On the collaboration support in Information Retrieval

Laure Soulier, Lynda Tamine

Abstract

Collaborative Information Retrieval (CIR) is a well-known setting in which explicit collaboration occurs among a group of users working together to solve a shared information need. This type of collaboration has been deemed as beneficial for complex or exploratory search tasks. With the multiplicity of factors impacting on the search effectiveness (e.g., collaborators' interactions or the individual perception of the shared information need), CIR gives rise to several challenges in terms of collaboration support through algorithmic approaches. More particularly, CIR should allow to satisfy the shared information need by optimizing the collaboration within the search session over all collaborators while ensuring that both mutually beneficial goals are reached and that the cognitive cost of the collaboration does not impact the search effectiveness. In this survey, we propose an overview of CIR with a particular focus on the collaboration support through algorithmic approaches. The objective of this paper is (a) to organize previous empirical studies analyzing collaborative search with the goal to provide useful design implications for CIR models, (b) to give a picture of the CIR area by distinguishing two main categories of models using the collaboration mediation axis, and (c) to point out potential perspectives in the domain.

1

On the collaboration support in Information Retrieval

Laure Soulier
Sorbonne Universités, UPMC Univ Paris 06,
UMR 7606, LIP6 - Paris - France
Lynda Tamine
University of Toulouse - UPS,
IRIT - 118 route de Narbonne - Toulouse - France

May 2, 2017

1 Introduction

In its premise, Information Retrieval (IR) has been characterized as a user-independent process in which fundamental ranking models have been proposed [102, 99, 91]. The core idea of these models relies on the fact that the user is exhaustively represented by his/her query and that the relevance is built upon a query-document topical matching. In this context, the literature includes a wide-range of system-based ranking approaches [102, 103, 99, 91, 115]. These models aim at identifying the set of relevant documents $\mathcal{D} = \{d_1, \cdots, d_i, \cdots, d_N\}$ with respect to an information need expressed using a query q_j. Accordingly, each document d_i receives a similarity score estimated through a Retrieval Status Value function $RSV(q_j, d_i)$. This approach fits with the static point of view in which the relevance is only system-oriented with no consideration of the user and its interactions with the search results.

However, a huge amount of work [73, 53] highlights the difficulty of the system to capture the search intent behind a query, which is generally short (less than four terms). To tackle this issue, a new evidence source, namely the user dimension, has been introduced [11, 47, 61]. Beyond the consideration of the user's interests and expertise [127, 131], learning from user's interactions [58] constitutes a new means for capturing the search intent, and accordingly adapting the document ranking model. From the IR formal point of view, this interactive framework gives rise to well-known challenges, such as the modeling

*We would like to thanks Benjamin Piwowarski for his carefully reading and advices helping to clarify the paper.

of user u_j and its consideration within the estimation of the document relevance $RSV(d_i|q_j, u_j)$. Thus, the IR paradigm shifts from a static point of view to an interactive one in which user-system interactions are exploited to enhance the effectiveness of the document ranking. One could see this user-driven framework as a user-system collaboration in which the user (respectively, the system) leverages the actions/outputs of the system (respectively, the user). However, one limitation of this approach relies on the fact that the ranking adaptation focuses on short-term retrieval in which successive queries are independent [58].

Recently, another line of work, called dynamic IR, and that could be seen as an extension of interactive IR, has been introduced [60, 132]. The document ranking process relies on the assumption that the search process consists of a sequence of users' interactions with the system (e.g., relevance feedback or query reformulation) allowing to optimize the search effectiveness over the search sessions. Accordingly, dynamic IR proposes a new IR paradigm aiming to learn dynamically from past user-system interactions and predicting the future in terms of relevance satisfaction [132]. In contrast to interactive IR which relies on a "one shot" framework, dynamic IR aims at optimizing the overall search session (from both a short and long-term point of view). Indeed, particularly effective for complex tasks, the goal of dynamic IR is to leverage the user-system collaboration over the different stages of the search session (or within multi-search sessions) [28, 60]. With this in mind, one could say that the relevance estimation is therefore impacted by an additional feature, namely the longitudinal facet of the session S, leading to formalize the Retrieval Status Value function as follows: $RSV(d_i|q_j, u_j, S)$.

Beyond the user-system collaboration, another form of collaboration that could be considered in IR concerns the user-user interactions. The user-user collaboration involves the consideration of social groups or communities, and could be found in the following research fields:

- Collaborative filtering [96, 76] in which the relevance depends on preferences of users with same interests,
- Social IR [3, 87] which takes into account users' interactions through the social network and social evidence sources, such as the users' betweenness or authority,
- Collaborative IR (CIR) [27, 30] in which several users work together with the goal of solving a complex information need.

While these two first approaches are characterized by an implicit collaboration leveraged to retrieve search results for a unique user, CIR is driven by an explicit collaboration framework in which user-user interactions contribute to enhancing the quality of the overall search effectiveness [23], as done in dynamic IR. CIR, which is the field we review in this paper, refers to the act of searching and retrieving information from an algorithmical point of view [105]. As shown

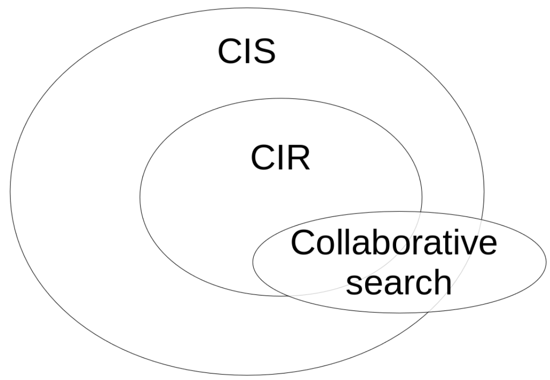

Figure 1: CIR in picture: a positionning with respect to CIS and collaborative search, inspired from [106]

in Figure 1, CIR is encapsulated within a broader field, namely Collaborative Information Seeking (CIS), which explores the search process regarding users' behaviors, as well as the information side. The CIS field covers a large range of search processes (searching, retrieving, browsing, sense-making, ...) which could be analyzed according to the behavioral point of view [35], or enhanced through algorithmic CIR approaches [89]. Both CIR and CIS take place in a group-based framework in which users collaborate to solve a shared goal. In the remaining paper, we use the term CIR in a broad sense, including also the particularly focus on the algorithmic side of collaborative search support.

With this in mind, CIR has been deemed as beneficial for solving complex or exploratory search tasks [27]. The benefit of such setting is that it enables to gather complementary skills to overpass the lack of knowledge of a unique user [128, 18, 105]. Collaborative search can be found in several application domains, e.g., medical, library, e-discovery, and academic. For instance, the collaborative process in the medical domain is characterized as a complex process due to the heterogeneity of patients, treatments, and the wide range of knowledge expertise related to the health domain [27, 48, 78]. Accordingly, a need has been risen toward collaborative information systems, such as CureTogether[1], PatientLikeMe[2], or EMERSE[3], enabling the exchange between patients, physicians, and health care workers [48, 78]. Collaboration is held in the context of the triangulation involving patients, physicians, and the web [129] or also between members of the health care team [48, 95]. The e-Discovery field concerns the management of electronic documents produced by organizations or companies in prevention of civil/criminal litigation or government inspection [16]. The

[1] http://curetogether.com/
[2] http://www.patientslikeme.com/
[3] http://project-emerse.org/

complexity of this management, despite high skills and qualifications of lawyers, lies in the necessity of a collaboration between stakeholders (companies, lawyers, and government inspections) and customers to identify privileged documents, to discuss sensitive issues, and to develop mutual awareness [5, 138]. Also, similarly to the diversity of application domains, collaborative search might include a heterogeneity of retrieved information, such as web pages [38, 89] or images [82, 81, 113].

Whether collaboration occurs in an application domain or for solving complex information needs, two main approaches of collaboration mediation arise [62]. The first one, more user-oriented, consists in adapting search interfaces to the multi-user context by supplying specific devices favoring exchanges, such as interactive tabletops [113] or shared workspaces with communication tools [101, 34]. The second approach, more system-oriented, relies either on (a) an algorithmic mediation [89, 25, 119] attempting to rank documents according to users' actions, characteristics, or roles, or on (b) the exploitation of IR techniques, such as clustering models to distribute documents among collaborators. The goal of these collaborative retrieval models is to favor a synergic effect within the overall search session through the optimization of users' actions ensuring that *"the whole is greater than the sum of all"* [105]. More particularly, this mediation is based on three main concepts [83, 26]:

- The *awareness* [20, 110] allows collaborators to perceive other collaborators' actions, and accordingly, facilitates the coordination of the group members. This concept is mainly dealt within the HCI field through the design of collaboration-devoted interfaces.
- The *division of labor* [26, 66] enables to split the task among collaborators to save time. The division of labor can be performed (a) at the user level [89] with distinct roles, or (b) at the document level through the exploration of different document subsets [25].
- The *sharing of knowledge* [133, 26] favors the information flow among collaborators using shared workspaces and communication tools [90], or through ranking models taking into account relevance judgements of the whole set of collaborators [25].

In this survey, we present a systematic review about CIR models and ranking algorithms including underlying impacting factors (e.g., role or expertise) by particularly focusing on the text retrieval task. Our primary objectives are specifically (a) to get a better understanding of the relevant factors that could be leveraged to optimize the collaborative search session effectiveness and (b) to propose a broad review of CIR models. It is worth mentionning that we do not focus on the evaluation framework in CIR since a relevant review is presented in [107]. While a significant amount of state-of-the-art work has been done in collaborative information seeking (CIS) in order to highlight the general context and the collaboration dimensions [79, 83, 105, 106, 107, 41], we deepen the literature review by focusing on the collaboration support in collaborative information retrieval (CIR) through algorithmic approaches. More particularly, in contrast to [79] who surveys search practices collected by means of questionnaires, we propose here to synthesize search practices observed in real user studies so as to highlight relevant factors for CIR models. Moreover, unlike work [83, 105] that presents a general overview of CIS including its definition, its dimensions, its application domains, and the existing interfaces, we focus here on the retrieval side by reviewing models supporting collaborative search. Finally, while [41] mainly review emerging topics and new application domains (e.g., learning, work tasks, and technology use) of CIS, our goal is to open up promising research directions under-explored in CIR, specifically regarding the relevance factors and new CIR paradigms. To the best of our knowledge, this article is the first attempt in organizing previous work particularly dealing with the algorithmic side of CIR. More specifically, this survey investigates the following aspects:

- The different forms of collaboration in the IR field. We present in Section 2 the underlying approaches according to the collaboration dimensions.
- The empirical studies analyzing the different factors of collaborative search in order to highlight some clues in terms of model design. The challenges and existing work on the collaboration support in CIR are therefore discussed in Section 3.
- The two main lines of work supporting collaboration through an algorithmic mediation are reviewed in Section 4. For this purpose, we distinguish CIR ranking models based on a system mediation from those in which the mediation is guided by users.
- The promising research directions in the domain are presented in Section 5.

2 Collaboration and Information Retrieval

In this part, our objective is specifically to propose a synthesis of the different forms of collaboration in IR, whether based on user-to-system [132] or user-to-user interactions [96, 30, 87]. Then, considering the particular scenario of CIR defined in [27, 30], we will introduce a general framework in terms of session modeling.

2.1 The Different Forms of Collaboration

Two main types of collaboration arise from the literature review in IR depending on the interaction level. A first form of collaboration defined in the IR field relies on the interactions between a user and an IR system. Taking into consideration users' clicks [11, 47, 61] and users' interests and expertise [127, 131], interactive

IR has emerged as a first solution to integrate the user dimension within the short-term retrieval process.

Recently, [132] define dynamic IR as the process of exploiting and modeling users' feedback to anticipate future actions. The dynamicity in the search session is multidimensional, including users whose behavior might evolve, documents which content might change, and their context-related relevance with respect to the information need. To tackle this challenge, several approaches aiming at leveraging sequential actions are proposed using Markov Decision models (POMDP) [139, 45, 75], pattern detection [97], or learning-to-rank methods [134, 50].

Orthogonally to interactive and dynamic IR, CIR also integrates the user-system collaboration for both ensuring the document relevance at the query level [89, 25] and optimizing the synergic effect of the search session [109].

In contrast or in complementarity, some IR approaches rely on a user-user collaboration, leveraging interactions of users belonging to groups/communities to enhance the retrieval effectiveness. Three main research fields in IR are based on a user-user collaboration:

- Collaborative filtering (CF) [96, 47, 72], also called recommendation, aiming at personalizing the search in response to a need expressed by a single user leveraging data generated implicitly by other users. The underlying intuition remains on the fact that users with similar profiles might be interested in the same or similar documents/items.

- Social IR (SIR) [3, 65, 15, 6] based on the analysis of social networks by modeling interactions between users and leveraging social indicators related to documents and users so as to improve the estimation of the document relevance [49].

- Collaborative IR (CIR) [128, 62] aims at solving information need shared by a set of users who are actively engaged in the search session through user-system and user-user interactions. The user-user interactions are prevalent here since they allow to both structure the collaboration [66] and to enrich the sensemaking process leading to a collective response with respect to the shared information need [25].

One main distinction between CIR and all other IR fields based on collaboration, whether user-system or user-user, depends on the user engagement, namely how the users are involved in the collaborative search process. Indeed, in CIR, users explicitly collaborate to solve a shared information need while other IR fields aim at leveraging implicit actions of users (collaboration and intent dimensions). Another characterization of collaborators' engagement is the level of their implication within the task, namely whether users are (a) active by performing explicit relevance feedback through annotations, bookmarks, and information exchanges (as leveraged in CF, SIR, and CIR) or (b) passive by only

Table 1: Synthesis of five main collaboration-based fields in IR

		Interactive IR	Dynamic IR	Collaborative filtering	Social IR	Collaborative IR
Collaboration	user-system	■	■	■	■	■
	user-user	□	□	■	■	■
Intent	implicit	■	■	■	■	□
	explicit	□	□	□	□	■
Implication	active	□	□	■	■	■
	passive	■	■	□	■	■
Mediation	user-driven	□	□	□	■	■
	system-mediated	■	■	■	■	■
Concurency	synchronous	■	■	□	□	■
	asynchronous	□	■	■	■	■
Location	co-located	□	□	□	□	■
	remote	□	□	■	■	■

submitting queries and reading information. In this last setting of implication, there are still relevance feedbacks that could be collected through submitted queries and visited pages, however the user engagement is less strong. This is generally the case for interactive and dynamic IR.

Other dimensions (e.g., the mediation level, the concurrency, and the location) distinguish these collaboration-based IR domains [30, 12, 105], as synthesized in Table 1. More particularly, the mediation level expresses the collaboration means between collaborators. We identify (a) the user-oriented mediation in which the entire achievement of the search session relies on the user (for instance, through search interfaces) and (b) the system-based mediation in which the collaboration is supported by algorithmic approaches. In the context of CIR, the mediation at the user level is generally supported by search interfaces, such as SearchTogether [80] or Coagmento [109], while the system-level restricts users' actions by providing rankings [25] or query suggestions [89]. The consideration of the user-driven mediation is a particular setting in CIR in which the system lets the user performing his/her task and applies dynamically the best suited adapted retrieval techniques (e.g., automatically expanding the query with specific words or suggesting query reformulations) with respect to his/her behavior. This setting has been recently introduced in CIR [118]. Assuming that other five fields generally adopt a constant retrieval strategy, they refer to a system-based mediation.

Another interesting dimension in CIR concerns the spatio-temporal axis characterizing the type of collaboration between users during the search. The first one denotes the concurrency of collaborators' actions, depending on whether users work simultaneously or not; this refers respectively to synchronous or asynchronous search [30]. The second aspect is the location of users within the collaborative process. If users are close to each other, the collaboration is collocated, whereas if users are in different places, this setting expresses a remote collaboration. Assuming that interactive and dynamic IR are characterized by a user-system collaboration, the location dimension is not relevant for those fields. However, the concurrency of user and system actions is synchronous for an immediate impact and, in the case of dynamic IR focusing on a long-term actions, might be combined with an asynchronous collaboration. For CF and SIR, the collaboration is generally asynchronous and remote while in CIR, all settings are possible.

2.2 The CIR Framework

CIR has been defined as a complex setting involving a group of users interacting each other in order to solve a shared information need [27]. Although three phases (before, during, after) in collaborative search have been highlighted [23], this is the "during" one which is generally prevalent in IR research. In the light of dynamic IR, CIR leverages collaborators' actions to enhance the overall performance of the search. Accordingly, a CIR setting is characterized as a search session which could be defined as a set of successive actions and interactions of collaborators' while seeking relevant information to solve a shared information need [23]. Collaborative search could be seen as an interactive environment in which two main categories of interactions co-exist [23, 88, 125]: (a) user-document interactions allowing collaborators to deepen their understanding of the topic; and (b) user-user interactions for structuring the collaboration and exchanging and organizing information. For instance, [137] outline different influence factors impacting the query reformulation. They also suggest that pages visited by a user and those judged as relevant by his/her collaborator(s) may impact on his/her own reformulated queries and increase the technicality or the diversity of the vocabulary.

In this context, [25] present a model of search session that involves, as illustrated in Figure 2:

1. A set of successive queries, as well as their temporal metadata, reflecting

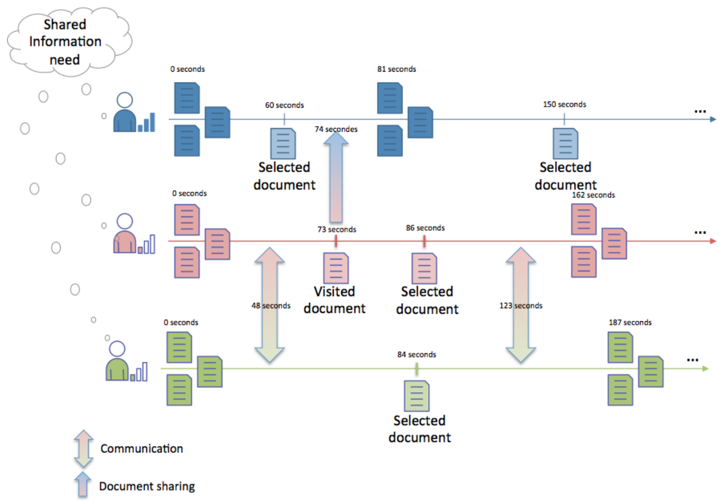

Figure 2: Example of CIR setting involving three users

reformulations of the shared information need.

2. A set of visited pages and other relevance judgments, with their temporal metadata, which denotes the users' assessments with respect to the relevance of documents towards the shared information need.

3 Towards the Integration of Collaboration in Document Ranking Models

The consideration of the collaborative aspect in IR models is not a simple task relying on basic IR techniques aggregating rankings of individual users. Indeed, the objective is rather to optimize the collaboration over successive rankings to ensure a synergic effect of search sessions while guaranteeing mutually beneficial goals. More particularly, the range of factors impacting on relevance is more important than for individual IR. For instance, user-system interactions might be enhanced by the consideration of user-user interactions involving communication, search strategies, or collaboration coordination. Accordingly, we believe that empirical studies provide a useful framework for highlighting relevance factors and search patterns that could be integrated within CIR models, and thus provide insights for the design of CIR models. In this section, we present an overview of the collaborative search framework to highlight, through empirical studies, the impacting factors that should be considered in the formalization of CIR models.

3.1 Empirical Understanding of Collaborative Search

Empirical studies [55, 109, 137] are a well-known framework in IR for understanding users' behavior, exploring the impact of features, and testing the effectiveness of ranking models [21]. In our particular context of explicit collaborative search, those studies have been used, for instance, to identify (a) the value of collaboration with respect to individual scenarios, (b) the manifestation of the collaboration in search tasks, and (c) the impact of different collaborative settings on the overall effectiveness of the search session. Following some collaboration dimensions based on the level of interactions (user-system and user-user), the spatio-temporal context (colocated/remote, synchronous/asynchronous), and the collaboration mediation through the role factor, we present below three main categories of empirical studies.

3.1.1 Modeling User-System and User-User Interactions

The analysis of users' interactions is essential in CIR in order to outline first search patterns and, then, search behaviors and interactions impacting the retrieval process.

The analysis of user-system interactions is a well-known research area since several behavioral models (e.g., ISP [70], ASK [7], the IS&R [57], or the Ellis model [22]) have been proposed for individual search. However, fundamental differences exist between individual and collaborative search sessions [62, 109, 9, 64] according to the diversity of the vocabulary, the interaction mode and the search effectiveness. Moreover, a collaborative search setting involves additional difficulties due to the important place of communication between collaborators, the complexity of the shared information needs, and the cognitive load of the sense-making process.

With the goal of building a picture of the collaborative search session stages, some authors [54, 55, 108] test the feasibility of the ISP model [70] in the collaborative context. On the one hand, Hyldegärd [54, 55] demonstrates that, although similarities with the ISP model exist between individual and collaborative search, some additional social and cognitive dimensions underlying the collaborative setting should be considered.

On the other hand, [108] perform a user study consisting in an exploratory task in which the process stages were quantitatively measured using user-system and user-user interaction features: (a) Initiation: number of chat messages before the stage initiation and between the different stages of the session, (b) Selection: number of chat messages discussing the strategy, (c) Exploration: number of search queries, (d) Formulation: number of visited webpages, (e) Collection: number of collected webpages, and (f) Presentation: number of moving actions for organizing collected snippets. The log analysis outlines that five stages of the ISP (namely, Initiation, Selection/Exploration, Formulation, Collection, and Presentation) could be clearly identified. The identification of these stages suggests that the ISP model is reasonable for modeling collaborative search. However, the study reveals that the Formulation, the Exploration, and the Collection stages are highly correlated (in terms of features) with quick switches between these stages. Both statements enable to infer that coordination-related stages are more prevalent in collaborative search than the search itself.

With a similar objective, [40] propose to map the collaborative search actions (split into document-related and human-related actions) with the IS&R generative behavioral model [57]. The latter is characterized by three main levels: (a) the work task level including the task initiation, the task preparation and the task completion; (b) the information seeking task level and (c) the information retrieval level. On the basis of a diary study involving nine patent engineers during two months, results show that the IS&R model could be applied to collaborative search and that the task preparation and the information seeking stages are prevalent in this context. In the information seeking task level, human activities (e.g. communication) constitute a high proportion of behaviors leading to retrieval-related actions, such as the query reformulation, that are performed before the user-document interactions.

While work presented above proposes to map collaborative models to individual search patterns, [126] analyze communication-based interactions between group of three users to identify collaborative sensemaking processes. Similarly to the IS&R model, findings outline that the latter covers the actions of structuring the task, searching for information, sharing information, and synthesizing information.

In a more abstract point of view, [136] exploit users' interactions to uncover the hidden Markov model underlying collaborative search. A comparative analysis between individual and collaborative search logs outlines that (a) collaboration is a more complex task since it requires a higher number of hidden states for modeling users, and (b) that collaboration involves more attention to sensemaking through communication-based interactions which occur throughout the collaborative search process. Interestingly, the study analyzing individual and collaborative scenarios reveals that individual search is very close to the ISP model [77] while collaborative search rather fits with the social search model presented by [23].

Besides the search pattern analysis, others studies focusing on search behaviors and interactions allow to understand the relevant factors impacting IR-related behaviors, such as query reformulation or relevance assessment. For instance, [137] investigate the process of query reformulation performed during two types of collaborative search session: an exploratory search task, namely an academic literature search, and a fact-finding task, namely a travel planning task. The authors first categorize the user's actions that are closely related to search and those that are related to search and collaboration. Using the log data, they quantitatively measure the query reformulation process with the aim of identifying the query term sources.

Results outline that query reformulations are mainly influenced by previously viewed/saved documents, as already noticed in individual search. However, results show that actions of collaborators also impact the query reformulation process, although the search is not performed on a shared workspace, limiting the collaboration coordination. Indeed, collaborators' queries and search histories have been highlighted as good evidence sources since, in average, at least one term included in collaboration-based actions have been identified in a user's reformulated query. Also, the chat analysis outlines that more than 60% of chat messages are connected to the query submission/reformulation process, leading to the exploration of subtopics or new topics.

3.1.2 Studying the Impact of Time and Space on Collaborative Search Performance

This second category of empirical studies highlights the impact of the spatial [109, 33] and temporal [32, 43] factors on the collaborative search effectiveness.

Concerning the spatial factor opposing remote and colocated search, [33] study the effect of the collaborators' location and the communication channels (text or audio-based instant messaging systems). Results highlight that, although the effectiveness is not significantly different between all studied settings (colocated/remote and text/audio-based communication), colocated groups seems to perform tasks with the lowest level of diversity, more particularly for the query reformulation process. Moreover, combining text and audio communication is more effective while a text-only communication leads collaborators to stop their search for interacting with their partners. Similar findings have been outlined in [109] with a particular distinction in terms of material environment for colocated users, working or not on the same computer. This environment variable is shown as an important factor in the search effectiveness, more particularly with respect to the recall value. This is explained by the nature of the designed task, which is exploratory and requires a wide topical range of documents. However, the authors underline that in the case of non-separable tasks, this setting would not be significant. Moreover, using coverage-based measures estimating the level of redundancy within the task [69], remote collaboration has been pointed as the best setting since it constrains users to perform useful interactions. This is more particularly interesting since the cognitive load of collaboration is not higher in this collaborative setting. We note that beyond the analysis of different collaborative settings, [109] outline the synergic effect of collaboration with respect to individual settings.

The concurrency of collaborators' actions is another search setting used in educational and organizational application domains [29, 32]. Similarly to the spatial constraint, the temporal factor is guided by the social presence theory [117] which strongly connects the social presence of collaborators to user-user interactions held through communication systems. In order to understand to what extent synchronous and asynchronous searches are different, [32] performed a user study which analyzes the communication, the search performance, and the cognitive load. Results reinforce previous statements about the importance of the communication exchanges between collaborators with a high number of messages and a good balance of messages within a collaborative group. Although a similar search effectiveness is obtained for both settings, asynchronous search is outlined as promising in terms of topic exploration, with a wider range of queries and a noticeable decrease of the cognitive load.

3.1.3 Investigating the Role-based Mediation in Collaborative Search

Several work [116, 66] has highlighted the benefit of leveraging collaborators' complementarity to enhance the collaborative effectiveness. With the purpose of structuring the collaborative search session, users' complementarity could be modeled through roles [66]. Some empirical studies [56, 123] analyze this factor in the collaborative context.

While previous work analyses intrinsic behaviors of collaborators, [123] focus on the comparison of users within a collaborative group. The goal of this study is to identify their respective roles, assuming that collaborators' complementarity could be beneficial for the collaboration coordination. With this in mind, the authors formulate two main research hypotheses: (a) collaborators behave differently from each other within a collaborative group, and, accordingly have different skills and might be complementary, and (b) collaborators' behaviors evolve throughout the search session.

In this context, the authors perform a user study, split into different settings depending on whether the pair of participants is guided by role guidelines (namely, prospector/miner roles [89] and gatherer/surveyor roles [111]) or not. Collaborators' behaviors are modeled through search features, such as the number of submitted queries or the number of visited pages. Three different analyses have been performed.
(a) The first one identifies behavioral differences between the different settings (with or without roles) and within the pair of roles. A more in-depth analysis of communication channels outlines that groups driven by predefined roles are more willing to exchange about document contents and the task topic while the other groups spent more time discussing about search strategies.
(b) The second analysis focuses on the complementarity identification through a temporal analysis of the correlations between the search feature differences of both collaborators during the search session. The obtained results reinforce the intuition that collaborators' search behaviors evolve throughout the search session. Moreover, collaborators not restricted to role guidelines are fastly able to coordinate their search strategies while this coordination is less obvious for collaborators restricted to gatherer/surveyor roles. Interestingly, the results also highlight a role drift for participants following the prospector/miner guidelines. These statements suggest that the role factor seems to constraint collaborators

in some skills that do not really fit.
(c) Combining these results with an effectiveness analysis, the authors observe that users without prior roles are able to bring out their skills through a user mediation without neglecting their effectiveness. Indeed, although users in both settings make fair judgments about relevance, users that define their roles themselves are more successfully willing to discard irrelevant page results because of a more in-depth reading. Accordingly, one interesting conclusion is that performing searches without prior roles may lead to more precision-oriented searches.

In the same spirit, [56] analyze the effect of explicit roles in the particular context of a travel planning task by comparing the search effectiveness and behaviors of collaborators involved in two settings (with or without roles). Considering the roles of *searcher* and *writer*, collaborators are free to choose their roles and swap them during the task. However, in contrast to [123] in which the search environment is similar between these two types of settings, the number of laptops varies according to the roles of collaborators (1 PC for groups driven by explicit roles and 2 PCs for non-explicit role-based groups). Results show that the role assignment, and accordingly the search environment, does not impact the search outcome of the collaborative travel planning task. However, similarly to previous work [123], groups with explicit roles seems able to quickly exchange valuable content-based information for solving the search task, while for non-explicit roles, the communication is rather oriented towards coordination or opinion confirmation.

3.2 Lessons Learned and Challenges in Collaborative Information Retrieval

According to the literature review of empirical studies related to collaborative search, one could highlight that the most common scenario is generally a small-group of users performing either exploratory or fact-finding tasks. In this context, several impacting factors have been highlighted, such as the search actions, the communication exchanges, the users' roles, or the spatio-temporal context. These observations allow to point out lessons for designing CIR models.

First, the collaborative setting implies the consideration of the group with multiple impacting factors on the search effectiveness; giving rise to coordination challenges according to both user-driven and algorithmic points of view.

Indeed, behavioral empirical studies have outlined that some phases of theoretical models of individual search are still valid for collaborative search [40, 137]. This suggests that well-known IR techniques, such as query reformulation, can be used in the collaborative context. However, behavioral models of individual search might be insufficient for modeling the collaborators' mediation since it requires the consideration of multiple factors. And conversely, collaborative mediation techniques (e.g., role-based mediation) might turn out to be ineffective in terms of document ranking. For instance, [56] and [123] point out that explicit roles lower the coordination cost and structure the collaboration, although not guaranteeing performance improvement in comparison to user-driven settings with implicit roles. One possible explanation might be that the role factor is not the only one that should be considered in CIR, suggesting that a multi-dimensional consideration of the collaborative context is desired. Accordingly, we believe that an hybrid approach combining the theoretical framework of individual IR with the paradigms of collaborative search would allow to build strong models for CIR. In other words, collaborative search is a complex setting involving several users, exchanging with each other, interacting with information sources, and in which the longitudinal aspect of the task is prevalent so as to accomplish a search session in a whole. More formally, this requires the integration of the "group-dependent user" as a novel variable in IR models, which is still challenging.

Second, as suggested in previous statements [109, 105], collaboration is effective in particular tasks depending upon complementary knowledge of collaborators to reach a synergic effect [109]. The task achievement is of great importance in collaborative search since it entails to aggregate individual actions of users to estimate the collective relevance of documents in the goal to build a collaborative response of the shared information need. Evaluation metrics relying on the notion of coverage [68] have been proposed and provides some guidelines for modeling the collective relevance. This new IR paradigm highlighting the estimation of the collective relevance is a consequence of the mutual beneficial goals underlying CIR [105].

With this in mind, we point out three main challenges:

1. *Learning from user-system and from user-user past interactions.* As an extension of the dynamic IR framework, collaboration requires the consideration of all types of interactions with the goal of optimizing the overall search session. However, while dynamic IR mainly focuses on user-system interactions, the collaboration setting opens new search activities, such as information sharing or sensemaking. Accordingly, IR models should consider the user-user interactions, standing through communications. In addition, interactions are time-related, leading to synchronous or asynchronous settings implying CIR models to support the temporal coordination of users' actions.

2. *Satisfying mutually beneficial goals.* In accordance to the collective relevance paradigm, the document relevance estimation in CIR might depend on the shared information need formulated by the whole group *(collective relevance)*. Indeed, beyond the users' satisfaction challenge considered in an individual setting, a CIR model might also deal with the collaborative group constraints in terms of the information need understanding and collaborators' coordination. One difficulty is to capture collaborators' intent to leverage from their perceived relevance within a retrieval model while considering their relevance feedback expressing their interests. More formally, we can argue that one emergent challenge underlying CIR concerns

the estimation of the collective relevance $RSV(d_i|q_j, u_j, g)$ of document d_i, that, beyond a personalized point of view (user u_j), integrates the group (g) as a dependent variable.

3. *Ensuring division of labor and sharing of knowledge.* In order to promote the synergic effect of collaborators within the search session, one challenge consists in leveraging from collaboration paradigms within the estimation of the collective relevance. Indeed, division of labor and sharing of knowledge are pivotal in collaboration since they allow to structure search and avoid redundancy between collaborators [66]. For instance, the former could be ensured at the document level by either splitting the document set [25, 119] or assigning tasks to collaborators according to role guidelines [89, 111].

4 Collaborative Information Retrieval Models

The goal of CIR is to support algorithmic mediation of users with the consideration of the collaboration features (e.g., role factor, interactions, time, and space constraints) through IR techniques and models [62]. In this section, we review the state-of-the-art CIR models to provide a picture of their hypotheses and peculiarities. In contrast to previous surveys focusing on the general context of collaborative search [79, 83, 105, 106, 41] and its evaluation [107], our objective here is to particularly point out the formal retrieval aspects through a synthesis of existing CIR models.

In what follows, we first present an overview of CIR models. Then, we describe previous work in CIR relying only on the system-based mediation or on a user-driven system-based mediation. For each model, we briefly discuss the experimental evaluation to present their advantages and limitations. All these models provide interesting peculiarities in terms of synergic effect with respect to individual settings. However, due to the complexity of evaluating real collaborative search tasks, experiments rely either on collaboration simulation or posterior log studies based on collaborative search tasks. Accordingly, the generability of the outlined statements are limited to the experimental setup, such as the task (exploratory search task in general) or the group size (two users for most studies).

4.1 Overview

Although relying on individual IR foundations, CIR is a young research area and has known a keen interest with the pioneers work of [89].

The proposed CIR models exploit user's search history data (e.g., click-through data or viewed SERPs) and user-user communication-based interactions to enhance the effectiveness of collaborative search sessions. To do so, two main mediation levels are used, as synthesized in Figure 3, leading to two categories of models. The first one relies on a system-based mediation which aims at harnessing collaboration directly within the document ranking step. In contrast,

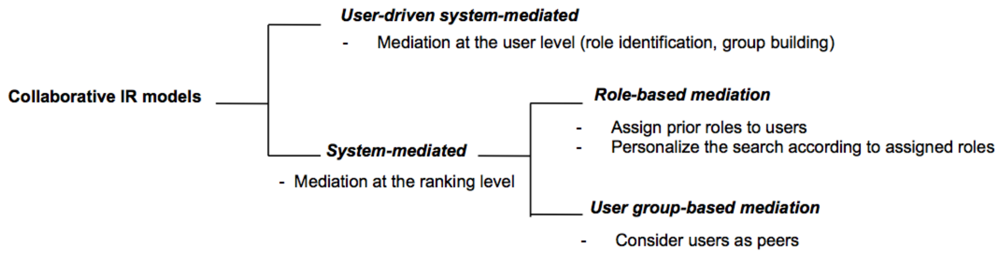

Figure 3: Taxonomy of CIR models.

the second category of models stands on a user-driven system-based mediation which occurs at the user level through, for instance, approaches building group or identifying roles of collaborators.

4.2 System-Mediated CIR Models

The system-mediated CIR models is based on the assumption that users need algorithmic approaches to mediate the document rankings to ensure division of labor and sharing of knowledge. This hypothesis enables to integrate the collaboration coordination directly within the document ranking step through, for instance, document distribution or query reformulation. More particularly, among system-mediated models, we can distinguish those that consider (a) users as peers and try to personalize the search according to their search history, or (b) users with fixed roles and try to personalize the search according to roles.

4.2.1 Group-based Mediation

The essence of these models is to consider users symmetrically (similarly to peer roles) [30]. Instead of taking into account users' search strategies, the division of labor is applied algorithmically through a distinction of the documentary spaces explored by collaborators to cover as much as possible the shared information need.

As a general framework of CIR models, [10] propose a cost model based on decision theory, called the collaborative-oriented probability ranking principle algorithm (cPRP). This cost function is equivalent to the collaborative-based probability ranking principle aiming at retrieving documents with respect to decreasing values of $p_{i,j}(1 - q_{i,\neg j})$ [98]. By taking into account the document discovery actions of collaborators, the authors argue that the exposed probability-based approach could be used as a justification of previous work in CIR (e.g., [25, 111]) ensuring division of labor. More particularly, the mediation is performed on the basis of collaborators' actions. It relies on the relevance probability $p_{i,j}$ of document d_i with respect to searcher u_j (namely, the collaborator who submits the query) and discovery likelihood $q_{i,\neg j}$ of document d_i given another collaborator $u_{\neg j}$. The underlying intuition is that the CIR task is recall-oriented due to the exploratory nature of underlying search tasks [111] and, accordingly, a division of labor between collaborators should ensure that a document retrieved for a team member should not be retrieved to another. With this in mind, the authors formulate a cost function of action α of user u_j given the selected document d_i:

$$EC(\alpha|d_i) = p_{i,j}[(1 - q_{i,\neg j})B + q_{i,\neg j}\bar{B}] \quad +(1 - p_{i,j})\bar{C} \tag{1}$$

$$where\ B = \mathcal{L}(select|relevant, not\ discovered) \tag{2}$$

$$\bar{B} = \mathcal{L}(select|relevant, discovered) \tag{3}$$

$$\bar{C} = \mathcal{L}(select|not\ relevant) \tag{4}$$

where $\mathcal{L}(a_1|s_1)$ is the loss function of the cost of performing action a_1 given state s_1.

The experimental evaluation performed through a user study highlights that:

- The query history seems to provide a better context in CIR than click-through history since collaboration involves several subtasks (in terms of topic exploration as well as actions) that could be captured by queries;
- The search context could be enhanced by the aggregation of the collaborators' search history, leading to the identification of the most relevant information.
- The consideration of user-user interactions through the exchanged message seems to be particularly effective in CIR. This is the case when the task requires an intensive collaboration with decision taken by a group (e.g., trip planning). In other tasks (e.g., exploratory search), the user-user interactions through chat messages rather offers a coordination means.

More oriented towards the estimation of the collective relevance of documents with respect to the information need, [25] propose a CIR model that aggregates the relevance judgments performed by the whole set of collaborators. More specifically, for a given query q, the aggregation of these collective relevance judgments is performed at the term weighting level. Therefore, the weight $purw(t_v)$ of term t_v refers to a partial-user relevance weighting and is estimated as follows:

$$purw(t_v) = log\frac{(\sum_{u_j=0}^{U-1}\alpha_{u_j}\frac{r_{u_j v}}{R_{u_j}})(1 - \sum_{u_j=0}^{U-1}\alpha_{u_j}\frac{n_v - r_{u_j v}}{N - R_{u_j}})}{(\sum_{u_j=0}^{U-1}\alpha_{u_j}\frac{n_v - r_{u_j v}}{N - R_{u_j}})(1 - \sum_{u_j=0}^{U-1}\alpha_{u_j}\frac{r_{u_j k}}{R_{u_j}})} \tag{5}$$

where n_v is the number of documents in which term v occurs, N expresses the number of documents in the collection. U is the set of users. $r_{u_j v}$ represents the number of documents including term t_v and assessed as relevant by user u_j. The number of documents assessed as relevant by user u_j is noted R_{u_j}. The coefficient α_{u_j}, also called authority factor, expresses the impact of user u_j in the term weighting, under the constraint that $\sum_{u_j=1}^{U} \alpha_{u_j} = 1$. The authority factor α_{u_j} can be (a) fixed within the experimental evaluation, or (b) dynamic according to the correlation between the weight of terms included in documents assessed by the user as relevant and the weights of terms in a set of relevant documents.

The term weighting function is used for two IR tasks:

- *The document ranking process* in which the partial-user relevance term weight is integrated in the probabilistic model proposed by [99]. The principle of division of labor is then ensured by retrieving only documents that have not been visited by other collaborators and/or that are not simultaneously displayed in their document lists.

- *The query expansion process* in which the partial-user relevance weight of terms are combined over all collaborators in order to obtain a partial-user offer weighting.

This model has been evaluated through the simulation of collaboration between pairs of users on the basis of individual search logs provided by the TREC Interactive [86]. The main results outline that division of labor allows to ensure diversity within search without discarding search effectiveness. Moreover, the aggregation of relevance judgments over the whole members, guaranteeing the sharing of knowledge among collaborators, has shown its effectiveness with respect to individual models. However, the absence of personalization of document rankings is a drawback limiting the impact of the model with respect to the challenge of the satisfaction of the mutually beneficial goals.

To tackle this personalization challenge, another approach is proposed by [84]. The authors present two collaborative ranking models based on the personalized score $perso(d_i, q_j, u_j)$, estimating the relevance of document d_i with respect to query q_j submitted by user u_j according to the user's interactions [127]. Below, we introduce these two ranking models that focus respectively on the user and the group level.

- *Smart-splitting.* The first model personalizes document rankings according to each individual user among all the collaborators. For this purpose, a personalized score is assigned to each document-query pair. Document d_i is assigned to collaborator u_j who obtains the highest personalized score $perso(d_i, q_j, u_j)$ over all the users of collaborative group g:

$$\forall d_i,\ \exists\ u_j^*;\ u_j^* = \arg\max_{u_j \in g} perso(d_i, q_j, u_j) \qquad (6)$$

- *Groupization.* The second model aims at building a collective response to the shared information need over all collaborators' relevance assessment.

Personalized scores $perso(d_i, q_j, u_j)$ are aggregated over all collaborators $u_j \in U$ to estimate the collective relevance of each document d_i. In addition, this score is linearly combined (α refers to the weighted coefficient) with the original document rank, noted $rank(d_i)$, to preserve the most important information. The final score $s_{coll}(d_i, q_j)$ of document d_i for query q_j is estimated as follows:

$$s_{coll}(d_i, q_j) = \alpha \sum_{u_j \in U} perso(d_i, q_j, u_j) + (1 - \alpha) rank(d_i) \qquad (7)$$

These models are integrated into the SearchTogether interface [80]. The experimental evaluation validates these models with log files of collaborative search tasks. Results show the synergistic effect of both models. Specifically, the "smart-splitting" ranking function allows a more effective division of labor among users at each query submission throughout the search session while the "groupization" algorithm could be assimilated to the sensemaking process aiming at estimating the collective relevance of documents.

While previous work focused on leveraging collaborators' relevance feedback, a recent model proposed by [38] opens the range of evidence sources to contextual data. Indeed, in the line of the contextual IR domain which generally exploits individual search history, the authors argue that an effective contextual support for CIR would be the collaborator's data (queries, results, bookmarks) and also collaboration behavior based on communication. Depending on the sources and types of user's histories, a 3-dimensional context can be built:

- *Individual search history* H_{QU}: each collaborator's self-search history (e.g., queries, SERPs, or bookmarks)
- *Collaborative group* H_{CL}: group search history (e.g., queries, SERPs, or bookmarks)
- *Collaboration* H_{CH}: collaboration behavior chat (communication)

The context of a user is modeled through a contextual language model θ_{H_x}, where $H_x \in \{H_{QU}, H_{CL}, H_{CH}\}$, estimated as a unigram language model. With this in mind, the contextual probability $p(w|H_x)$ of word w is obtained as follows:

$$p(w|H_x) = \frac{1}{K} \sum_{k=1}^{K} p(w|X_k) \qquad (8)$$

$$p(w|X_k) = \frac{c(w, X_k)}{|X_k|} \qquad (9)$$

where K expresses the number of user histories (e.g., the number of queries or the number of chat messages) of type $X \in \{QU, CL, CH\}$. The number of occurrences of word w in user histories X_k is noted $c(w, X_k)$ and $|X_k|$ is the word count in X_k.

A re-ranking process is then performed by computing the KL divergence between the traditional document language model θ_d [91] and the language model of each context type:

$$D(\theta_d, \theta_{H_x}) = -\sum_{w \in q} p(w|\theta_d)\, log\, p(w|H_x) \tag{10}$$

In order to rank documents with respect to user's contexts, a learning to rank algorithm is used based on two main features: prior document rank and the KL divergence values $D(\theta_d, \theta_{H_x})$ between document d and the collaborative contexts H_x.

4.2.2 Role-based Mediation

The approaches presented previously consider users as peers, having similar search strategies and objectives with respect to the shared information need. However, other work [89, 111, 119, 120, 118] assumes that users might act according to asymmetric roles. These roles attempts to structure and organizes the group members in the collaborative process [66]. In contrast to previous CIR models in which the number of collaborators is unlimited, the models presented below are generally based on a pair of users. In this context, the latter are mostly identified by roles belonging to a taxonomy [30] that describes the potential roles of users in a CIR context, namely:

- The role of *Peer* denoting collaborators acting independently to each other and combining their search results manually. Collaborators' roles are, therefore, symmetric, allowing users to freely drive their own session.
- The roles of *Domain A expert/Domain B expert* highlighting collaborators with different domain expertise.
- The roles of *Search expert/novice or Domain expert/novice* referring to the expertise level of collaborators in an application domain or search skills.
- The roles of *Search expert/Domain expert* characterized by the type of contribution to the search process is asymmetric. The search expert is more willing to select documents and formulate queries while the domain expert has a higher ability to assess the document relevance with respect to his/her background.
- The roles of *Prospector/Miner* for which the goal of the former is to favor the search diversity whereas the Miner is interested in the document relevance and explores those related to the main topic.

In what follows, we distinguish three main types of asymmetric roles, as considered in CIR models. First, we introduce models based on roles guided by distinct search strategies [89, 111]. Then, we present CIR models structuring collaboration through domain expertise-based roles [119, 120].

A) CIR models relying on search strategy-based roles In this category of models, roles are used to split tasks among collaborators. The objective is therefore to divide the labor with complementary search actions by, for instance, favoring result diversity (*prospector* role) while ensuring that the main aspect of the shared information need is covered (*miner* role). We review here CIR models based on this type of roles.

[89] propose a collaborative ranking model based on the asymmetric roles of *prospector* and *miner*, defined in the role taxonomy [30]. The former, namely the *prospector*, favors the diversity in the search results by opening new exploration fields in the information space, while the latter, namely the *miner*, ensures the quality and richness of the explored documents. To model these roles, the authors take into account the previously ranked documents as well as collaborators' relevance judgments. Both roles rely on the assumptions that if rankings include a lot of relevant documents and few documents not already visited, the exploration track is likely effective and should favor search diversity and relevance towards the query. Thus, the authors identify two factors, namely relevance and freshness, estimated for document list L_j, retrieved for query q_j:

- The relevancy factor $w_r(L)$ measures the ratio of the number of relevant documents in retrieved list L to the number of irrelevant documents in list L.
- The freshness factor $w_f(L)$ measures the ratio of the number of documents not already visited in list L to the number of documents visited in list L.

For each role, the CIR model includes a ranking function in adequation with the role search strategies:

1. The mediation function connected to the role of *prospector* aims at supporting the diversity of the explored search results and consists in a term suggestion for the query reformulation. For each term t_v from all documents in all retrieved lists $\mathcal{L}$ to both users u_j and $u_{j'}$, a score is estimated as:
$$score(t_v) = \sum_{L \in \mathcal{L}} w_r(L) w_f(L) rlf(t_v, L) \tag{11}$$
where $rlf(t_v, L)$ represents the number of documents in list L that include term t_v.
2. The role of *miner* requires to look deeper into the document content and to identify those that fit the most with the shared information need. Thus, documents d_i not assessed as relevant by the *Prospector* are re-ranked and assigned to the *miner* according to a score estimated as follows:
$$score(d_i) = \sum_{L \in \mathcal{L}} w_r(L) w_f(L) borda(d_i, L) \tag{12}$$
where function $borda(d_i, L)$ expresses a document voting score.

Two prototypes, namely *Cerchiamo* [2, 29] and *Querium* [19], integrate this model based on the roles of *prospector* and *miner*. Relying on the same interface functionalities (query submission, document comments, and annotations), *Cerchiamo* and *Querium* are distinguished by the nature of the exploited information since the former is devoted to identifying videos while the latter aims at retrieving textual information.

Experiments are carried out through a user study performed on the Querium prototype with the exploratory information need issued from the TREC Interactive dataset. Similarly to previous models, results outline the synergic effect of the prospector-miner model compared to an individual scenario. Moreover, an analysis of the search outcomes with respect to the topic difficulty reveals that this model is particularly effective for solving difficult topics.

[111] propose a CIR model relying on the following pair of roles: (a) the *gatherer* which aims at quickly selecting relevant documents; and (b) the *surveyor* which has the objective to cover a wide range of results to better understand the nature of the information need to explore the potential exploratory fields as well as detect why queries are not optimal. These roles are complementary since the *gatherer* can search relevant information alone, but the *surveyor* requires the collective intelligence for a topical diversity for a better understanding of the information need. The mediation between these two roles is based on a merging and splitting of search results retrieved for two queries, respectively submitted by both collaborators. More specifically, the model responds to queries submitted by both collaborators. Let's consider query q_j (resp. $q_{j'}$) submitted by user u_j (resp. $u_{j'}$) and the associated retrieved list L_j (resp. $L_{j'}$) of documents. The model proceeds in two steps:

1. *The merge step* which builds a unified and ranked list $L_{jj'}$ of documents by merging both lists L_j and $L_{j'}$ through the CombSUM function that respectively normalizes the score $RSV(d_i, q_j)$ and $RSV(d_i, q_{j'})$ for each document d_i retrieved in lists L_j and $L_{j'}$. The intuition behind this step is that the combination of the two ranked lists of documents increases the effectiveness of the unified list $L_{jj'}$ since it aggregates the scores $RSV(q_j, d_i)$ and $RSV(q_{j'}, d_i)$ of document d_i, respectively obtained in response to queries q_j and $q_{j'}$.

2. *The split step* relies on a classification algorithm, namely the 2-means one, applied to the merged list $L_{jj'}$. This results in two document classes associated to the *gatherer* and the *surveyor* roles according to the following criteria:

    (a) The class with the highest centroid enables to build the list L_{Gath} retrieved for the *gatherer*.

    (b) The other class, with the lowest centroid, enables to build the list L_{Surv} retrieved for the *surveyor*.

Experiments focus on determining the optimal functions of fusion and separation, namely CombSUM and the classification algorithm k-means. The analyzes also compare the model with an individual scenario in order to validate its synergic effect.

B) CIR models relying on domain expertise-based roles Generally speaking, two users, involved in the same search session, are characterized by a relative difference of domain expertise level towards the shared information need. With this in mind, Soulier et al. [119, 120] consider collaborators' roles based on different domain expertise levels, as suggested in the role taxonomy [30]. More particularly, the authors distinguish two dimensions of the collaborators' expertise:

- A vertical distinction which assumes that one collaborator has more expertise than the other one. This assumption is reinforced by search behavior analysis [131] which has highlighted differences between *domain experts* and *domain novices*. One main divergence is that *domain experts* are more familiar with technical vocabulary while *domain novices* who need more term suggestions for getting a better insight of the domain [51, 131].
- A horizontal distinction in which collaborators are seen as a group of experts with different knowledge expertise and points of view with respect to the same information need. More particularly, this setting is used for solving multi-faceted information need in order to leverage the users' different knowledge expertise and assign them implicit knowledge-based roles towards at least one query facet.

Concerning the vertical dimension of the domain expertise resulting in the expert and novice roles, [120] propose a two-step CIR model for ranking documents according to the domain expertise-based roles. This approach assumes that the most experienced user for the query topic would most likely be interested by documents with a high level of (a) specificity [51, 131] and (b) novelty with respect to the user's domain knowledge [114]. This model introduces a generic approach which first estimates the role-based score of each document-user pairwise and then, distribute documents to the most likely fitted user. The first step aims at estimating scores $P(d_i|u_j, q_j)$ of documents d_i with respect to the query q_j and the role of each collaborator u_j. These scores include a personalized probability $P(\pi(u_j)|\theta_{d_i})$ which integrates document specificity $Spec(d_i)$ [67] and novelty $Nov(d_i, \mathcal{D}(u_j))$ towards the set $\mathcal{D}(u_j)$ of document already selected by user u_j [13] within a language model smoothing.

$$\begin{aligned} P(\pi(u_j)|\theta_{d_i}) &= \prod_{(t_v, w_{vj})\in\pi(u_j)} [\lambda_{ij} P(t_v|\theta_{d_i}) + (1-\lambda_{ij}) P(t_v|\theta_C)]^{w_{vj}} \quad (13) \\ with\ \lambda_{ij} &= \frac{Nov(d_i, \mathcal{D}(u_j)) \cdot Spec(d_i)^{\beta}}{\max_{d_{i'}\in\mathcal{D}} Nov(d_{i'}, \mathcal{D}(u_j)) \cdot Spec(d_{i'})^{\beta}} \end{aligned}$$

where β depends on the users' role and is respectively 1 and -1 for the expert and the novice. θ_{d_i} and θ_C are the parameters of the language model of document

d_i and collection C, w_{v_j} represents the weight of term t_v in user profile $\pi(u_j)$. λ_{ij} is the smoothing parameter.

In order to optimize scores previously estimated over all collaborators, the second step consists in a document allocation to user roles using the Expectation-Maximization (EM) learning method [17]. Relying on the document relevance scoring, this step assigns documents to the most likely suited user. Finally, division of labor is reinforced by ensuring that currently displayed document lists do not include the same documents.

The model is evaluated through a collaboration simulation-based framework as done in [25]. The model is pointed out as effective for both relative and absolute expertise within the group, in other words, when both (a) users have different levels of expertise, without being necessarily identified as expert or novice and (b) users are clearly labeled as expert and novice. A deeper analysis at the role level outlines that the model is able to improve the search experience of novice by displaying document lists more specific over the search session.

In another model relying on a horizontal distinction within collaborators' domain expertise levels, [119] propose to support collaboration between a group of domain experts aiming at solving a multi-faceted information need. This approach allows to leverage users' different knowledge expertise and assigns them implicit knowledge-based roles towards at least one query facet. These facets are modeled through document and user topical-based representations using the LDA generative model [8]. The proposed algorithm includes two main steps. The first one estimates the document relevance according to each expert with respect to his/her facet expertise and the shared information need. For this purpose, the authors combine (a) the document relevance probability $p(\pi(u_j)|d_i)$ with respect to the user profile $\pi(u_j)$ and (b) the document relevance probability $p(q_j|d_i)$ depending on the BM25 score and a LDA-based score [42]:

$$p(q_j|d_i) = \lambda RSV_{LDA}(q_j|d_i) + (1-\lambda) RSV_{BM25}(q_j, \theta_{d_i}) \tag{14}$$
$$with\ RSV_{LDA}(q_j|d_i) = \prod_{t_v \in q_j} \sum_{t=1}^{T} p(t_v|t).p(t|d_i)$$

where $p(t_v|t)$ represents the probability of term t_v given topic t over set T of topics. $p(t|d_i)$ is the probability of topic t given document d_i.

Similarly to the previous algorithm, documents are allocated to the best-suited experts using the Expectation-Maximization algorithm and retrieved to collaborators by ensuring the division of labor policy.

The experimental evaluation follows a similar framework than the one used for the previous model based on expertise [120], with a small variation in the group building. Indeed, this model is not restricted to a pair of collaborators since it allows larger groups in which the topical expertise-based role of users are identified according to clickthrough data. Results outline the search effectiveness of the model with respect to individual-oriented baselines. Moreover, results highlight the model robustness for all sizes of the groups, ranging between 2 and 6 users. This result is promising and shows that collaboration could be effectively mediated by algorithms, even though for larger groups.

4.3 User-Driven System-Mediated CIR Models

In this second category of work, we propose to review algorithmic mediations focusing on the user level. Instead of ranking documents according to the division of labor and the sharing of knowledge concepts, another approach is to proactively help users to perform an effective search session by making inference on the basis of their search logs. This challenge is tackled in the literature according to two main approaches: the recommendation of users willing to collaborate [36] and the dynamic identification of collaborators' roles [118, 122]. We detail these two approaches below.

The first category of models provides an interesting insight on collaborative groups and how they could be built to perform an effective group. While previous work assumes that collaborative groups are already formed, building a collaborative group for a given task is an outstanding challenge. A first approach, called "pseudo-collaboration", has been proposed by [36] aiming at evaluating the collaboration opportunity between two users performing individual search tasks on the same topic. The strength of this model is to also capture the best moment in which two users should collaborate by measuring the benefit/cost ratio of the pseudo-collaboration.

To do so, a collaborative search session is simulated by aligning and combining individual search sessions performed by a pair of potential collaborators. Evaluation performance of the simulated collaborative search are estimated at different timestamps by considering two metrics: (a) the search effectiveness expressing the precision of information found and identified as useful; and (b) the search efficiency normalizing the search effectiveness by the number of queries to capture the cognitive effort of the user to identify useful information. The collaboration between those two users is then characterized as helpful if the ratio of both metrics (namely, search effectiveness and efficiency) is higher than a threshold. The pair of collaborators who maximized this ratio are then retained and one could easily imagine that collaborative ranking algorithms could thus be launched so as to mediate the collaboration between users.

Experiments of the model are carried out using a collaboration-simulation framework relying on individual logs in which the approach is compared with real collaborative search logs. Evaluation performances are estimated on the basis of well-known CIR evaluation metrics defined in [109]. Results outline that mediating collaboration between users allows to reach a synergic effect. Moreover, another interesting insight is that the effectiveness of pseudo-collaboration overpasses the one obtained by real collaborative groups. These two statements reinforce the intuition that collaboration is not always useful for a whole search session and that capturing the best collaboration timing is more effective. However, it is worth highlighting that the authors do not consider the issue of detecting users' interests, but rather assume that users have similar information need. This limitation is still challenging today. In addition, this

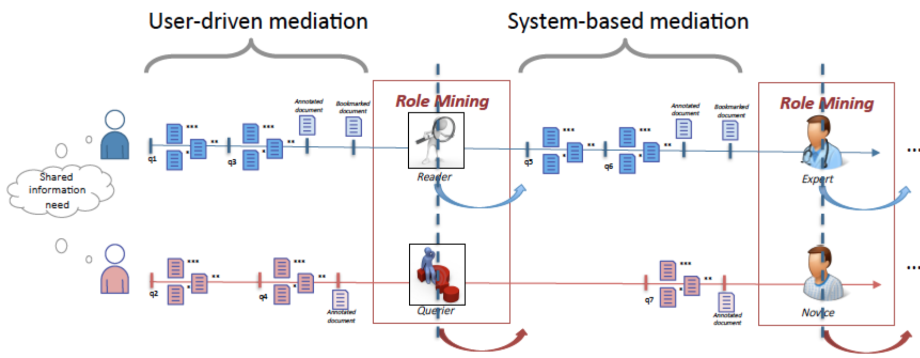

Figure 4: Hybrid search session guided by user-driven and system-oriented mediation

work presents several challenges such as the protection of users' privacy since pseudo-collaboration requires tracking users over time. Also, once two users have been identified as candidate collaborators, the challenge of connecting and coordinating these "strangers" for performing an effective collaborative search remains.

The second category of models [118, 122] is a variant of the group building ones since it assumes that collaborators already work together and mediate their search through the coordination of their roles. The proposed approaches rely on the assumptions that (a) skill or preference difference is one of the motivations that make users collaborate, and (b) collaborators' roles might evolve throughout the session [123] (in contrast to previous work based on fixed roles [89, 111, 119]).

With this in mind, [118] propose to leverage collaborators' real-time actions to identify the most suited roles of collaborators throughout the search session. This model opens a new vision of the collaborative search by letting the collaborators structure themselves their session and considering their actions to enhance algorithmically the retrieval effectiveness of the search session through a role mining approach. This refers to as user-driven system-mediated CIR, as shown in Figure 4.

In this context, the authors analyze how users are different, and, accordingly suggest roles given their complementarity that optimizes the collaborative outcome. The approach relies on a temporal representation of collaborators' search behaviors $S^{(t_l)}_{u_j}$ based on set F of search features f_k which estimates the cumulative value of each feature over each one-minute timestamp from the beginning of the search session until timestamp t_l. Differences $\Delta^{(t_l)}_{u_j,u_{j'}}(f_k)$ between collaborators with respect to the temporal representation of each search feature f_k are then estimated while complementarities and similarities/dissimilarities between participants are identified through correlation $C^{(t_l)}_{u_j,u_{j'}}(f_k, f_{k'})$ between search feature differences $\Delta^{(t_l)}_{u_j,u_{j'}}(f_k)$ and $\Delta^{(t_l)}_{u_j,u_{j'}}(f_{k'})$ of collaborators u_j and $u_{j'}$. The negative correlations computed between the feature difference values enable to bring out complementarities within search skills in which participants are the most effective.

Finally, the best suited roles with respect to a role pool are identified using a minimization problem estimating the distance between the correlation matrix $C^{(t_l)}_{u_1,u_2}$ of the collaborators and the role pattern $F^{R_{m,n}}$ of a predefined pair of roles $R_{m,n}$:

$$argmin\ _{R_{1,2}}||F^{R_{1,2}} - C^{(t_l)}_{u_1,u_2}|| \tag{15}$$

under the constraint: $\forall_{(f_k,f_{k'})\in\mathcal{K}^{R_{1,2}}}\ F^{R_{1,2}}(f_k,f_{k'}) - C^{(t_l)}_{u_1,u_2}(f_k,f_{k'}) > -1$

with $||.||$ representing the Frobenius norm.

The experimental evaluation of this model is carried out on search log of two collaborative IR user studies involving dyads. Results highlight the following trends. First, leveraging from role mining within a CIR session overpasses the aggregation of an individual scenario, showing the necessity of analyzing users' differences and mining their roles to optimize the collaboration. Second, the role mining-based CIR framework provides better results than scenarios relying on fixed roles. However, the effectiveness towards the pair of Prospector-Miner roles has not been demonstrated. One possible explanation might be that, similarly to our work, the Prospector-Miner-based ranking model also analyzes the retrieved documents and submitted queries from the beginning of the session for providing optimized document rankings. Third, the role mining methodology seems to be more effective than a scenario in which roles are assigned randomly. This emphasizes the reliability of the role mining approach which seems to accurately match users' search behaviors, and optimize the collaboration.

However, this previous approach has some limitations. Indeed, it might suffer from the fact that predefined roles do not allow to leverage the best of collaborators since roles might be under-constrained for their actions during the task. Accordingly, a very recent model, *MineRank* [122], overpasses the framework of predefined roles constraining users in search skills they might not really fit and focuses on latent roles of collaborators. At each query submission, the two-step model (a) evaluates the complementarity of collaborators, that could be assimilated to latent roles, in an unsupervised manner using various search behavior-related features for each individual involved; and (b) re-injects these latent roles to collaboratively rank documents.

The first step relies on assumptions formulated in [118] enabling to model complementarity of collaborators according to a correlation matrix $C^{(t_l)}_{u_1,u_2}$ according to a set of feature F. With this in mind, the complementarity is exhibited by the identification of the most discriminant features allowing: (a) to avoid redundancy between features measured through correlations $C^{(t_l)}_{1,2}(f_k,f_{k'})$ of features f_k and $f_{k'}$; and (b) to provide good indicators of the document assignment to users within the collaborative document ranking. The latter hypothesis is measured by the quality of the classification (namely, the recall $\mathcal{R}ec_{1,2}^{(t_l)}(f_k)$) performed over collaborators using the feature f_k as a criterion. This intuition can be translated into an optimization problem with multi-objectives determining the feature set F^p of size p and transformed into a unique objective optimization problem by linearly combining both optimized functions:

$$\begin{array}{lll} max_\alpha \sum_{k=1}^{n} \mathcal{R}ec_{1,2}^{(t_l)}(f_k) & \cdot\alpha_k - \gamma(\sum_{k=1}^{n}\sum_{k'=1}^{n} C_{1,2}^{(t_l)}(f_k, f_{k'}) \cdot \alpha_k \cdot \alpha_{k'}) & \\ \qquad\qquad subject\ to & \alpha_k \in \{0,1\};\ k = 1, ..., n & (16) \\ \qquad\qquad\qquad and & \sum_{k=1}^{n} \alpha_k = p & \end{array}$$

where α is the vector of size n where each element α_k is a Boolean indicator specifying whether feature f_k is included in the feature subset $F_{k1,2}^{(t_l)}$ at timestamp t_l. γ is a decay parameter expressing the level of behavior complementarity taken into account in the latent role mining algorithm.

In order to solve the optimization problem, the authors assume that the discriminant feature selection could be seen as an adapted maximum clique extraction algorithm, called Coll-Clique, using graph theory. The discriminant features identified through the Coll-Clique algorithm are then re-injected within a collaborative ranking relying on logistic regression classification. The idea is to use the most discriminant features to first assign documents to the most suited collaborator, and then ensure the division of labor principle.

The obtained results highlight three main contributions: (a) the MineRank model enables participants to benefit from the synergic effect of collaboration; (b) ranking documents with respect to meta-roles gives an additional value to a CIR model based only on the behavioral analysis of collaborators; (c) mining meta-roles for collaborators seems to be more effective than a CIR scenario in which roles are fixed throughout the search session. A more in-depth analysis enables to show that the meta-role varies between successive query submissions in the beginning of the session; reinforcing the need of mining role dynamically.

4.4 Synthesis and Recommendations

In Table 2, we sum up all collaborative-oriented algorithmic approaches presented beforehand (system-based and user-driven system-based mediation, noted respectively SBM and UDSBM). We characterize these models according to three dimensions, namely the relevance, the evidence sources and the collaboration hypotheses (division of labor and sharing of knowledge).

From a general point of view, most of CIR models aim at satisfying both individual and collective information needs. Indeed, their goal is to simultaneously build document rankings in response to a query issued by an individual user considering the group members through division of labor or sharing of knowledge. As outlined in the literature [25, 66, 107], the latter allow to avoid redundancy within the task or favor the information flow between users so as to enhance the synergic effectiveness of the collaborative group. The approaches proposed by [84] have two different goals, respectively focusing on the users' or the group satisfaction, and therefore are less oriented towards the satisfaction of mutual beneficial goals.

In terms of evidence sources, one could see that relevance feedback is the most used feature since it provides interesting insight on the collaborator's understanding of the information need, following paradigms underlying interactive and dynamic IR. Indeed, although collaborators are guided by a shared information need, they might have their own perception of the topic due to the diversity of knowledge, interests, or search skills. First, the role of collaborators is used in several work, whether fixed over the session [89, 104, 120] or dynamic [118]. As mentioned in [66], roles allow to structure and organize the search session, and constitute an important variable in the collaboration coordination. Although the behavior might be used in most of CIR work, this second evidence source is used differently in user-driven approaches [118, 36] so as to let the users being the owner of the search session achievement. This also open barriers of controlled search tasks that could be designed in role-based CIR models, limiting the range of collaborators' actions in skills/topics they might not exactly fit with [122, 123].

Concerning the collaboration hypotheses, one could see that division of labor is prevalent in CIR models [111, 10], probably because this is the easiest to model by avoiding overlapping between document rankings or assigning roles with distinct strategies. In contrast, sharing of knowledge [25] is less present since the information flow might be more difficult to formalized through IR techniques. Indeed, adapted search interfaces constitute a better environment since they allow to share documents or information with respect to the task.

To put these work into perspectives, we propose some recommendations in terms of model uses according to the search setting (e.g., group size, task, or user peculiarities).

As seen in empirical studies, collaborative search is mostly studied by considering pairs of users. Accordingly, several CIR models [89, 111, 120, 118, 36, 122] are designed for dyads. However, larger groups could be supported by CIR models as proposed in [25, 84, 119, 10, 38] with no restriction on the group size. CIR models based on the role of Prospector-Miner [89] or Gatherer-Surveyor [111] could be easily extended to groups with more than two users. For instance, in [111], the voting function estimating the document relevance could be computed over a large number of users. Similarly, the document redistribution carried out through a k-means is also possible by specifying $k > 2$, where k denotes the number of users.

Regarding the task, all CIR models are adapted for exploratory search. However, some work might be inappropriate for fact-finding tasks, such as travel planning ones, or online shopping [79]. Indeed, for those tasks, it is worth mentioning that additional support might be needed, such as a shared workspace. Accordingly, algorithmic mediations relying on search interface functionalities (such as the search history and communication [38]) harness the whole context

Table 2: Synthesis of collaborative ranking models.

		[25]	[84] “smart-splitting”	[84] “groupization”	[38]	[10]	[89]	[111]	[119, 120]	[36]	[118]	[122]
		SBM								UDSBM		
Relevance	collective	■	□	■	■	■	■	■	■	■	■	■
	individual	■	■	□	■	■	■	■	■	■	■	■
Evidence source	feedback	■	■	■	■	■	■	■	■	□	■	■
	expertise	□	□	□	□	□	□	□	■	□	□	□
	behavior	□	□	□	□	■	□	□	□	■	■	■
	role	□	□	□	□	□	■	■	■	□	■	■
	communication	□	□	□	■	□	□	□	□	□	□	□
Hypotheses	division of labor	■	■	□	□	■	■	■	■	■	■	■
	sharing of knowledge	■	□	□	■	□	■	■	□	□	□	□

of the collaborative session. Also, CIR models relying on functional roles (such as writer/reader [56]) would allow to overpass the different search skills required by those tasks while knowledge-based roles [111, 119] are better adapted for exploratory tasks. Moreover, while most of work aims at ranking documents at each query submission, the “groupization” algorithm [84] proposes an interesting point of view since it estimates the collective relevance at the group level. Accordingly, this model could be adapted for performing a synthesis of the collaborative search session by aggregating the users’ feedback over the whole session, and thus providing a unique ranking denoting the collective answer to the information need.

Finally, users also constitute an important variable that should be taken into account in the choice of CIR models. On the one hand, if users are not able to specify skills/roles in which they fit, models considering users as peers [84, 25, 38] constitute the most basic solution. However, user-driven system-mediated approaches should also be considered since they allow users to bring out their skills in real time. On the other hand, combining users’ skills and task requirement might allow to identify the best suited model. For instance, for an exploratory search between a pair of users with different levels of knowledge, we suggest considering the CIR model based on the role of expert and novice

[120]. Another expertise-based model [119] is also interesting for multifaceted information need and groups larger than two users. When skills are more distinguishable with respect to search skills (rather than knowledge), functional roles of Prospector-Miner [89] and Gatherer-Surveyor [111] seem well adapted. Moreover, in the light of the user-driven system-mediated approach [118], a combination of roles evolving with users' behaviors is also possible.

5 Conclusion and Promising Perspectives

Although generally perceived as a solitary process, IR becomes more and more collaborative due to the task complexity and the amount of available information, requiring skill and knowledge complementarity of users. In this survey, we reviewed the literature surrounding collaborative search. After defining the different collaboration forms in the IR field, we presented an overview of the main empirical studies performed to analyze the collaborative search task. We particularly reviewed these work by focusing on their impact on the IR field so as to highlight relevant factors that should be integrated in CIR models. Then, we focused on the particular field of CIR and attempt to give a broad overview of document ranking models. We showed first that there are two main categories of CIR models aiming at solving a shared information need and relying on relevance feedback, users' roles, expertise, and skills. The first category proposes ranking algorithms for collaborative search while the second category of models presents novel approaches focused on the collaboration coordination through group building or role mining algorithms.

Although an intensive implication of researchers in collaborative information seeking regarding the behavioral understanding [23, 64] and the interface design [113, 80], they are still open directions for CIR, whether related to theoretical aspects of IR, or to new paradigms of collaboration. Therefore, we believe that a new generation of CIR models could be proposed, leading to promising perspectives in the CIR field. In what follows, we present some open issues that may give rise to relevant investigations in the field.

New models of collaboration mediation in IR Previous CIR models presented in this survey open several perspectives in terms of collaboration mediation. Two promising approaches have been recently proposed, respectively at the document ranking or the collaborative group building levels. Specifically, mediating document ranking is no more simply viewed as a passive process in which the system guides the collaborative session, but rather as an active process in which collaborators are the owner of the search session achievement [118, 122]. We believe that such user-driven algorithmic mediation would enhance the retrieval effectiveness since it lets the user perform in skills he/she is the best while being supported by an algorithmic mediation that captures his/her implicit search intent. Although some work has been done, they are on their premise. Indeed, the latent role of collaborators formalized in [122] implies that the best-suited IR technique to his/her role is a document ranking. However, as shown in [30], functional rules might be considered to better fit with the whole set of actions performed by collaborators within a search session [23]. The second mediation approach initiates the premise of a collaborative group building [36], which still remains an opened issue for collaborative search. Indeed, the authors assume that candidate users are interested in the same topic. However, this assumption is not really naturalistic since users' interest might be multi-dimensional and that other features might impact on the collaboration likelihood (e.g., users' availability or complementarity). Accordingly, several challenges are remaining: recommending collaborators, collaborative groups, and once is done, connecting these strangers to perform an effective search task.

Theoretical foundations for CIR A significant set of relevant works have been carried out in CIR, built upon individual IR foundations (e.g., probabilistic models [98] or query reformulation [100]). These work outlines that multiple variables (e.g., users' profile, users' role, search actions, relevance judgments, or information need complexity) might impact the search effectiveness while being characterized by a combined effect. Therefore, further work is necessary to build theoretical foundation of CIR and to define heuristics related to users' roles/behavior/search strategies and collaboration hypotheses (division of labor and sharing of knowledge). A preliminary work has been done by [10] with respect to the division of labor but this effort need to be pursued to take into account the overall features or collaborative constraints. Accordingly, there is an important need to investigate axiomatic approaches [63, 24] in CIR. In addition, we could enumerate the following research objectives that should be considered in collaborative-based axiomatic approaches: (a) optimizing the overall performance of the collaborative session (as suggested in dynamic IR [132]) while ensuring the satisfaction of mutual beneficial goals [37], (b) guaranteeing the synergic effect without neglecting the cognitive cost/effectiveness balance [109, 135], (c) leveraging user-user interactions [38] and user-system ones [89, 132]) performed throughout the search session.

Longitudinal CIR models Although empirical studies outline interesting results [32], the temporal factor underlying the interactive setting of collaborative search is still under-explored in the formalization of collaborative ranking models. Most of work consider a search session as a unique entity assuming that the information need is solved in the end. However, the integration of the time factor in CIR models is still challenging since it allows to understand the search process, and accordingly impacts the document relevance. We list in what follows the different ways of dealing with time in CIR. First, the sequence of submitted queries or relevance feedback contributes to the sense-making process [23]. As analyzed in [120], the knowledge of users increases throughout the session allowing users to get a better understanding of the information need and a better perception of the desired outcomes. This longitudinal information surrounding collaborators' actions highlights the search strategy carried out by users. Such statements could be injected in a query expansion process, for instance, that would enhance incrementally the understanding of the search according to the evolving actions within the search session. Second, collaboration might be performed within long intervals (e.g., multiple sessions) requiring the distinction of long/short-term preferences and interests, similarly to individual search [74]. Third, while most of work considers a synchronous search session in which users interact each other simultaneously, collaborative search might occur asynchronously. This particular setting might require appropriate interfaces and models that consider the time intervals of collaborators' actions and switching actions.

Multi-level CIR models In some application domains, the information access is not obvious due to confidentiality and privacy concerns of data. For instance, [39] explain that in the military or diplomatic fields, information is classified and users are only permitted to reach a certain granularity level of the document collection. Also, the collaboration could be perceived in a consumer/supplier environment (such as patient/doctor [48], company/laywer in the e-Discovery domain [5], ...) in which the latter disposes of the access to the whole collection and is asked by the former to give some otherwise inaccessible information. In this case, the main challenge deals with the sensemaking process underlying the heterogeneous accessibility of the collection that prevents the user to have a clear understanding of his/her information need or its context. An interesting issue in CIR could be to support this particular type of collaboration to overcome the multi-level information access in collaborative search while preserving the data privacy and security issue underlying some particular application domains.

Social embedded CIR models Collaborative search is generally considered to happen within a controlled environment, involving a set of users interacting with each other through a search interface [89]. With the fruitful use of social networks in IR [65, 87, 15] and the relatedness of social interactions in both social and collaborative IR [112], we believe that a substantial effort could be done to extend CIR to social environments. Indeed, in addition to benefit from the crowd, a social-oriented CIR model would enable to connect larger groups, eventually on social platforms, and let them work together to leverage from a wider range of complementarity between collaborators. In this context, some approaches have been proposed [85, 121]. However, further work addressing three main tasks remains:

- *Recommending collaborators.* Identifying and recommending experts or users in social-media platforms is a well-known challenge in IR. But, from the collaboration point of view, it leads to promising perspectives since it would allow favoring interactions between the information provider and the recommended users. The most intuitive framework in which collaboration occurs is community question-answering [92]. Also, in social networks (such as Twitter or Facebook), researchers [59] highlight the increasing trend toward friendsourcing for answering a question, which remains in majority unsolved. In this context, two approaches are emerging: (a) recommending users to mention in order to stimulate the information flow between users and favor implicit collaboration [130, 31], and (b) recommending users willing to answer based on explicit collaboration, as done in SearchBuddies [46] or Aardvark [52]. These research directions would constitute a first step to a social-collaborative IR paradigm in which users would leverage from the crowd to collaborate.

- *Building the right group of collaborators.* A further step towards collaboration consists in recommending a cohesive and relevant group of users willing to collaborate to solve a task. For instance, in [4, 112, 124], the authors investigate the value of a collaborative activity in social networks or question-answering platforms to achieve a search task. One interesting challenge in this field is to build the collaborative group according to users' compatibility, availability, and expertise [14, 85, 94, 121]. Regardless of the task goal, another emerging line of work focuses on collaborative task optimization in crowd-sourcing platforms [71, 93, 1]. The main issue here is to optimize the task cost by performing the right task-to-user and user-to-group assignment according to both user and task peculiarities.

- *Mediating document ranking in social-collaborative environment.* Assuming that collaborative groups are built, the next step is to coordinate the search session, and more particularly the document ranking, by leveraging the social indicators. This issue is not tackled yet, but we believe that a CIR model optimizing the collaboration among users and harnessing of social signals would allow to enhance the estimation of the collective relevance of documents with respect to the shared information need. Also, since the community in social networks might be too large to perform a collaborative task, sub-groups of users might be extracted with the goal of solving sub-tasks as suggested in [44]. These sub-tasks might be aggregated to build the final answer, resulting in a significant gain in the task achievement and completeness.

CIR is a young research area that provides promising challenge. Despite the growing interest of researchers in this particular IR domain, it remains under-explored. We hope that this survey surrounding existing work will create synergic effects in the community to produce more effective and better user-adapted ranking algorithms as well as a depth thought in evaluation frameworks adapted for collaborative search.

References

[1] Ittai Abraham, Omar Alonso, Vasilis Kandylas, Rajesh Patel, Steven Shelford, and Aleksandrs Slivkins. How many workers to ask?: Adaptive exploration for collecting high quality labels. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, SIGIR '16, pages 473–482. ACM, 2016.

[2] John Adcock, Jeremy Pickens, Matthew L Cooper, Lisa Anthony, Francine Chen, and Pernilla Qvarfordt. FXPAL Interactive Search Experiments for TRECVID 2007. In *Proceedings of the TRECVid Conference*, TREC '07, 2007.

[3] Sihem Amer-Yahia, Michael Benedikt, and Philip Bohannon. Challenges in Searching Online Communities. *IEEE Data Engineering Bulletin*, 30(2):23–31, 2007.

[4] Ashton Anderson, Daniel P. Huttenlocher, Jon M. Kleinberg, and Jure Leskovec. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *The 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, pages 850–858. ACM, 2012.

[5] Simon Attfield, Ann Blandford, and Stephann Makri. Social and interactional practices for disseminating current awareness information in an organisational setting. *Information Processing & Management (IP&M)*, 46(6):646–655, 2010.

[6] Ismail Badache and Mohand Boughanem. Harnessing social signals to enhance a search. In *Proceedings of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence*, WI '14, pages 303–309, 2014.

[7] Nicholas J Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, (5):133–143, 1980.

[8] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research (JMLR)*, 3:993–1022, 2003.

[9] Thilo Böhm, Claus-Peter Klas, and Matthias Hemmje. Supporting Collaborative Information Seeking and Searching in Distributed Environments. *Proceedings of the "Lernen, Wissen, Adaption" LWA Conference*, pages 16–23, 2013.

[10] Thilo Böhm, Claus-Peter Klas, and Matthias Hemmje. *Collaborative Information Seeking: Best Practices, New Domains and New Thoughts*, chapter A Probability Ranking Principle for Collaborative Search, pages 167–191. Springer International Publishing, 2015.

[11] Chris Buckley, Gerard Salton, and James Allan. The Effect of Adding Relevance Information in a Relevance Feedback Environment. In *Proceedings of the Forum at the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pages 292–300. ACM, 1994.

[12] Robert Capra, Javier Velasco-Martin, and Beth Sams. Levels of "Working Together" in Collaborative Information Seeking and Sharing. In *Proceedings of the International Workshop on Collaborative Information Seeking*, CSCW '10. ACM, 2010.

[13] Pablo Castells, Saul Vargas, and Jun Wang. Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance. In *Proceedings of the International Workshop on Diversity in Document Retrieval (DDR 2011)*, ECIR '11. Springer, 2011.

[14] Shuo Chang and Aditya Pal. Routing questions for collaborative answering in community question answering. In *Advances in Social Networks Analysis and Mining 2013*, ASONAM '13, pages 494–501. ACM, 2013.

[15] Sergiu Viorel Chelaru, Claudia Orellana-Rodriguez, and Ismail Sengör Altingövde. Can social features help learning to rank youtube videos? In *Proceedings of the 13th International Conference on Web Information Systems Engineering*, WISE '12, pages 552–566, 2012.

[16] Jack G Conrad. E-Discovery Revisited: A Broader Perspective for IR Researchers. In *Proceedings of the DESI (EDD) Workshop*, ICAIL '07, pages 321–345, 2007.

[17] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[18] Peter J Denning and Peter Yaholkovsky. Getting to "We". *Communications of the ACM (CACM)*, 51(4):19–24, 2008.

[19] Abdigani Diriye and Gene Golovchinsky. Querium: A session-based collaborative search system. In *Proceedings of the European Conference on Advances in Information Retrieval*, ECIR '12, pages 583–584. Springer, 2012.

[20] Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the Conference on Computer Supported Cooperative Work*, CSCW '92, pages 107–114. ACM, 1992.

[21] Susan T. Dumais. Putting searchers into search. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '14, pages 1–2. ACM, 2014.

[22] David G. Ellis. A Behavioral Approach to Information Retrieval System Design. *Journal of Documentation*, 45(3):171–212, 1989.

[23] Brynn M Evans and Ed H Chi. An elaborated model of social search. *Information Processing & Management (IP&M)*, 46(6):656–678, 2010.

[24] Hui Fang and ChengXiang Zhai. An exploration of axiomatic approaches to information retrieval. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 480–487. ACM, 2005.

[25] Colum Foley and Alan F Smeaton. Synchronous Collaborative Information Retrieval: Techniques and Evaluation. In *Proceedings of the European Conference on Advances in Information Retrieval*, ECIR '09, pages 42–53. Springer, 2009.

[26] Colum Foley and Alan F Smeaton. Division of Labour and Sharing of Knowledge for Synchronous Collaborative Information Retrieval. *Information Processing & Management (IP&M)*, 46(6):762–772, 2010.

[27] Jonathan Foster. Collaborative information seeking and retrieval. *Annual Review of Information Science & Technology (ARIST)*, 40(1):329–356, 2006.

[28] Norbert Fuhr. A probability ranking principle for interactive information retrieval. *Information Retrieval Journal*, 11(3):251–265, 2008.

[29] Gene Golovchinsky, John Adcock, Jeremy Pickens, Pernilla Qvarfordt, and Maribeth Back. Cerchiamo: a collaborative exploratory search tool. *Proceedings of the Demo in Computer Supported Cooperative Work*, 2008.

[30] Gene Golovchinsky, Pernilla Qvarfordt, and Jeremy Pickens. Collaborative Information Seeking. *IEEE Computer*, 42(3):47–51, 2009.

[31] Yeyun Gong, Qi Zhang, Xuyang Sun, and Xuanjing Huang. Who will you "@"? In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 533–542. ACM, 2015.

[32] Roberto González-Ibáñez, Muge Haseki, and Chirag Shah. Time and space in collaborative information seeking: the clash of effectiveness and uniqueness. In *Proceedings of American Society of Information Science and Technology*, ASIST '12, pages 1–10, 2012.

[33] Roberto González-Ibáñez, Muge Haseki, and Chirag Shah. Let's search together, but not too close! An analysis of communication and performance in collaborative information seeking. *Information Processing & Management (IP&M)*, 49(5):1165–1179, 2013.

[34] Roberto González-Ibáñez and Chirag Shah. Coagmento: A system for supporting collaborative information seeking. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–4, 2011.

[35] Roberto González-Ibáñez, Chirag Shah, and Natalia Cordova-Rubio. Smile! Studying expressivity of happiness as a synergic factor in collaborative information seeking. *Proceedings of the ASIS&T Annual Meeting*, 48(1):1–10, 2011.

[36] Roberto González-Ibáñez, Chirag Shah, and Ryen W. White. Capturing collabportunities: A method to evaluate collaboration opportunities in information search using pseudocollaboration. *Journal of the Association for Information Science and Technology*, 66(9):1897–1912, 2015.

[37] Barbara Gray. *Collaborating: finding common ground for multiparty problems*. Jossey Bass Business and Management Series. Jossey-Bass, 1989.

[38] Shuguang Han, Daqing He, Zhen Yue, and Jiepu Jiang. Contextual support for collaborative information retrieval. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, CHIIR '16, pages 33–42. ACM, 2016.

[39] Mark J. Handel and Emily Y. Wang. I can't tell you what i found: Problems in multi-level collaborative information retrieval. In *Proceedings of the International Workshop on Collaborative Information Retrieval*, CIKM '11, pages 1–6. ACM, 2011.

[40] Preben Hansen and Kalervo Järvelin. Collaborative information retrieval in an information-intensive domain. *Information Processing & Management (IP&M)*, 41(5):1101–1119, 2005.

[41] Preben Hansen, Chirag Shah, and Claus-Peter Klas. Collaborative information seeking: Best practices, new domains, new thoughts. pages 1–226. The Computer-Supported Cooperative Work (CSCW), Springer-, 2015.

[42] Morgan Harvey, Ian Ruthven, and Mark Carman. Ranking social bookmarks using topic models. In *Proceedings of the Conference on Information and Knowledge Management*, CIKM '10, pages 1401–1404. ACM, 2010.

[43] Müge Haseki, Chirag Shah, and Roberto González-Ibáñez. Time as a trigger of interaction and collaboration in research teams: A diary study. In *Proceedings of the Conference on Computer Supported Cooperative Work Companion*, CSCW '12, pages 95–98. ACM, 2012.

[44] Ahmed Hassan Awadallah, Ryen W. White, Patrick Pantel, Susan T. Dumais, and Yi-Min Wang. Supporting complex search tasks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pages 829–838. ACM, 2014.

[45] Yin He and Kuansan Wang. Inferring search behaviors using partially observable markov model with duration (pomd). In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 415–424. ACM, 2011.

[46] Brent Hecht, Jaime Teevan, Meredith Ringel Morris, and Daniel J. Liebling. Searchbuddies: Bringing search engines into the conversation. In *Proceedings of the Sixth International Conference on Weblogs and Social Media*, ICWSM '14. The AAAI Press, 2012.

[47] Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. An Algorithmic Framework for Performing Collaborative Filtering. In *Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, pages 230–237. ACM, 1999.

[48] Morten Hertzum. Collaborative information seeking: The combined activity of information seeking and collaborative grounding. *Information Processing & Management (IP&M)*, 44(2):957–962, 2008.

[49] Paul Heymann, Daniel Ramage, and Hector Garcia-Molina. Social tag prediction. In *Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 531–538. ACM, 2008.

[50] Katja Hofmann, Shimon Whiteson, and Maarten de Rijke. Balancing exploration and exploitation in learning to rank online. In *Advances in Information Retrieval - 33rd European Conference on IR Research*, ECIR 2011, pages 251–263. Springer, 2011.

[51] Christoph Hölscher and Gerhard Strube. Web search behavior of Internet experts and newbies. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 33(1-6):337–346, 2000.

[52] Damon Horowitz and Sepandar D Kamvar. The Anatomy of a Large-scale Social Search Engine. In *Proceedings of the International Conference on World Wide Web*, WWW '10, pages 431–440. ACM, 2010.

[53] Jian Hu, Gang Wang, Fred Lochovsky, Jian-tao Sun, and Zheng Chen. Understanding user's query intent with wikipedia. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 471–480. ACM, 2009.

[54] Jette Hyldegärd. Collaborative Information Behaviour: Exploring Kuhlthau's Information Search Process Model in a Group-based Educational Setting. *Information Processing & Management (IP&M)*, 42(1):276–298, 2006.

[55] Jette Hyldegärd. Beyond the Search Process - Exploring Group Members' Information Behavior in Context. *Information Processing & Management (IP&M)*, 45(1):142–158, 2009.

[56] Marika Imazu, Shin-ichi Nakayama, and Hideo Joho. Effect of Explicit Roles on Collaborative Search in Travel Planning Task. In *Proceedings of the Asia Information Retrieval Societies Conference*, AIRS '11, pages 205–214. Springer, 2011.

[57] Peter Ingwersen and Kalervo Järvelin. *The Turn: Integration of Information Seeking and Retrieval in Context (The Information Retrieval Series)*. Springer-Verlag New York, Inc., 2005.

[58] Bernard J Jansen, Danielle L Booth, and Amanda Spink. Determining the Informational, Navigational, and Transactional Intent of Web Queries. *Information Processing & Management (IP&M)*, 44(3):1251–1266, 2008.

[59] Jin-Woo Jeong, Meredith Ringel Morris, Jaime Teevan, and Daniel J. Liebling. A crowd-powered socially embedded search engine. In *Proceedings of the Seventh International Conference on Weblogs and Social Media*, ICWSM 2013. The AAAI Press, 2013.

[60] Xiaoran Jin, Marc Sloan, and Jun Wang. Interactive Exploratory Search for Multi Page Search Results. In *Proceedings of the International Conference on World Wide Web*, WWW '13, pages 655–666. ACM, 2013.

[61] Thorsten Joachims. Optimizing Search Engines Using Clickthrough Data. In *Proceedings of the SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, pages 133–142. ACM, 2002.

[62] Hideo Joho, David Hannah, and JoemonM. Jose. Revisiting IR Techniques for Collaborative Search Strategies. In *Proceedings of the European Conference on Advances in Information Retrieval*, ECIR '09, pages 66–77. Springer, 2009.

[63] Murat Karamuftuoglu. Collaborative information retrieval: Toward a social informatics view of ir interaction. *Journal of the American Society for Information Science*, 49(12):1070–1080, 1998.

[64] Arvind Karunakaran, Madhu C Reddy, and Patricia Ruma Spence. Toward a model of collaborative information behavior in organizations. *Journal of the Association for Information Science and Technology (JASIST)*, 64(12):2437–2451, 2013.

[65] Bastian Karweg, Christian Huetter, and Klemens Böhm. Evolving social search based on bookmarks and status messages from social networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 1825–1834. ACM, 2011.

[66] Ryan Kelly and Stephen J Payne. Division of labour in collaborative information seeking: Current approaches and future directions. In *Proceedings of the International Workshop on Collaborative Information Seeking*, CSCW '13. ACM, 2013.

[67] Giyeong Kim. Relationship Between Index Term Specificity and Relevance Judgment. *Information Processing & Management (IP&M)*, 42(5):1218–1229, 2006.

[68] Nikolaos T Korfiatis, Marios Poulos, and George Bokos. Evaluating authoritative sources using social networks: an insight from Wikipedia. *Online Information Review*, 30(3):252–262, 2006.

[69] R Korhage. Information storage and retrieval. *John Wiley and Sons Inc.*, 1997.

[70] Carol C Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the Association for Information Science and Technology (JASIST)*, 42(5):361–371, 1991.

[71] Hongwei Li, Bo Zhao, and Ariel Fuxman. The wisdom of minority: Discovering and targeting the right group of workers for crowdsourcing. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 165–176. ACM, 2014.

[72] Qing Li and Byeong Man Kim. Clustering approach for hybrid recommender system. In *Proceedings of the International Conference on Web Intelligence*, WI '03, pages 33–38. IEEE Computer Society, 2003.

[73] Xiao Li, Ye-Yi Wang, and Alex Acero. Learning query intent from regularized click graphs. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 339–346. ACM, 2008.

[74] Dong Liu, Xian-Sheng Hua, Meng Wang, and Hong-Jiang Zhang. Image Retagging. In *Proceedings of the International Conference on Multimedia*, MM '10, pages 491–500. ACM, 2010.

[75] Jiyun Luo, Sicong Zhang, Xuchu Dong, and Hui Yang. Designing states, actions, and rewards for using POMDP in session search. In *Advances in Information Retrieval - 37th European Conference on IR Research*, ECIR 2015, pages 526–537. Springer, 2015.

[76] Hao Ma, Michael R Lyu, and Irwin King. Learning to Recommend with Trust and Distrust Relationships. In *Proceedings of the Conference on Recommender Systems*, RecSys '09, pages 189–196. ACM, 2009.

[77] Gary Marchionini. *Information Seeking in Electronic Environments*. Cambridge University Press, 1995.

[78] Meredith R. Morris and Dan Morris. Understanding the Potential for Collaborative Search Technologies in Clinical Settings. In *Proceedings of the International Workshop on Collaborative Information Retrieval*, CIKM '11. ACM, 2011.

[79] Meredith Ringel Morris. A survey of collaborative web search practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1657–1660. ACM, 2008.

[80] Meredith Ringel Morris and Eric Horvitz. SearchTogether: an interface for collaborative web search. In *Proceedings of the Symposium on User Interface Software and Technology*, UIST '07, pages 3–12, 2007.

[81] Meredith Ringel Morris, Jarrod Lombardo, and Daniel Wigdor. WeSearch: supporting collaborative search and sensemaking on a tabletop display. In *Proceedings of the Conference on Computer Supported Cooperative Work*, CSCW '10, pages 401–410. ACM, 2010.

[82] Meredith Ringel Morris, Andreas Paepcke, and Terry Winograd. TeamSearch: Comparing Techniques for Co-Present Collaborative Search of Digital Media. In *Proceedings of the International Workshop on Horizontal Interactive Human-Computer Systems*, Tabletop '06, pages 97–104. IEEE Computer Society, 2006.

[83] Meredith Ringel Morris and Jaime Teevan. *Collaborative Web Search: Who, What, Where, When, and Why*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009.

[84] Meredith Ringel Morris, Jaime Teevan, and Steve Bush. Collaborative Web Search with Personalization: Groupization, Smart Splitting, and Group Hit-highlighting. In *Proceedings of the Conference on Computer Supported Cooperative Work*, CSCW '08, pages 481–484. ACM, 2008.

[85] Besmira Nushi, Omar Alonso, Martin Hentschel, and Vasileios Kandylas. Crowdstar: A social task routing framework for online communities. In -*Proceedings of the 15th International Conference on the Engineering the Web in the Big Data Era*, ICWE '15, pages 219–230. Springer, 2015.

[86] Paul Over. The TREC interactive track: an annotated bibliography. *Information Processing & Management (IP&M)*, 37(3):369–381, 2001.

[87] Aditya Pal and Scott Counts. Identifying topical authorities in microblogs. In *Proceedings of the Conference on Web Search and Data Mining*, WSDM '11, pages 45–54. ACM, 2011.

[88] S A Paul and M C Reddy. A Framework for Sensemaking in Collaborative Information Seeking. In *Proceedings of the International Workshop on Collaborative Information Seeking*, CSCW '10. ACM, 2010.

[89] Jeremy Pickens, Gene Golovchinsky, Chirag Shah, Pernilla Qvarfordt, and Maribeth Back. Algorithmic Mediation for Collaborative Exploratory Search. In *Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 315–322. ACM, 2008.

[90] Steven Poltrock, Jonathan Grudin, Susan Dumais, Raya Fidel, Harry Bruce, and Annelise Mark Pejtersen. Information Seeking and Sharing in Design Teams. In *Proceedings of the International SIGGROUP Conference on Supporting Group Work*, GROUP '03, pages 239–247. ACM, 2003.

[91] Jay M Ponte and W Bruce Croft. A language modeling approach to information retrieval. In *Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 275–281. ACM, 1998.

[92] Marie L. Radford, Lynn Silipigni Connaway, Stephanie Mikitish, Mark Alpert, Chirag Shah, and Nicole A. Cooke. Shared values, new vision: Collaboration and communities of practice in virtual reference and sqa. *Journal of the Association for Information Science and Technology (JASIST)*, 2016.

[93] Habibur Rahman, Senjuti Basu Roy, Saravanan Thirumuruganathan, Sihem Amer-Yahia, and Gautam Das. "the whole is greater than the sum of its parts": Optimization in collaborative crowdsourcing. *CoRR*, abs/1502.05106, 2015.

[94] Suhas Ranganath, Jiliang Tang, Xia Hu, Hari Sundaram, and Huan Liu. Leveraging social foci for information seeking in social media. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 261–267, 2015.

[95] Madhu C Reddy and Patricia Ruma Spence. Collaborative Information Seeking: A Field Study of a Multidisciplinary Patient Care Team. *Information Processing & Management (IP&M)*, 44(1):242–255, 2008.

[96] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of the Conference on Computer Supported Cooperative Work*, CSCW '94, pages 175–186. ACM, 1994.

[97] Stephen E. Robertson. Threshold setting and performance optimization in adaptive filtering. *Information Retrieval Journal*, 5(2-3):239–256, 2002.

[98] Stephen E Robertson, S Walker, S Jones, M Hancock-Beaulieu, and M Gatford. Okapi at TREC-3. In *Proceedings of the Text retrieval conference-3 (TREC-3)*, TREC '95, pages 109–126, 1995.

[99] Stephen E Robertson and Stephen Walker. Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. In *Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pages 232–241. ACM, 1994.

[100] J J Rocchio, editor. *Relevance Feedback in Information Retrieval*. Prentice Hall, 1971.

[101] Jesus A. Rodriguez Perez, Stewart Whiting, and Joemon M. Jose. CoFox: A visual collaborative browser. In *Proceedings of the International Workshop on Collaborative Information Retrieval*, CIKM '11. ACM, 2011.

[102] G Salton. A comparison between manual and automatic indexing method. *Journal of American Documentation*, 2(1):61–71, 1971.

[103] G Salton and M J McGill. *Introduction to modern information retrieval*. McGraw-Hill computer science series. McGraw-Hill, 1983.

[104] Chirag Shah. Toward collaborative information seeking (cis). In *Proceedings of the Joint Conference on Digital Libraries*, JCDL '08. ACM, 2008.

[105] Chirag Shah. Collaborative Information Seeking - The Art and Science of Making the Whole Greater than the Sum of All. pages I–XXI, 1–185, 2012.

[106] Chirag Shah. Collaborative information seeking. *Journal of the Association for Information Science and Technology (JASIST)*, 65(2):215–236, 2014.

[107] Chirag Shah. Evaluating collaborative information seeking - synthesis, suggestions, and structure. *Journal of Information Science (JIS)*, 40(4):460–475, 2014.

[108] Chirag Shah and Roberto González-Ibáñez. Exploring Information Seeking Processes in Collaborative Search Tasks. In *Proceedings of the ASIS&T Annual Meeting*, ASIS&T '10, pages 60:1–60:10, 2010.

[109] Chirag Shah and Roberto González-Ibáñez. Evaluating the Synergic Effect of Collaboration in Information Seeking. In *Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 913–922. ACM, 2011.

[110] Chirag Shah and Gary Marchionini. Awareness in collaborative information seeking. *Journal of the Association for Information Science and Technology (JASIST)*, 61(10):1970–1986, 2010.

[111] Chirag Shah, Jeremy Pickens, and Gene Golovchinsky. Role-based results redistribution for collaborative information retrieval. *Information Processing & Management (IP&M)*, 46(6):773–781, 2010.

[112] Chirag Shah, Marie L. Radford, and Lynn Silipigni Connaway. Collaboration and synergy in hybrid q&a: Participatory design method and results. *Library & Information Science Research*, 37(2):92 – 99, 2015.

[113] Alan F Smeaton, Colum Foley, Cathal Gurrin, Hyowon Lee, and Sinead McGivney. Collaborative Searching for Video Using the Fischlar System and a DiamondTouch Table. In *Proceedings of the International Workshop on Horizontal Interactive Human-Computer Systems*, Tabletop '06, pages 151–159. IEEE Computer Society, 2006.

[114] Ian Soboroff and Donna Harman. Novelty Detection: The TREC Experience. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 105–112. Association for Computational Linguistics, 2005.

[115] Fei Song and W Bruce Croft. A General Language Model for Information Retrieval. In *Proceedings of the Conference on Information and Knowledge Management*, CIKM '99, pages 316–321. ACM, 1999.

[116] Diane H Sonnenwald, Kelly L Maglaughlin, and Mary C Whitton. Designing to support situation awareness across distances: an example from a scientific collaboratory. *Information Processing & Management (IP&M)*, 40(6):989–1011, 2004.

[117] Diane H. Sonnenwald and Linda G. Pierce. Information behavior in dynamic group work contexts: interwoven situational awareness, dense social networks and contested collaboration in command and control. *Information Processing & Management (IP&M)*, 36(3):461–479, 2000.

[118] Laure Soulier, Chirag Shah, and Lynda Tamine. User-driven System-mediated Collaborative Information Retrieval. In *Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '14, pages 485–494. ACM, 2014.

[119] Laure Soulier, Lynda Tamine, and Wahiba Bahsoun. A Collaborative Document Ranking Model for a Multi-faceted Search. In *Proceedings of the Asia Information Retrieval Societies Conference*, AIRS '13, pages 109–120. Springer, 2013.

[120] Laure Soulier, Lynda Tamine, and Wahiba Bahsoun. On domain expertise-based roles in collaborative information retrieval. *Information Processing & Management (IP&M)*, 50(5):752–774, 2014.

[121] Laure Soulier, Lynda Tamine, and Gia-Hung Nguen. Answering twitter questions: a model for recommending answerers through social collaboration. In *Proceedings of the 25rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '16, page to appear, 2016.

[122] Laure Soulier, Lynda Tamine, and Chirag Shah. Minerank: Leveraging users latent roles for unsupervised collaborative information retrieval. *Information Processing & Management*, 5(6):1122–1141, 2016.

[123] Lynda Tamine and Laure Soulier. Understanding the impact of the role factor in collaborative information retrieval. In *Proceedings of the ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 43–52. ACM, 2015.

[124] Lynda Tamine, Laure Soulier, Lamjed Ben Jabeur, Frederic Amblard, Chihab Hanachi, Gilles Hubert, and Camille Roth. Social media-based collaborative information access: Analysis of online crisis-related twitter conversations. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, HT '16, pages 159–168. ACM, 2016.

[125] Yihan Tao and Anastasios Tombros. An Exploratory Study of Sensemaking in Collaborative Information Seeking. In *Proceedings of the European Conference on Advances in Information Retrieval*, ECIR '13, pages 26–37. Springer, 2013.

[126] Yihan Tao and Anastasios Tombros. How collaborators make sense of tasks together: A comparative analysis of collaborative sensemaking behavior in collaborative information-seeking tasks. *Journal of the Association for Information Science and Technology*, 2016.

[127] Jaime Teevan, Susan T Dumais, and Eric Horvitz. Personalizing Search via Automated Analysis of Interests and Activities. In *Proceedings of the Annual International SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '05, pages 449–456. ACM, 2005.

[128] Michael B Twidale, David M Nichols, and Chris D Paice. Browsing is a Collaborative Process. *Information Processing & Management (IP&M)*, 33(6):761–783, 1997.

[129] Hedy S Wald, Catherine E Dube, and David C Anthony. Untangling the WebThe impact of Internet use on health care and the physicianpatient relationship. *Patient Education and Counseling*, 68(3):218–224, 2007.

[130] Jian Wang, Cheng Zeng, Chuan He, Liang Hong, Liang Zhou, Raymond K Wong, and Jilei Tian. Context-aware Role Mining for Mobile Service Recommendation. In *Proceedings of the Annual Symposium on Applied Computing*, SAC '12, pages 173–178. ACM, 2012.

[131] Ryen W White and Susan T Dumais. Characterizing and Predicting Search Engine Switching Behavior. In *Proceedings of the Conference on Information and Knowledge Management*, CIKM '09, pages 87–96. ACM, 2009.

[132] Hui Yang, Marc Sloan, and Jun Wang. Dynamic information retrieval modeling. In *The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '14, page 1290. ACM, 2014.

[133] Ke-Thia Yao, Robert Neches, In-Young Ko, R Eleish, and S Abhinkar. Synchronous and asynchronous collaborative information space analysis tools. In *Proceedings of the International Workshops on Parallel Processing*, ICPPW '99, pages 74–79. IEEE Computer Society, 1999.

[134] Yisong Yue and Thorsten Joachims. Interactively optimizing information retrieval systems as a dueling bandits problem. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 1201–1208. ACM, 2009.

[135] Zhen Yue. Investigating search processes in collaborative exploratory web search, June 2014.

[136] Zhen Yue, Shuguang Han, and Daqing He. Modeling search processes using hidden states in collaborative exploratory web search. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, pages 820–830. ACM, 2014.

[137] Zhen Yue, Shuguang Han, Daqing He, and Jiepu Jiang. Influences on Query Reformulation in Collaborative Web Search. *IEEE Computer*, 47(3):46–53, March 2014.

[138] Zhen Yue and Daqing He. Exploring collaborative information behavior in context: a case study of e-discovery. In *Proceedings of the International Workshop on Collaborative Information Seeking*, CSCW '10. ACM, 2010.

[139] Zeyuan Allen Zhu, Weizhu Chen, Tom Minka, Chenguang Zhu, and Zheng Chen. A novel click model and its applications to online advertising. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 321–330. ACM, 2010.