



# On domain expertise-based roles in collaborative information retrieval



Laure Soulier\*, Lynda Tamine, Wahiba Bahsoun

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

## ARTICLE INFO

### Article history:

Received 12 August 2013

Received in revised form 6 April 2014

Accepted 7 April 2014

Available online 15 May 2014

### Keywords:

Collaborative information retrieval

Domain expertise

Ranking model

Learning-method

## 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.

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

An information retrieval (IR) task is generally perceived as an individual process (Catledge & Pitkow, 1995). However, the analysis of user intent within an IR task has underlined the increasing need of collaboration to answer multifaceted (Kashyap, Hristidis, & Petropoulos, 2010) and multitopical queries (Castells, Vargas, & Wang, 2011) where result diversity is expected. Indeed, for such queries and more generally, for complex and exploratory ones (Denning & 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 & Smeaton, 2010; Morris & 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 & Morris, 2011), academic (Moraveji, Morris, Morris, Czerwinski, & Henry Riche, 2011) or political (Mascaro & 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, Hannah, & Jose, 2009; Shah, Hansen, & Capra, 2012). A role taxonomy

\* Corresponding author. Tel.: +33 561557440.

E-mail addresses: [soulier@irit.fr](mailto:soulier@irit.fr) (L. Soulier), [tamine@irit.fr](mailto:tamine@irit.fr) (L. Tamine), [wbahsoun@irit.fr](mailto:wbahsoun@irit.fr) (W. Bahsoun).

(Golovchinsky, Qvarfordt, & Pickens, 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, Golovchinsky, Shah, Qvarfordt, & Back, 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, Blandford, & Makri, 2010; Privault, O'Neill, Ciriza, & Renders, 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 & Rudd, 1986; Twidale, Nichols, & Paice, 1997) in which users and librarians collaborate for satisfying users' bibliographic information; and (4) the question–answering domain (Horowitz & Kamvar, 2010; White & Richardson, 2012) in which users collaborate with the asker for solving his/her own information need. Moreover, previous work surrounding user search behavior domain (Allen, 1991; Hembrooke, Granka, Gay, & Liddy, 2005; White, Dumais, & Teevan, 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 retrieval 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 simultaneous 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 first step integrates document specificity and novelty according to user roles within a language model smoothing. Then, the Expectation–Maximization (EM) learning method (Dempster, Laird, & Rubin, 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.

Download English Version:

<https://daneshyari.com/en/article/515842>

Download Persian Version:

<https://daneshyari.com/article/515842>

[Daneshyari.com](https://daneshyari.com)