris, M.R., Teevan, J., 2009. Collaborative Web Search: Who, What, Where, When, and Why. Morgan & Claypool, UK.

Morris, M.R., Teevan, J., Bush, S., 2008. Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting, in: CSCW Conference Proceedings, pp. 481–484.

Over, P., 2001. The trec interactive track: an annotated bibliography. Inf. Process. Manage. 37, 369–381.

Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., Back, M., 2008. Algorithmic mediation for collaborative exploratory search, in: SIGIR Conference Proceedings, pp. 315–322.

Podichetty, V.K., Booher, J., Whitfield, M., Biscup, R.S., 2006. Assessment of internet use and effects among healthcare professionals: a cross sectional survey. Postgrad Med J 82, 274–9.

Privault, C., O'Neill, J., Ciriza, V., Renders, J.M., 2010. A new tangible user interface for machine learning document review. Artif. Intell. Law 18, 459–479.

Robertson, S.E., Sparck Jones, K., 1976. Relevance weighting of search terms. Journal of the American Society for Information Science 27, 129–146.

Rodriguez Perez, J.A., Whiting, S., Jose, J.M., 2011. Cofox: a visual collaborative browser, in: International workshop on CIR of CIKM Conference, pp. 29–32.

Rudd, J., Rudd, M., 1986. Coping with information load: strategies and implications for librarians. College & Research Libraries 47, 315–322.

Sebastiani, F., 2002. Machine learning in automated text categorization. ACM Comput. Surv. 34, 1–47.

Shah, C., 2012a. Coagmento - A Case Study in Designing User-Centric Collaborative Information Seeking System. IGI Global; Hershey, PA.

Shah, C., 2012b. Collaborative Information Seeking: The Art and Science of Making The Whole Greater Than The Sum of All. Springer.

Shah, C., Hansen, P., Capra, R., 2012. Report on the second workshop on cis: New orleans, 12 october, 2011, Information Research.

Shah, C., Pickens, J., Golovchinsky, G., 2010. Role-based results redistribution for collaborative information retrieval. Inf. Process. Manage. 46, 773–781.

Smeaton, A.F., Foley, C., Gurrin, C., Lee, H., McGivney, S., 2006. Collaborative searching for video using the fischlar system and a diamondtouch table, in: TABLETOP Conference Proceedings, pp. 151–159.

Smirnova, E., Balog, K., 2011. A user-oriented model for expert finding, in: ECIR Conference Proceedings, pp. 580–592.

Soboroff, I., Harman, D., 2005. Novelty detection: the trec experience, in: HLT Conference Proceedings, pp. 105–112.

Soulier, L., Tamine, L., Bahsoun, W., 2013. A collaborative document ranking model for a multi-faceted search, in: AIRS Conference Proceedings, pp. 109–120.

Taylor, R., 1968. Question-negotiation and information seeking in libraries. College & Research Libraries 29, 178–194.

Toomela, A., 2007. Sometimes one is more than two: When collaboration inhibits knowledge construction. Integrative Psychological and Behavioral Science 41, 198–207.

Twidale, M.B., Nichols, D.M., 1996. Collaborative browsing and visualisation of the search process. Aslib Proceedings 48, 177–182.

Twidale, M.B., Nichols, D.M., Paice, C.D., 1997. Browsing is a collaborative process. Inf. Process. Manage. 33, 761–783.

Vakkari, P., Pennanen, M., Serola, S., 2003. Changes of search terms and tactics while writing a research proposal a longitudinal case study. Inf. Process. Manage. 39, 445–463.

Wald, H.S., Dube, C.E., Anthony, D.C., 2007. Untangling the Web–the impact of Internet use on health care and the physician-patient relationship. Patient education and counseling 68, 218–224.

Wang, J., Soergel, D., 2010. A user study of relevance judgments for e-discovery, in: ASIS&T Conference Proceedings, pp. 74:1–74:10.

White, R.W., Dumais, S.T., Teevan, J., 2009. Characterizing the influence of domain expertise on web search behavior, in: WSDM Conference Proceedings, pp. 132–141.

White, R.W., Richardson, M., 2012. Effects of expertise differences in synchronous social q&a, in: SIGIR Conference Proceedings, pp. 1055–1056.

White, R.W., Richardson, M., Liu, Y., 2011. Effects of community soze and contact rate in synchronous social q&a, in: CHI Conference Proceedings, pp. 2837–2846.

White, R.W., Ruthven, I., Jose, J.M., Rijsbergen, C.J.V., 2005. Evaluating implicit feedback models using searcher simulations. ACM Trans. Inf. Syst. 23, 325–361.

Wielhorski, K., 1994. Teaching remote users how to use electronic information resources. The Public-Access Computer Systems Review 5, 5–20.