Understanding the Impact of the Role Factor in Collaborative Information Retrieval

Lynda Tamine
Université de Toulouse UPS IRIT
118 Route de Narbonne
F-31062 Toulouse, France
tamine@irit.fr

Laure Soulier
Université de Toulouse UPS IRIT
118 Route de Narbonne
F-31062 Toulouse, France
soulier@irit.fr

ABSTRACT

Collaborative information retrieval systems often rely on division of labor policies. Such policies allow work to be divided among collaborators with the aim of preventing redundancy and optimizing the synergic effects of collaboration. Most of the underlying methods achieve these goals by the means of explicit vs. implicit role-based mediation. In this paper, we investigate whether and how different factors, such as users' behavior, search strategies, and effectiveness, are related to role assignment within a collaborative exploratory search. Our main findings suggest that: (1) spontaneous and cohesive implicit roles might emerge during the collaborative search session implying users with no prior roles, and that these implicit roles favor the search precision, (2) role drift might occur alongside the search session performed by users with prior-assigned roles.

Categories and Subject Descriptors

H.3.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Information Search and Retrieval - *Clustering, Retrieval models, Search process*

General Terms

Collaborative Information Retrieval

Keywords

Collaborative information retrieval, user behavior analysis, user study

1. INTRODUCTION

In [11], the authors define collaborative information seeking (CIS) as "*the study of the systems and practices that enable individuals to collaborate during the seeking, searching, and retrieval of information*". Broadly speaking, collaboration allows people to create and share collective knowledge within a work team to identify and resolve a shared complex problem [24].

CIKM'15, October 19–23, 2015, Melbourne, Australia.

ACM 978-1-4503-3794-6/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/XXX.XXXXXXX .

Collaboration outlines three main paradigms (division of labor, awareness and sharing of knowledge) that avoid redundancy between collaborators and favor the synergic effects of collaboration [10, 25]. Division of labor has been highlighted as the most important paradigm for allowing a work team to cover different facets of a search while minimizing the redundancy between the actions of the involved members [25, 26]. To address these concerns, CIS tools are based on three main approaches to division of labor: user-driven approaches [20, 7], system-mediated approaches [10, 23, 29], and user-driven system-mediated approaches [31]. Focusing on the commonalities that exist between them, we can observe that all of them rely on the user's role as the means of work coordination toward the achievement of the overall task goal. The primary difference among these approaches is the underlying users' role specifications and uses during the collaborative task. In user-driven approaches to the division of labor, users' roles can be either (1) fixed (e.g., searcher/writer) and then assigned to users before the search or (2) freely negotiated and created by the collaborators during the search according to their preferences and search and knowledge skills. In both cases, communicative and interactive interfaces are used to ensure search awareness and coordination. In contrast, in the case of system-mediated approaches, users' roles are automatically leveraged via algorithmic mediation to optimize the search session effectiveness. However, this mediation adapts the search toward the same fixed roles regardless of the user's involvement in the search activity. To maintain the advantages of the two former approaches and address their limitations, a hybrid approach, which is referred to as the user-driven system-mediated approach [31], has been proposed. In this approach, based on users' activities, the roles (from within a set of fixed roles) in which they fit best are dynamically learned, and then the search is algorithmically mediated accordingly.

Based on these approaches and the related studies, we conjecture that regardless of the underlying mediation policy, proper specification of users' roles, whether performed in advance or dynamically during the search, is critical because it might alter the synergic effects of the collaborative task. How the synergic effects are altered depends on the gap between (1) how the users are expected to behave under (system or human) role assignment and (2) how they really behave during the collaborative search activity. This potential alteration is problematic for CIS designers because it leads to the creation of processes and technologies that do not fit with natural, collaborative information behavior. Relatively few fundamental studies have addressed this issue

[19, 25], and to the best of our knowledge, this work is the first empirical comparison of an assigned, role-based collaborative information search and a non-assigned-role-based one in terms of users' search behavior and effectiveness. Through close examination of users' interaction processes and search strategies, we thoroughly study the importance of role factor within a collaborative search process, with the goal of achieving useful and effective collaboration. The results can inform the design of future CIS/collaborative information retrieval (CIR) systems.

The remainder of this paper is organized as follows. Section 2 presents the background for our study, including previous work regarding CIS and the research questions addressed in the paper. Section 3 describes the study design. Section 4 details the study results and presents related discussions. Section 5 highlights the study implications, and Section 6 concludes the paper.

2. BACKGROUND

2.1 Related Work

A collaborative search task can be defined as a complex process that involves several users who interact with each other with the goal of addressing a shared information need. Previous studies [28, 29] have demonstrated the benefits of such a setting for gathering the complementary skills of individual users to solve complex tasks, such as fact-finding tasks (e.g., travel planning) [19] or exploratory search tasks [29]. Within the latter, insufficient knowledge or skills of individual users is the main reason that motivates them to collaborate, which allows collaborators to attain a synergic effect in which *"the whole (is) greater than the sum of all"* [28]. CIS results in collaborative information behavior processes, such as information sharing, evaluation, synthesis, sense-making and utilization of the retrieved information [14]. Whereas search behavior model in an individual search context is well-studied [32, 17], few research works have focused on the collaborative setting [15, 18, 8]. Evans and Chi [8] examined the collaboration process in social searches and defined three main phases in which collaborators interact with each other to optimize the task outcome. Before a search, collaborators' interactions mainly focus on refining the information need and structuring the task guidelines. During the search, users interact within a sense-making process via querying or information foraging, with the aim of increasing the search effectiveness. After the search, they collaboratively interact to assess the collective relevance of the results. Other works investigated a deeper level in the search behavior understanding by focusing either on the collaboration search pratices [19] or the impact of communication channels on collaborative search [13]. Using qualitative and quantitative analysis, the authors highlighted the important need to propose adapted collaborative systems (e.g., relying on audio tools) which would avoid redundant search actions or limit the cognitive effort of collaborators.

To optimize search behaviors among collaborators, three main principles guide the collaboration process: (1) avoiding redundancy between users' actions (division of labor), either at the document level [9] or role level [23, 29], (2) favoring information flow among users (sharing of knowledge), either implicitly by search inference [10] or explicitly by collaborative-based interfaces [22, 26], and (3) informing users of other collaborators' actions (awareness) [26]. In this paper, we focus specifically on the division of labor principle, which can be traced through three levels of mediation that result in specific user behaviors [25]. First, coordination between collaborators can be naturally induced through explicit discussions or exchanges. For this reason, interfaces that allow the users to be freed from specific actions by providing communication and organization-based support have been proposed [28]. The authors examined the communication process in collaborative search settings and found that even if it does not impact the search effectiveness, communication helps collaborators to address many aspects of the information needs. Second, designing adaptive interfaces for collaboration is a key challenge, as noted by Joho et al. [16]. Some of the proposed interfaces [20, 22] facilitate division of labor by allowing users to split and manually distribute the search results among collaborators. Third, algorithmic mediation [9, 23, 29] focuses on the interface background; this process consists of automatically ranking documents in response to queries formulated by the users. Collaborators are no longer the owners of the search session progress because their search results are the outcomes of a ranking algorithm that takes their context into consideration. Recently, a hybrid mediation-based approach was proposed [31]. The core of the underlying method is to assign roles to users based on the differences in their search actions, with the aim of optimizing information retrieval performance. For all of these approaches, division of labor can be exercised based on collaborators' roles, thus enabling structuring of the search process [25] among collaborators and minimizing redundant actions. Golovchinsky et al. [12] define a role taxonomy that involves several pairs of roles, such as domain A expert/domain B expert, search expert/search novice and prospector/miner (PM). Within the last pair of roles, the *prospector* favors diversity in the search results by opening new exploration fields in the information space, whereas the *miner* ensures the quality and richness of the explored documents. Practically, the role of the prospector is to identify several facets of the information need by skimming documents, whereas the miner is responsible for in-depth reading of the documents to select the most relevant ones. In [29], the authors define the pair of gatherer/surveyor (GS). Whereas the *gatherer* aims at quickly selecting relevant documents, the *surveyor* has the objective of covering a wide range of results to better understand the nature of the information needed to explore the potential exploratory fields and to determine why queries are suboptimal. These roles are complementary because the gatherer can search relevant information alone, whereas the surveyor needs the collective intelligence to ensure topical diversity, which enables a better understanding of the information need landscape.

2.2 Research Questions

The two main conclusions that one can draw from the literature review are the following: (1) the synergic effects of collaboration may be achieved through coordination approaches regardless of whether roles are explicitly assigned to collaborators, (2) a collaborative search task results in a large range of specific user behavioral facets (e.g., communicating with peers, formulating queries, skimming or reading in-depth documents, selecting/splitting results, and sense-making). To design efficient support processes for CIS and CIR, we need to (1) fully understand how users' search behavior changes in a collaborative search session

with respect to the division of labor policy and (2) measure how these changes impact the synergic effects of collaboration. However, there is a lack of studies that address role specification as a critical factor that can either facilitate or inhibit the collaboration effectiveness. Unlike previous work that focused on the synergic effects of collaboration compared with individual searches [28], we want to extend our understanding of the differences in users' behaviors in role-oriented and non-role-oriented CIR scenarios. In particular, we wish to focus on an exploratory web search task for which we believe that role-based coordination is challenging. Indeed, several studies have previously demonstrated that role-based coordination particularly benefits from synergic collaboration among searchers to address different aspects of the search results [6, 23, 26]. We address this issue by performing a user study that allows comparing and contrasting both search scenarios. In the study, the role-based scenario relies on well-known, state-of-the-art roles, namely the PM [23] (defined in the role taxonomy of [12]) and GS [29] pairs. Our study provides insight into the real manifestations of division of labor policies in each collaborative search scenario and enables us to emphasize the strengths of each, which could be combined to achieve synergic collaboration. We address the following research questions:

- RQ1: How do user search behaviors that result from collaborating with assigned roles differ from those of users that collaborate without pre-defined roles?
- RQ2: To what extent, if any, does the coordination of a collaborative group with assigned roles differ from the coordination of a collaborative group without roles?
- RQ3: Do users with assigned roles respect the role guidelines? If not, what are the main underlying reasons that they do not?
- RQ4: What is the effect of role assignment on the search results effectiveness?

To answer these research questions, we postulate the following hypotheses:

- H1: The behaviors of collaborators are complementary with respect to division of labor policies [30].
- H2: Differences in the behaviors of members that belong to the same working group represent complementarity signals [31].

3. METHODOLOGY

Our user study involved 75 pairs of users that performed a collaborative exploratory search task. With respect to the study objective, users were assigned to different collaborative search scenarios: (1) search with no-prior roles and (2) search with a prior fixed role, either the prospector and miner [12, 23] or the gatherer and surveyor [29] pair of roles. The log data from the user study are available under the conditions specified[1]. Below, we describe the experimental protocol, including participants, system, search tasks, and processed data.

3.1 Participants

In this study, we recruited participants by dyads, under the constraint that they already knew each other and had already worked on collaborative projects together, to facilitate collaboration during the experiments. The collaborators, who were mainly undergrad and Ph.D. students, were between 18 and 30 years old. All of them had experience in browsing the web and using search engines. For their involvement in the user study, participants received material compensation (worth $20), and the three most effective groups received an additional gift (worth $50). Participants were randomly assigned to one of the search settings that we considered: 25 dyads for the scenario in which there was no prior explicit role assignment, which we refer to as W/oRole, and 50 dyads for the scenario in which there was fixed role assignment prior to the search, which we refer to as W/Role. The dyads were split into PM and GS pairs.

Table 1: **Collaborative task example**

The mayor of your countryside village must choose between building a huge industrial complex or developing a nature reserve for animal conservation. As forest preservationists, you must raise awareness about the possibility of wildlife extinction surrounding such an industrial complex. ***Yet, before warning all citizens, including the mayor, you must do extensive research and collect all the facts about the matter. Your objective is to create a claim report together, outlining all the possible outcomes for wildlife should the industrial complex be built.*** *Your focus is on wildlife extinction. You must investigate the animal species involved, the efforts done by other countries and the association worldwide to protect them and the reasons we, as humans, must protect our environment in order to survive.* ***You must identify all relevant documents, facts, and pieces of information by using bookmarks, annotations, or saving snippets. If one document discusses several pieces of useful information, you must save each piece separately using snippets. Please assume that this research task is preliminary to your writing, enabling you to provide all relevant information to support your claims in your report.***

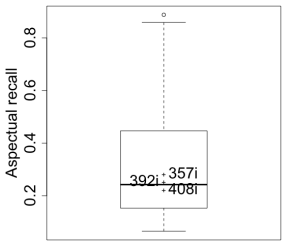


Figure 1: **Aspectual recall analysis over TREC Interactive topics**

3.2 System

To support collaborative search actions, we used an extension of the lab-study open-source Coagmento collaborative search system[2] [27], which we downloaded in December 2013. This version consists of a Firefox plugin that enables tracking of collaborators' actions throughout a web search session and consists of a toolbar that allows participants to rate, annotate and snip web pages. We extended the open-source version by replacing the bookmark option with a page rating tool and including a sidebar with a chat system and an action viewer that enable participants to see the results of the actions (namely, bookmarked, annotated and snipped pages) of their collaborators. To capture all of the participants' actions, we asked them to exchange information only through the chat system.

[1]contact authors: tamine,soulier@irit.fr

[2]https://github.com/kevinAlbs/CoagmentoCollaboratory

Table 2: **Pre-task and post-task questionnaires**

	Questions	Possible responses
Pre-Task Questions	1. How old are you?	a) 15-20 years old b) 21-25 years old c) 26-30 years old d)+30 years old
	2. What is your gender?	a) Female b) Male
	3. How often do you use search engines?	a) More 10 times/day b) 2-3 times/day c) Less than 1 time/day d) Less than 1 time/week
	4. How much experienced in web browsing are you?	From 1 (very experienced) to 4 (not experienced at all)
	5. What kind of tools do you use for web search?	a) Search engines b) Social networks c) Direct browsing d) Other
	6. Which web searcher do you use the most?	a) Bing b) Google c) Yahoo d) Other
	7. Are you used to collaboratively work on project or web search?	a) Daily b) Weekly c) Monthly d) Yearly e) Never
	8. What do you think about collaborative work?	a) Enriching b) Useful c) Useless d) Conflictual
Post-Task Questions	1. Have you already participated in such user study? If yes, please describe it.	Free-answer
	2. What do you think about this collaborative manner of seeking information?	Free-answer
	3. What was the level of difficulty of the task?	a) Easy (Not difficult) b) Moderately difficult c) Difficult
	4. What was task difficulty related to?	Free-answer
	5. Could you say that the collaborative system supports your search?	a) Yes b) Not totally c) Not at all
	6. How could we improve this system?	Free-answer

3.3 Tasks

All of the groups within each search setting aimed to solve the same synchronous, collocated, exploratory information search task for which information needs were extracted from the Text Retrieval Conference (TREC) 6-7-8 Interactive track. Because the TREC Interactive topics are designed for exploratory search tasks and lead TREC participants to cover all the aspects (i.e., facets) of the information needs [21], they have already been used in previous CIR research [9, 23]. According to the study objective, we modified the formulation of the TREC topics into collaborative search tasks. An example task description is presented in Table 1. The collaborative facets of the task are highlighted in bold. To limit the effects of knowledge on the search behaviors and effectiveness [32], we allowed the groups that belonged to different search settings to select the topics to be explored. However, to control the effect of the search topic across the different search settings, we first identified homogeneous topics. To achieve this goal, we used the TREC Interactive search logs to select topics that had similar aspectual recall scores [21] that were near the median, as illustrated in Figure 1 (topics 329i, 347i and 408i). Finally, we assigned the topics to the participants with their agreement.

Role guidelines were also supplied to guide the division of labor policy within the search task. We used the PM and GS role pairs, which were designed in previous work for solving exploratory search tasks. The pairs of roles are complementary in the sense that each pair corresponds to one of the two dimensions of the division of labor (search activity [12, 23] and content exploration [29]), and they enable collaboration between dyads or within a small group of collaborators. To allow for fair comparisons between the W/oRole and each of the W/Role pairs, namely PM and GS, we considered collaborative settings based on dyads. Below, we describe the peculiarities of each search setting:

- *W/oRole* (topic 408i): This setting did not include any role guidelines and thus allowed participants to freely organize collaboration throughout the search session.
- *PM* (topic 392i): We used the role guidelines described in [23]: (1) the prospector "opens new fields for exploration into a data collection" and (2) the miner "ensures that rich veins of information are explored".
- *GS* (topic 347i): The role guidelines were extracted from [29]: (1) the gatherer "scan[s] results of the joint search activity of team members to discover the most immediately relevant information", (2) the surveyor "browse[s] a wider diversity of information to obtain a better understanding of the nature of the collection being searched, to understand where the current queries might be failing, and to identify potential avenues of exploration".

The role guidelines were first presented to participants in a written format and then orally explained during a separate face-to-face meeting to highlight the main differences: (1) both the gatherer and miner aim to identify the most relevant documents for the core of the information need, but the former briefly analyses documents, whereas the latter looks deeper into the document content. (2) The surveyor and prospector both focus on diversity, but the role of the surveyor is more than a simple "new content tracker"; rather, the surveyor aims to explore a large area of the information need landscape.

3.4 Data

The user study consisted of a 4-step task:

1. *A sign-up and pre-task questionnaire*, which is described in Table 2, in which the pairs of participants registered their team and provided some demographic information, such as their age, gender and browsing behavior. To assign the roles for the most suited member in the team, the questionnaire included 2 additional questions extracted from [23, 29]. For the GS role allocation, we asked "Who is the most likely to fit with these behaviors: (1) quickly scan pages displayed by a web searcher? and (2) browse documents not necessarily related to the main topic?" For the PM, the question was, "Who is the most likely to fit with these behaviors: (1) identifying the most highly relevant documents? and (2) browse documents not necessarily related to the main topic?"

2. *A training step* during which participants completed a system tutorial that introduced the collaborative system and let them test its functionalities for a few minutes.

3. *The 30-minute search task.* The automatically collected data included participants' search log data, which were issued from physical actions such as visiting pages or expressing feedback (annotations, snippets, and bookmarks)

Table 3: **Behavior-based features**

Feature	Description
npq	Average number of visited pages by query
dt	Average time spent between two visited pages
nf	Average number of relevance feedback information (snippets, annotations & bookmarks)
qn	Average number of submitted queries
ql	Average number of query tokens
qo	Average ratio of shared tokens among successive queries
nbm	Average number of exchanged messages within the search groups

Table 4: **Chat message categories [13]**

	Description	Examples issued from our studies
(TC)	Task coordination: All types of statements which involve decision making about how the task should be performed	Let's divide the work. Look at the definition, I look on the issues.
(TN)	Task content: All types of statements which include information assessment, layout, structure, and revision of report	I found something interesting *www.nhc.noaa.gov/*
(TS)	Task social: All types of statements that concern group functioning, effort, or latitude, as well as opinions regarding to information obtained or information sources	What do you think about this definition? We are doing well
(NT)	Non-task related: All statements that are not related to the assignment or regarding technical issues of system being used	I can't see my snippets How to remove snippets?
(NC)	Non-codable: All types of statements that do not belong to any category specificities	Ok Thanks

according to the features presented in Table 3. These features, which were defined based on the literature [1], model users' behaviors and search experience. All of the behavioral features analyzed in our study were tested for homogeneity in the different collaborative settings by performing outlier detection based on the Chebyshev inequality [2]. To exploit the chat messages exchanged between the collaborative group members, we manually annotated the entire pool of messages following the categorization proposed in [13] and presented in Table 4. Each message was qualitatively examined for content by two graduate students and then categorized. For categorization, strong evidence regarding the participants' feelings had to be expressed in the message, according to the category description. The agreement level between the two assessors was estimated using Cohen's kappa coefficient. We obtained a value equal to 0.71, which indicated substantial agreement. To check the coding consistency, the students met to make consensual decisions about the final categorization.

4. A *post-task questionnaire* (Table 2) that collected participants' feedback after the search task. In this study, we focused on questions 3 and 4 of the post-task questionnaire, which concerned the level and type of difficulty, respectively, that participants encountered during the search. We note that question 4 was a free-answer question for which a coding scheme was constructed with the six main categories presented in Table 5. Because one answer could be labeled in multiple categories, the agreement level between the two assessors was estimated using the Jaccard index; the value was 0.77, which indicates substantial agreement. The assessors discussed the cases of disagreement to assign a consensual category.

Table 5: **Categorization of participants' responses about the nature of task difficulty**

Categ.	Description	Examples
Page	All statements about web page relevance	Identify relevant information within noisy pages
Comm.	All statements including communication concerns	Exchange through the chat system instead of by voice
Org.	All statements focusing on task organisation	Split search tasks
Topic	All statements concerning the task topic	Focus on the task and do not disgress
Technic.	All statements related to system technical issues	The chat system window is too small
Role	All statements connected to role assignment	It is difficult to keep my role

4. RESULTS AND DISCUSSION

In this section, we present and discuss the results. Then, we enumerate the primary findings that emerged from them. We used both quantitative and qualitative methods for the analysis, as detailed below.

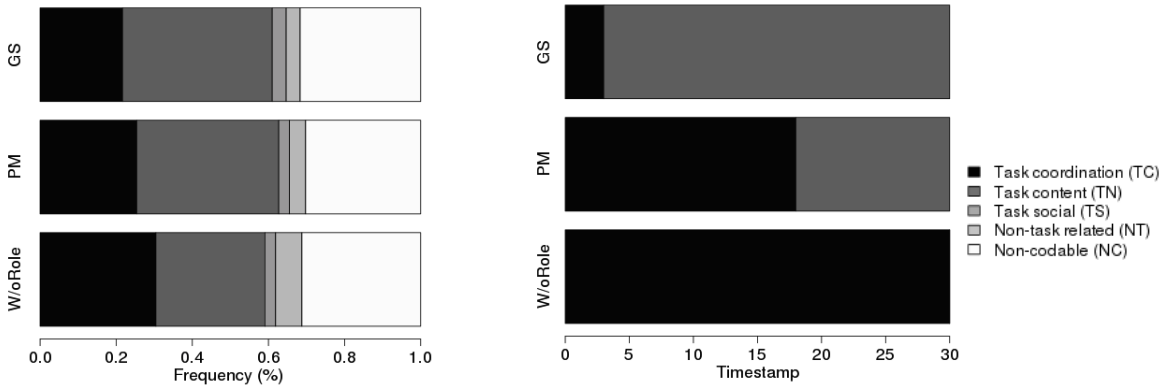
4.1 Analyzing users' behavioral differences

Our aim here was to examine the effects of the role factor on the users' behavior. The search groups' behavioral activities were inferred from physical actions such as visiting pages or expressing feedback (annotations, snippets, and bookmarks). For this purpose, we collected search log data and modeled the users' behavior according to the features presented in Table 3. Following [31], we employed a temporal, feature-based representation of each user's u search session using a matrix $S_u^{(t)} \in \mathcal{R}^{t \times n}$, where n is the number of features. Each element $S_u^{(t)}(t_l, f_k)$ corresponds to the cumulative value of feature f_k for user u aggregated from the beginning of the search session until timestamp t_l. Consequently, we performed Kolmogorov-Smirnov tests to assess the significance of the differences between the *W/oRole* scenario and each of the *W/Role* search scenarios. Table 6 provides a summary of groups' search behavior feature values and the associated standard statistical indicators at the end of the search session ($t_l = 30$). To check wether the *W/Role* intra-group users behaved differently as they were assigned different explicit roles, we computed the differences in their intrinsic search behavioral features ($IGDiff^p$ where p represents the significance value of users' difference estimated through a Student test). To further determine whether the observed differences between the search scenarios varied because of the general underlying division of labor policy, either *W/oRole* or *W/Role*, regardless of the role specifications, we lumped the PM and GS pairs of roles together and performed a one-way analysis of variance (ANOVA). Specifically, because the settings *W/oRole* and *W/Role* had unequal sample sizes, which can lead to inhomogeneous variances, we used a one-way Welch ANOVA [5]. The significance of each role factor as determined by the obtained p-value is reported in the last row of Table 6.

We confirm from Table 6 that intra-group users performing a *W/Role* search had significantly different behavioral facets ($IGDiff^p$). We can also see from Table 6 that the search behaviors collected from *W/oRole* collaborative

Table 6: **Search behavior analysis. Mean(s.d.): Mean value of the feature over the groups (standard deviation). $IGDiff^p$: Mean value and significance value of the feature difference between collaborators. p-value/role-base setting (PM,GS): significance of the average differences between the no-role-based setting and the role-based setting; ANOVA p-val.: significance of the role factor ($W/Role$ vs. $W/oRole$). ($0.01 < p < 0.05$): moderately significant *, ($0.001 < p < 0.01$): significant ** , ($p < 0.001$): highly significant *****

			Mean(s.d.)						
			npq	*dt*	*nf*	*qn*	*ql*	*qo*	*nbm*
W/Role	GS	Group	1.719(1.06)	9.993(3.37)	58.522(27.13)	65.913(31.54)	4.640(1.11)	0.438(0.18)	20(14.50)
		$IGDiff^p$	-0.524	-3.469***	1.304***	2.087***	1.160***	0.139***	2.230***
	PM	Group	1.884(1.53)	10.469(3.11)	56.308(27.95)	56.308(27.95)	2.791(0.70)	0.391(0.08)	15(12.88)
		$IGDiff^p$	0.242***	1.453***	-2.423***	-1.692***	0.058***	0-0.232***	0.045***
W/oRole		Group	2.085(1.01)	13.159(3.92)	24.125(12.81)	43.583(16.28)	3.673(0.67)	0.446(0.10)	19(11.34)
		p-value/GS	***	***	***	***	***	***	
		p-value/PM	***	***	***	***	***	***	*
W/Role vs. *W/oRole*		ANOVA p-val.		**	***	**		*	



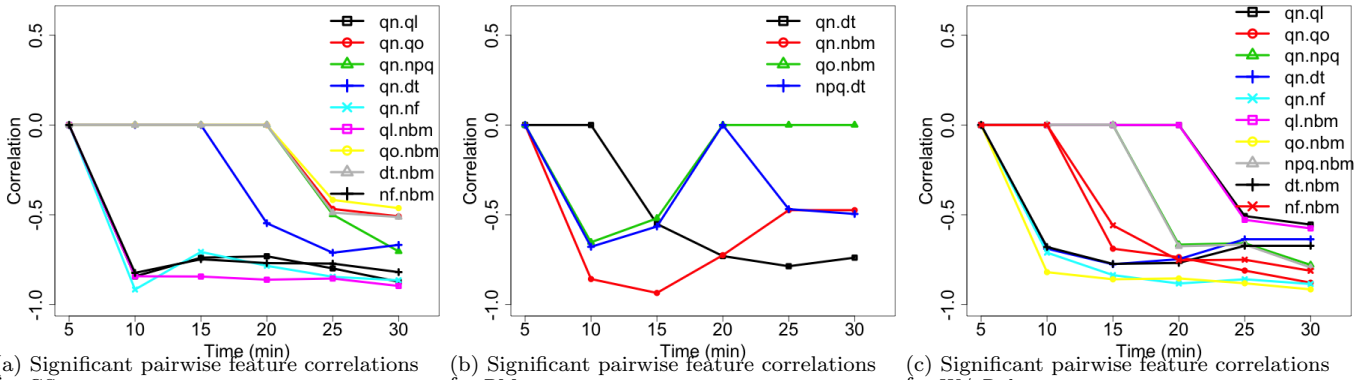
(a) Proportion of message categories for each search setting

(b) Distribution of the most important message categories over the search session timeline

Figure 2: **Communication analysis.**

search sessions differed significantly from those obtained from the $W/Role$ search scenarios for almost all features and that the differences in the search policies for the two scenarios explained most of the observed differences. In particular, we observed that groups without prior roles spent a longer time on results pages ($Mean(dt) = 13.159$) than did groups with assigned roles ($Mean(dt) \leq 10.159$), although the former annotated, bookmarked, and snipped less pages ($Mean(nf) = 24.125$) vs. $Mean(nf) \geq 56.308$). This result suggests that given that the users that performed a $W/oRole$ search scenario had no search policy, they spent more time scanning pages before selecting a result or refining the query. Therefore, in the same limited search time (30 minutes) as groups with prior roles, groups without prior roles provide less feedback on pages. This result is also confirmed by the analysis of the average number of pages visited per query (*npq*). This slow and exhaustive evaluation style has been demonstrated to be more common among less-experienced users [3]. Thus, we may conclude that the users in the $W/oRole$ scenario relied on their limited knowledge about the topic to select web pages according to the relevance of their content, whereas the others used role guidelines to quickly select or discard web pages. Through examination of the query length feature (*ql*), we can note that in the $W/oRole$ scenario, the query lengths are significantly different from those of the $W/Role$ scenario. The GS (PM) dyads had longer (shorter) query lengths than those in the $W/oRole$ scenario. To gain a better understanding of the cause of this result, we analyzed the average difference in the intra-query length features within each collaborative group involved in each scenario. We found relatively important and significant differences for GS search scenarios ($Mean = 1.116, p < 0.001$) but very low differences for both PM ($Mean = 0.058, p < 0.001$) and $W/oRole$ ($Mean = 0.549, p < 0.001$) search scenarios. As evident from the role specification, the GS pair of roles is more inclined toward verbose queries because they favor searching diversity [4]. Combining these observations, we can conclude that the differences in query length can be explained by the differences in the role policies involved in the $W/Role$ search scenarios, not by the prior assignment of explicit roles. A similar conclusion can be obtained from analysis of the query overlap feature (*qo*). Regarding communication through chat message exchanges (*nbm*), there were few, mostly non-significant differences between the different search scenarios.

To understand why participants communicate, we analyzed the chat messages, categorized as detailed above. Figure 2a reports the proportion of each message category within the communication stream of the collaborative groups in the different search settings. The general outcome of this analysis highlights that collaboration, either



(a) Significant pairwise feature correlations for GS

(b) Significant pairwise feature correlations for PM

(c) Significant pairwise feature correlations for W/oRole

Figure 3: **Collaborators' behavior-based correlation analysis.**

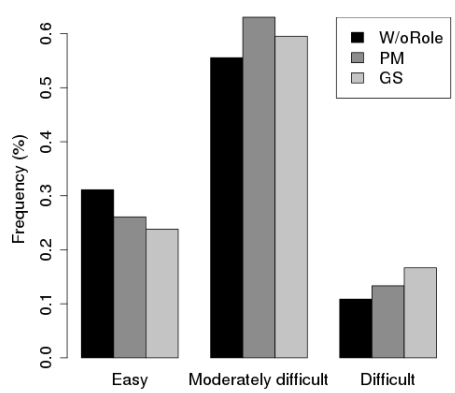
with or without assigned roles, mainly triggered coordinative and content messages. More particularly, we can observe that the most important communication category in the *W/oRole* search scenarios was task coordination (*30.43%*); the fraction of messages in this category was significantly greater than in both the *PM* (*25.42%*) and GS (*21.66%*) search scenarios. This difference reflects the specific need of the users involved in *W/oRole* to bridge the gap, through communication messages, between individual and collaborative perception of the search task as no explicit policy guided their involvement. In contrast, communicating about results was less common in the *W/oRole* searches (*28.64%*) than in the *W/Role* searches (*PM: 37.28%*; *GS: 39.33%*). This could be obviously explained by the increase of coordinative messages in the case of *W/oRole* scenario as shown above. Another assumption is that the role assignment is more functional oriented than content oriented; thus users within role condition are more likely to look for additional knowledge from their peers in order to achieve their involvement in the task. In contrast, users within no role condition are more willing to negotiate their roles according to their skills, which presumably indicates that they are more likely to have a more adequate level of knowledge (than users without prior roles) to perform the task. No apparent differences were observed for the remaining categories. To check whether this tendency was evident over the entire search session, Figure 2b shows the highest proportion of categories for each timestamp $(1-30)$. We can confirm that coordination was the main goal underlying communication for collaborators involved in the *W/oRole* scenario and that it dominated all the other categories within the entire timeline of the search. Interestingly, Figure 2b reveals that although the *PM* and *GS* scenarios were quantitatively not significantly different in terms of task coordination and content, they were qualitatively different. The users involved in the *PM* setting needed to coordinate much more than those involved in the *GS* setting. This result suggests that some roles can lead to role ambiguity, which induces overlap in the functional assignments. This may lead users to clarify, through coordination messages, their responsibilities and involvement in the search.

4.2 Analyzing users' division of labor strategies

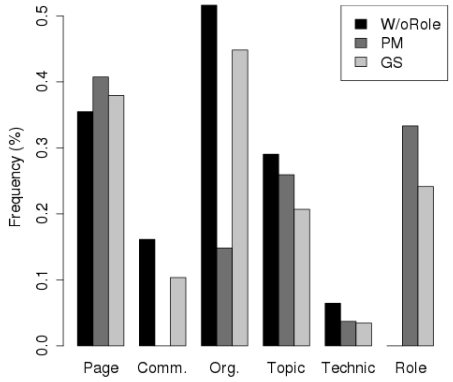
Here, we analyze how the intra-group members' search strategies differed alongside the search session among the studied search scenarios. Following [31] and according to our research hypotheses ($H1$ and $H2$), our particular objective was to investigate and contrast the complementarities of the user's behaviors toward the achievement of the collaborative task within the different settings by computing the significance of the differences between their behavioral features. To achieve this, we computed the correlation between the participants' cumulative behavior pairs of features alongside the search session for each one-minute timestamp as follows [31]:

$$C_{1,2}^{(t_l)}(f_k, f_{k'}) = \begin{cases} \rho(\Delta_{1,2}^{(t_l)}(f_k), \Delta_{1,2}^{(t_l)}(f_{k'})) & if\ p(\Delta_{1,2}^{(t_l)}(f_k)) < \theta \\ & and\ p(\Delta_{1,2}^{(t_l)}(f_{k'})) < \theta \\ 0 & otherwise \end{cases} \quad (1)$$

where $\Delta_{1,2}^{(t_l)}(f_k) = S_{u_1}^{(t_l)}(f_k) - S_{u_2}^{(t_l)}(f_k)$ represents the cumulative difference in the feature values of user u_1 and user u_2 and θ is a significance threshold. Negative correlations between the difference values, rather than the feature values themselves, at the session level signify complementarities, i.e., search skills for which collaborators exhibit differences and are effective [31]. Figures 3a, 3b and 3c represent the significant negative correlations of pairwise feature differences for the GS, PM and *W/oRole* search scenarios, respectively, for each five-minute timestamp (5-30). The sets of the negative correlations over the pairwise features evolved across the session, thus highlighting the participants' involvement. We conjecture that the stability of the correlation curve trend across time indicates behavior convergence, thus suggesting that participants achieved complementary skills at the corresponding search stage. We can observe that a higher number of significant correlated features resulted from the *W/oRole* setting (10) than in both of the *W/Role* scenarios (4). This observation indicates that self-organization provided users without roles, with a higher flexibility in their behavior leading to the emergence of a wider range of complementarity skills. We emphasize that according to Figure 3b, 2 of the 4 features implied in the PM correlations varied between negative and null correlations over the search session. For instance, the correlations between the number of pages per query (*npq*) and the page dwell time feature (*dt*) differences were not significant at the beginning of the session, became negative between the 10^{th} and 20^{th} minutes, were non-significant from minutes 20 to 25, and



(a) Proportion of users assessing task difficulty for each search setting



(b) Proportion of answers belonging to each category of difficulty across answers assessing that task is either difficult or moderately difficult

Figure 4: **Qualitative analysis of task difficulty**

finally became negative again at the end of the session. Furthermore, from Figure 3a, we can observe that in the GS scenario, 55% (5/9) of the pairwise features exhibited significant negative correlations at 20 minutes, and all of the implied features converged to significant correlations after 25 minutes. In contrast, Figure 3c indicates that collaborators in the *W/oRole* scenario, 80% (8/10) of the feature pairs reached correlation convergence earlier in the session, after 15 minutes. This indicates that roles have weakened the interdependence between the users and it appears that this fact had a negative effect on intra-group coordination stability, especially in the PM setting, in which collaborators seemed to exhibit role drift. To extend our investigation of the features that characterized collaborators' complementarities across time, we examined the pairwise features that were significantly correlated in the *W/oRole* setting but not in the GS or PM settings. These features are listed in the legends of Figures 3a, 3b and 3c. Although there was no significant difference between the GS and *W/oRole* scenarios, we can observe that the most frequent pairwise features, as indicated by the set of non-significantly correlated features within the PM setting, were those that could be connected to the role characteristics: the miner should identify the most relevant documents (a behavior that can be characterized by the number of feedbacks (nf)), and the prospector should favor diversity by submitting a higher number of queries (qn) with low overlap (qo). All of the above observations suggest the following: (1) the no role condition offers to the users a wider range (than the role condition) of opportunities allowing them to give rise to complementarities in which they are the most effective toward the achievement of the shared goal; and (2) roles do not guarantee a cohesive intra-group search.

To gain a deeper qualitative understanding of these observations, we analyzed the post-task questionnaires. Using the provided codes (Table 2, post-task question 3), we assessed the number of participants involved in each scenario that assessed the task as *easy*, *moderately difficult* or *difficult*; the results are shown in Figure 4a. As we can see, most of the users, regardless of the search scenario, found that the task was moderately *difficult*. However it seems that it was more likely that the task is assessed as *easy* when performed in a no role condition. In contrast, it was more likely that the task is assessed as *difficult* when performed in a role condition. However, a χ^2 test between the role and the difficulty level factors ($\chi^2(2, N = 150) = 0.58, p > 0.05$) found that the reported differences are not significant. Figure 4b shows the qualitative differences in terms of what made the task particularly *difficult* or *moderately difficult* within each setting as determined from the participants' answers (Table 2, post-task questions and related codes (Table 5)). We observed that the most important underlying reasons were related to communication (*Comm*), organization (*Org*) and roles (*Role*). It is clear that the latter are linked to the overall task coordination. We can conclude the following: (1) setting up useful communication and organization was particularly and relatively easy within the PM setting; this fact could be explained by the high proportion of chat messages exchanged between users with PM roles, as revealed by the communication analysis, (2) establishing roles was easy in the *W/oRole* setting, whereas fitting pre-assigned roles was difficult in the *W/Role* setting, especially in the PM scenario. Four participants involved in the PM scenarios explicitly mentioned that they had difficulty following the role guidelines (e.g., "*It is difficult to keep my role*" and "*restrict myself to my role*"). These observations confirm our previous expectations that were based on quantitative analysis of the feature-based behavior correlations. We can also observe from Figure 4b that the other types of difficulty, namely *Page*, *Topic* and *Technical*, were mentioned with similar frequencies by the participants in the different settings.

4.3 Search Effectiveness

To evaluate the effectiveness of the various collaborative search settings, we calculated some common metrics and then averaged the values over all of the groups that were in a given search setting. The common metrics used were the following [28]:

- The search precision of collaborative group g:

$$Precision(g) = \frac{RelevantCoverage(g)}{Coverage(g)} \quad (2)$$

where $Coverage(g)$ is the set of unique pages annotated, bookmarked and snipped by group g. $RelevantCoverage(g)$ denotes the set of unique annotated, bookmarked and snipped pages among those of $Coverage(g)$ with a minimal agreement level of 2 participants belonging to distinct groups.

- The search recall of collaborative group g:

$$Recall(g) = \frac{RelevantCoverage(g)}{GT} \quad (3)$$

where $GT = \bigcup_g RelevantCoverage(g)$ is the ground truth computed as the total number of unique annotated, bookmarked and snipped pages over all the search groups with a minimal agreement level of 2 participants belonging to distinct groups.

- The F-measure of a group g combining precision and recall:

$$F(g) = \frac{2 \times Precision(g) \times Recall(g)}{Precision(g) + Recall(g)} \quad (4)$$

The goal of the recall-precision analysis was to determine the impact of role assignment on the search effectiveness. Table 7 presents statistics regarding the obtained precision, recall and F-measure scores. Because the test queries differed from one search setting to another, we centered and reduced the effectiveness measures to make the comparisons of the studied settings fair. The statistical measure transformation was performed over each group within each search setting. The last row reports the significance values of the role factor within each search setting obtained by performing an ANOVA on the transformed scores. We can observe that the precision was significantly higher when searches were performed without prior role assignment ($Mean = 0.557$ vs. $GS : Mean = 0.435/PM : Mean = 0.422$); however, no significant differences were found for the recall and F-measure scores. According to the behavior analysis, we observed that users without prior roles (1) were more likely to spend a longer time on the results pages, (2) were more involved in making assessments (annotations, bookmarks, and snippets) on the results pages, which were considered as relevance signals, and (3) were less communicative about the document content suggesting that spontaneous roles are expected to emerge according to their knowledge skills. Combining all these facts about the search effectiveness and the users' behavior, we can conclude that although users in both settings were able to make fair judgments about relevance, which led to similar levels of recall, users that defined their roles themselves were more successful at discarding irrelevant page results because of their more in-depth page visits and understanding of the search topic. Thus, we hypothesize that performing searches without prior roles may lead to more precision-oriented searches.

Table 7: **Search effectiveness analysis. Mean(s.d.): Mean value of the effectiveness metric over the search session (standard deviation). ANOVA p-value: significance of the role factor ($W/Role$ vs. $W/oRole$). ($0.01 < p < 0.05$): moderately significant *, ($0.001 < p < 0.01$): significant **, ($p < 0.001$): highly significant *****

		Mean(s.d.)		
		Precision	*Recall*	*F-measure*
W/Role	GS	0.435(0.14)	0.042(0.02)	0.075(0.03)
	PM	0.422(0.16)	0.038(0.02)	0.069(0.03)
W/oRole		0.557(0.17)	0.042(0.01)	0.074(0.03)
	ANOVA/GS	***		
	ANOVA/PM	***		

5. DESIGN IMPLICATIONS

In this section, we discuss the potential future directions for designing CIS/CIR systems that are revealed by our results.

5.1 Mining latent evolving roles

Our findings indicate that users in prior-role-based settings found it difficult to coordinate among themselves according to the role guidelines. In contrast, user-driven roles enabled users to achieve a reasonable level of coordination, even though this coordination relied on intensive communication. The implication of these observations is that it would be helpful to design processes that are able to determine the most effective roles for users in a "just-in-time" fashion based on their evolving behavior and then automatically re-inject the determined roles in the ranking model to mediate the subsequent searches. Recently, a method for determining roles based on users' behavior has been proposed [31]. However, this method relies on fixed taxonomic roles that may not be ideal or relevant to users' search skills. Considering our findings, there is room for combining system-based and user-based mediation approaches to design algorithms building unlabeled and latent roles based on collaborators' differences to appropriately personalize collaborative searches. Such approaches may reduce both collaborative search drift and communication costs.

5.2 Providing task-based role templates

The qualitative analysis of communication reported in section 4.1 demonstrated that in both role-based and no-role-based search settings, the chat messages exchanged between users mostly concerned task coordination. The division of labor costs required to build the search task background over the collaborators would be small if the system offered content-based role templates (in addition to functional explicit roles if any) that are appropriate for different types of collaborative search tasks. For instance, in exploratory search tasks in which guidance in exploring unfamiliar topics is helpful, the system could play a virtual role in facilitating query refinement and mediation among collaborators by suggesting the most relevant information nuggets according to the differences in users' knowledge regarding the information landscape of the topic. Accordingly, there is an opportunity for designing systems that are able to detect the collaborative search intent and the individual collaborator's level of knowledge, and then to use this information accordingly in methods that engage the user to better drive the search.

5.3 Enhancing role-based awareness

The analysis of the division of labor strategies and post-task questionnaires highlighted the difficulty that collaborators faced in strictly following their role guidelines. Although roles seem to constrain the users too much, they can structure collaboration and reduce communication costs, and they might be required for some tasks. One underlying challenge concerns the interface design for enhancing collaborators' awareness of their ability to follow role guidelines. Studying users' search behavior across a number of taxo-

nomic collaborative tasks may lead to a set of indicator(s) associated with each pair of task-role (for instance, query overlap for prospectors) that could help CIS/CIR systems automatically detect role drift and monitor support for users in order to let them better suit with role specifications.

6. CONCLUSION

In this paper, we have presented a user study of collaborative, exploratory web-search behavior, and effectiveness. This study is the first empirical investigation of how well roles, either implicit or explicit, explain users' behavior and their methods of using division of labor to achieve a shared exploratory task. Our overall conclusion was that role, either implicitly build by the users or explicitly assigned to the users, appears to have a significant effect on users' behavior, collaborative search strategy, and search results effectiveness. The results of both the quantitative and qualitative analyses reported here particularly that: (1) users without prior roles had a more slow and exhaustive search behavior and worked less independently than users with prior roles, (2) users with no prior roles were more likely to broaden their complementarities and structure their roles relatively early in the search session while users with prior roles had difficulties to fit with their roles, (3) roles limited the precision of the search results. Finally, we discussed several methods that hold promise perspectives for improving the synergic effect of collaborative mediation and coordination. In this work, we only studied the effect of role in collaborative exploraty search within dyads, which might be different than other search settings. However, we hope that this study will improve our understanding of collaborative information behavior from the user role perspective and enable the emergence of a new generation of CIS/CIR systems that bridge the gap between system-centered and user-centered approaches to collaboration.

7. ACKNOWLEDGMENTS

We thank the undergrad and PhD students from Université Paul Sabatier who participated to the user study.

8. REFERENCES

[1] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *ACM SIGIR*, pages 3–10, 2006.

[2] B. Amidan, T. Ferryman, and S. Cooley. Data outlier detection using the chebyshev theorem. In *Aerospace Conference, 2005 IEEE*, pages 3814–3819, 2005.

[3] A. Aula, P. Majaranta, and K.-J. Räihä. Eye-tracking reveals the personal styles for search result evaluation. In *INTERACT*, pages 1058–1061, 2005.

[4] M. Benderski and B. Croft. Do longer queries retrieve more diverse results? Research report 764, Center for Intelligent Information Retrieval, University of Massachusetts, 2010.

[5] R. Day and G. Quinn. *Ecol. Monogr.*, 59:433–463, 1989.

[6] P. J. Denning and P. Yaholkovsky. Getting to "we". *Commun. ACM*, 51(4):19–24, 2008.

[7] T. Erickson. A social proxy for collective search. In *CIS Workshop at ACM CSCW*, 2010.

[8] B. M. Evans and E. H. Chi. An elaborated model of social search. *Inf. Process. Manage.*, 46(6):656–678, 2010.

[9] C. Foley and A. F. Smeaton. Synchronous collaborative information retrieval: Techniques and evaluation. In *Springer ECIR*, pages 42–53, 2009.

[10] C. Foley and A. F. Smeaton. Division of labour and sharing of knowledge for synchronous collaborative information retrieval. *Inf. Process. Manage.*, 46(6):762–772, 2010.

[11] J. Foster. Collaborative information seeking and retrieval. *Annual Review of Information Science and Technology*, 40(1), 2006.

[12] G. Golovchinsky, O. Qvarfordt, and J. Pickens. Collaborative information seeking. *IEEE Computer*, 42(3):47–51, 2009.

[13] R. González-Ibáñez, H. Muge, and S. Chirag. Let's search together, but not too close! an analysis of communication and performance in collaborative information seeking. *Inf. Process. Manage.*, 49(5):1165–1179, 2013.

[14] P. Hansen and K. Järvelin. Collaborative information retrieval in an information-intensive domain. *Inf. Process. Manage.*, 41(5):1101–1119, 2005.

[15] J. Hyldegard. Collaborative information behaviour-exploring kuhlthau's information search process model in a group-based educational setting. *Inf. Process. Manage.*, 42(1):276 – 298, 2006.

[16] H. Joho, D. Hannah, and J. M. Jose. Revisiting ir techniques for collaborative search strategies. In *Springer ECIR*, pages 66–77, 2009.

[17] C. C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5):361–371, 1991.

[18] T. Kussmann, S. Elbeshausen, T. Mandl, and C. Womser-Hacker. Discovering ellis' phases of information seeking behavior in collaborative search processes. In *CIS Workshop at ACM CSCW*, 2013.

[19] M. R. Morris. A survey of collaborative web search practices. In *ACM SIGCHI*, pages 1657–1660, 2008.

[20] M. R. Morris and E. Horvitz. Searchtogether: an interface for collaborative web search. In *UIST*, pages 3–12, 2007.

[21] P. Over. The trec interactive track: an annotated bibliography. *Inf. Process. Manage.*, 37(3):369–381, 2001.

[22] S. A. Paul and M. R. Morris. Cosense: Enhancing sensemaking for collaborative web search. In *ACM SIGCHI*, pages 1771–1780, 2009.

[23] J. Pickens, G. Golovchinsky, C. Shah, P. Qvarfordt, and M. Back. Algorithmic mediation for collaborative exploratory search. In *ACM SIGIR*, pages 315–322, 2008.

[24] S. Poltrock, R. Fidel, and H. Bruce. Information seeking and sharing in design teams. In *GROUP*, pages 239–247, 2003.

[25] K. Ryan and P. Stephen J. Division of labour in collaborative information seeking: current approaches and future directions. In *CIS Workshop at ACM CSCW*, pages 4–8, 2013.

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

[27] C. Shah and R. González-Ibáñez. Coagmento - a system for supporting collaborative information seeking. In *ASIST*, pages 9–12, 2011.

[28] C. Shah and R. González-Ibáñez. Evaluating the synergic effect of collaboration in information seeking. In *ACM SIGIR*, pages 913–922, 2011.

[29] C. Shah, J. Pickens, and G. Golovchinsky. Role-based results redistribution for collaborative information retrieval. *Inf. Process. Manage.*, 46(6):773–781, 2010.

[30] D. H. Sonnenwald. Communication roles that support collaboration during the design process. *Design Studies*, 17(3):277–301, 1996.

[31] L. Soulier, C. Shah, and L. Tamine. User-driven system-mediated collaborative information retrieval. In *SIGIR*, pages 485–494, 2014.

[32] I. Xie and S. Joo. Factors affecting the selection of search tactics: Tasks, knowledge, process, and systems. *Inf. Process. Manage.*, 48(2):254 – 270, 2012.