

# Exploring Collaborative Information Behavior in Context: A Case Study of E-discovery

Zhen Yue  
School of Information Sciences  
University of Pittsburgh  
zhy18@pitt.edu

Daqing He  
School of Information Sciences  
University of Pittsburgh  
dah44@pitt.edu

## ABSTRACT

The University of Pittsburgh team participated in the interactive task of Legal Track in TREC 2009. We conducted a pilot study to investigate into the collaborative information behavior (CIB) of a group of people working on an e-discovery task provided by Legal Track in TREC 2009. Through the pilot study, we proposed a model for understanding CIB in e-discovery.

**Categories and Subject Descriptors:** H.3.3 [Information Storage and Retrieval]: Search Process; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces—Computer-supported cooperative work, Theory and models

**General Terms:** Design, Experimentation, Human Factors

**Keywords:** Collaborative Information Behavior, Model, E-discovery

## 1. INTRODUCTION

It is common for people to collaborate when dealing with complex problem. In information seeking environment, collaboration is important and useful when the task is too complex for a single individual to tackle. E-discovery can be viewed as such type of complex information seeking task that requires the help from collaboration. In real e-discovery processes, there is a lead attorney who is in charge of overseeing a large document-review effort and for vouching for the completeness and accuracy of the produced collection. The attorney often hires an e-discovery firm or team to gather all the relevant documents from the full document collection implicated by the matter. The design of interactive task of Legal Track in TREC 2009 resembles this situation, and the research team involving in the task acts as the hired e-discovery firm, and the track organizers assign a lead attorney for each task to provide authority on the issues related to the relevance and the scope of the discovery process.

Therefore, e-discovery tasks in general, and the interactive task in the Legal Track in specific, are challenging and complex information seeking tasks that requires a group of people to collaborate. Our study conducted in the context of the Legal Track of Text REtrieval Conference (TREC) in 2009 investigated the manifestation of collaboration in e-discovery. The goal of the research work presented in this paper is to explore the essence of collaboration in information seeking process and then toward building a model for understanding CIB in e-discovery.

## 2. METHODOLOGY AND STUDY SETTING

### 2.1 Tasks and Study Procedure

Our investigation of CIB activities in e-discovery involved a

*Copyright is held by the author/owner(s).*

*CSCW Workshop on Collaborative Information Seeking, Feb 6-10, 2010, Savannah, Georgia, USA.*

group of people working on the e-discovery task provided by TREC Legal Track 2009. The objective of the task was to find all the relevant documents from Enron email collection. The Enron collection is a collection of email messages, with attachments, produced by Enron in response to requests by the Federal Energy Regulatory Commission. In our study, the participants worked on the following e-discovery topic from the legal track interactive task:

*All documents or communications that describe, discuss, refer to, report on, or relate to the Company's engagement in structured commodity transactions known as "prepay transactions"<sup>1</sup>*

To resemble the complex search process in e-discovery setting, we designed our study to have multiple sessions, and there was no time limit for the participants to finish the task. They could do whatever they could to complete the task. Once they felt that they had reached certain point and wanted to stop, we would call it a session. The participants could use as many sessions as they want until they could finally complete the task. In the end, it took our participants four search sessions and one final session of wrapping up the results to complete task. Each search session lasted about two hours and the wrapping up session took about half an hour.

Throughout each session, we observed the behaviors of our participants and took notes about the actions of the participants and any issue we found significant during the study. At the end of each session, the participants worked collaboratively to write the subtasks they had done during the session and then filled out questionnaires separately about their satisfaction to the search experience and search results. Also the participants were asked to answer several open-ended questions about the details of conducting the task and their feelings and thoughts involved after each session. The study was structured to observe and record the types of interactions between the participants when they worked as a team. Specifically we were interested in the flow of information between the two participants which contributed to accomplishing the team's task.

### 2.2 Participants and Tools

In our study, a topic authority was assigned by the TREC legal track organizer based on the topic we chose. The topic authority is a real lawyer who has extensive e-discovery experience. The topic authority played an important role because he/she defined the information need and had the right to judge the relevancy of the results. Besides the topic authority, we recruited two participants to work collaboratively as the searchers in the study. Surowiecki (2004) emphasized that one of the requirements for a successful collaboration is the "diversity of opinion" that comes from different background and expertise. Therefore, we recruited a law

---

<sup>1</sup> TREC Legal Track 2009 Topic 201

school student who is an expert with legal background and the other participant from the School of Information Sciences who is an information retrieval expert. Both of the searchers had experiences in e-discovery.

The participants were provided with a search tool and a set of collaborative tools. Each of the two searchers could use the search tool to access the collection and obtaining results. In the mean time, they could use the collaborative tools to communicate with each other or they could use a common or shared interface. Also, they had a shared space where they could store and organize their results. The topic authority shared information need with the searchers and communicated about relevancy information. However, the topic authority did not conduct search himself.

The search tool used was “email explorer”<sup>2</sup>, which is a search engine specially designed for email searching with a graphic user interface. We used two collaborative tools in this study. One was Skype, of which two functions were used: 1) sending and receiving text messages with unlimited size; 2) screen sharing remotely to show what is on one’s screen to the other. The second collaborative tool was wiki. It was designed as a common space to share search strategies, results and other information that the participant want to post and share. The two participants were in the same room, back to back, about 10 feet apart. Verbal communication was available between them. However, for the interaction with the topic authority, they had to use conference phone. Each of the searchers had a computer and they had full control of their own physical devices such as keyboard and mouse.

### 2.3 Data Analysis Method

We used qualitative method to analyze the data obtained from the study. The data included those collected in post-session questionnaires about the complexity of subtasks, the satisfaction on search experience and search results. It also included the participants’ answers to open-ended questions and the reports recorded from a focus group study after the study.

When the search task was completed, the participants and the observer formed a focus group to review the study, discuss any particular point in the study and share experience with each other. In this way, a better understanding of various behaviors in the study was obtained. The outcomes of the focus group were recorded for the qualitative analysis.

## 3. DATA ANALYSIS AND DISCUSSIONS

Overall the participants thought that the e-discovery task was a complex information seeking task for which collaboration was necessary. They claimed that they discovered more than they would have had if they had worked separately in the study. This means that the collaboration was helpful. However, the participants were not completely satisfied with the final results because they didn’t have access to the attachment (limitation of the search tool) which greatly affected the number of documents produced for the final results. They also mentioned that if they had access to the attachment, it wouldn’t be a task that they could finish within 10 hours for two people. Such tasks require much more time and even more people in collaboration in order to complete the task effectively and efficiently.

<sup>2</sup> Developed by the legal track team of the University of Maryland at College Park.

In last year’s study that we conducted for Legal Track 2008, we identified three major characteristic of CIB in e-discovery. Those are 1) frequent communication is an essential component of CIB in e-discovery; 2) the division of labor is common in the collaborative task of e-discovery; and 3) it is important for collaborators to keep an “awareness” of each other’s activities to make sure that the collaboration goes well. In this study, besides these characteristics which we still found existing and important, we also focused on the aspects that reveal the essence of collaboration throughout the whole information seeking process.

### 3.1 Subtasks Analysis

In total, there were four study sessions for information seeking and one last session for final results wrapping up. In the first two sessions, the two participants got familiar with the e-discovery task and the background information by reading the materials<sup>3</sup> and made conference phone call with the topic authority. They explored the collection and consulted the topic authority to make sure that they had a comprehensive understanding of the information need and chose the right strategy and path. At the end of the second session, they developed two search strategies. Based on the communication with the topic authority, they thought both of the strategies were correct and they decided to pursue independent paths. Therefore, in the third and fourth session, each of the participants focused on their own path and tried their best to produce results in good quality. In the final stage, the two participants worked together to combine their results.

At the end of each session, we asked our participants to summarize the subtasks they had been doing during the session. For each subtask, they needed to identify whether it was an individual subtask or a collaborative subtask and then assigned complexity level (easy, medium or difficult) to each subtask. For example, “reading prelim documents on the topic and notes from topic authority” was recognized as an easy individual subtask. “Looking for special purpose entities” was another individual subtask of which the complexity level was difficult. “Conference phone call to topic authority” was an instance of easy collaborative subtask. For complex collaborative subtask we had “Setting up new search approaches for relevant documents.” The distribution of subtasks according to sessions is indicated in the Table 1.

**Table 1: Subtasks distribution**

Subtask type		Session				
		S 1	S 2	S 3	S 4	Final
Individual Subtasks	Easy	2	1	1	0	0
	Medium	0	1	2	4	0
	Difficult	1	0	2	2	0
Collaborative Subtasks	Easy	2	2	0	0	1
	Medium	1	0	0	0	1
	Difficult	0	1	0	0	0
Total		6	5	5	6	2

Table 1 shows that all the collaborative subtasks were in the first two sessions and the final session. Most of the individual subtasks were in the third and fourth session. Although there were no

<sup>3</sup> Background information provided by Legal Track 2009 and the topic authority

collaborative subtasks identified in sessions 3 and 4, the communication between participants were still frequent as they constantly contributed to each other with valuable information. The subtasks were relatively easy in the first two sessions because they were working with the general characteristics of the task. Once they started to have to make decisions about that the relevance of the documents, it became even more difficult. This wasn't a collaborative issue but a judgment call about what they thought the topic authority wanted. Additionally, as they exploring the collection more and more in the later sessions, they found that there were indications that many more relevant documents could exist in the collection, but they didn't have the right tool to obtain and evaluate them (i.e., can't access to the attachments). This made them felt the subtasks becoming even more difficult.

For each subtask, we asked the participants to evaluate their satisfaction on the result and search experience respectively on a 1-10 scale. We found that the participants generally were more satisfied with the result and experience when conducting collaborative subtasks than individual subtasks. In particular, the score difference for difficult subtasks was the biggest. This indicated that collaboration produced much better result and experience for complex tasks.

### 3.2 Collaboration Advantages and Costs

In general, the e-discovery task is a complex information seeking task that requires collaboration. Collaboration, therefore, had shown its advantages. No individual participant in our study could claim to have comprehensive knowledge of the task. That's why the participants felt that the collaboration made the completion of the task more efficient and effective, and they felt more satisfied with the collaborative results and experience. One participant mentioned that the collaboration made the learning of search lessons twice as fast. Whenever one participant learned something first, the other one could benefit by just knowing the fact. In addition, we found that the different background of the two participants contributed to the success of collaboration. For example, one participant was more familiar with legal terminologies and it was easy for him to come up with synonyms. The other participant felt that those legal terminologies helped him a lot during the search. Then, the second participant who is an information retrieval expert used network analysis to assist the information searching and this strategy inspired the first participant to produce more relevant results.

However, in spite of all these advantages, we still need to pay attention to the extra cognitive loads and collaboration costs compare to individual information seeking. As in a collaborative environment, the tools are more complex and information flow is not just between system and users, but also between users. In the remaining of this section, we will present some discussions on the costs of collaboration we found in our study.

**Adaption to the tools.** Many researchers have attempted to address the issue of cognitive load induced by a system (Fidel etc, 2004). Compare to individual information seeking where user only needs to interactive with the search tool, users in collaborative information seeking need to interact with the collaboration tools as well. In our study, the collaboration tool we selected were Skype and Wiki which our participants were familiar with already. However, knowing how to use a tool does

not necessarily means there is no extra cost adopting the tool in the task. Our participants pointed out a disadvantage of the tools that there were not enough screen space for all the search tools and collaboration tools. Sometime they focused on one screen and totally neglected what was going on with the other tools. A very interesting phenomenon in our study was that, for the first 2 sessions, the two participants used Skype to send and receive messages a lot. However, in the following sessions, the usage of Skype text exchanging became less and less. This means our participants didn't adopt Skype in a long run as they discovered that Skype actually slow down their communication pace.

**Collaboration between searcher the topic authority.** This is a type of collaboration which is unique to an e-discovery scenario. In a typical individual information seeking scenario, the information need is often initiated by the searcher himself/herself. In a e-discovery task, the information need is not proposed by the searcher but the topic authority. Therefore, there will be an extra cost of "transferring" the information need. This is a very important step for the whole task as it decides the success of the following search steps. In our study, because the topic authority could only be remotely accessed by phone calls, sometime the searchers felt that there were not enough in time support from the topic authority. At the initial stage, the discussion with topic authority was very helpful for the searchers to understand the task requirement. However, in the actual search process, they felt that they could make better use of the support from the topic authority for relevancy judgment if the support can be accessed immediately.

**Collaboration between searchers.** In individual information seeking, one only need to focus on his/her own activity. However, in collaborative information seeking, there could be several sources of extra cognitive load: (1) contributing to each other (sending or posting important messages that they think their partner should know), (2) keeping aware of what the other people is doing (3) pay attention to the group history in addition to the individual search history. While these activities may be the key beneficial of collaboration for the completing of task, they are still extra cost for users. Therefore, a balance between the collaboration benefit and cost is crucial to the success of collaboration. Also, there is an issue of getting familiar with each other. At the beginning of our study, our two participants used text to chat a lot, verbal communication was very rare. Later, the verbal communication occurred more and more frequently and eventually almost replaced the texting. This means there was a process of getting familiar with each other and gaining trust. Although in our study, the unfamiliarness did not cause untrust because they believed each other's expertise in their own field, this may not be the case for other collaboration scenarios.

## 4. AN INITIAL CIB MODEL FOR E-DISCOVERY

Based on the result analysis of the pilot study, we proposed the following initial model for collaborative information behavior in e-discovery in Table 2. We found that there were stages in collaborative information seeking. Users often started the task with intensive collaboration. As in this stage, they need to explore the task, make sure they have the same understanding of the information need and set up search strategies. Once they were clear with the executing process of the task, a division of labor

**Table 2 CIB model**

<b>Stages</b>	Initial & Exploring -----> Division of labor -----> ..... ----->Final result		
<b>Cognitive Load</b>	-----> adaption to the system -----> getting familiar with other collaborators		
<b>Collaboration Levels</b>	Collaborative -----> Individual -----> Collaborative ↑↓ Coordination                      Communication Cooperation                      Contribution		
<b>Collaboration Type</b>	Synchronous -----> Asynchronous -----> Synchronous		
<b>Subtask complexity</b>	Easy -----> Difficult -----> Easy		
<b>Types of supports needed</b>	Support for frequent communication and awareness	Implicit collaborative support (recommendation); Support for relevancy judgment (topic authority)	Shared space for common results

was common for the next stage as they would like to allocate subtasks and pursue paths independently. Although in our study, the exploring stage and division of labor stage only occurred once, in a general case, the transition of stages is not necessarily to be one-direction. Circles may occur from the division of labor back to exploring.

When users felt that they had reached certain point of completion on searching, there would be a final stage for results wrapping up when they combined their results together. During the whole process, the cognitive load of collaborators was not only reflected as adaption to the system, including searching tools and collaborative tools, but also reflected as getting familiar with other collaborators, including gaining trust and making contribution to each other.

Shah (2008) proposed a five layers model of collaboration: (1) communication (2) contribution, (3) coordination (4) cooperation and (5) collaboration. Each of the higher layer requires more cost than the lower layer. In our study, we found that different levels of collaboration existed in different stages. At the initial stage when the main goal was to explore the dataset and set up search strategies. It involved more coordination and cooperation. In this situation, synchronous collaboration support was needed as the participants need to keep awareness in time. However, when the participants set up search strategies and decided to pursue different paths, communication and contribution played more important roles. Under such circumstance, asynchronous collaboration support was needed as the participant need to focus on their own task.

In our study, the subtasks in initial and exploring stage were recognized as relatively easier than that in division of labor stage. This means in synchronous collaboration when collaborators were making decisions together, the subtasks were easier to complete as long as the communication and awareness were well supported. For asynchronous collaboration, when user need to make judgment call by himself/herself more often, the subtasks were harder to complete and they could really make better use of extra implicit collaboration support like recommendations and support for relevancy judgment. In the final stage for result wrapping up, the subtasks were relatively easy as the collaborators only need to combine the results and solve controversial problem if any. In this

stage, a shared space for common results was very helpful for supporting the collaboration.

## 5. CONCLUSION

In this report, we present a case study of CIB in the context of e-discovery. Through participating in the interactive tasks in Legal Track, we investigated the manifestation of collaboration in e-discovery. Based on the analysis of results obtained from this pilot study, we proposed an initial model for understanding CIB in e-discovery. The model incorporates collaboration type, collaboration level and subtask complexity into different stages. It reveals the essence of collaboration in information seeking process. Also, this model could be used to recognize different types of supports needed for different stages in a collaboration information seeking process.

The limitation of this model is that it is based on a single case study. In the future, we planned to conduct large-scale surveys among employees in e-discovery firms to verify and improve the model.

## 6. ACKNOWLEDGEMENT

This work was partially supported by the National Science Foundation under Grant No. 0704628.

## 7. REFERENCES

- [1] Fidel, R., Pejtersen, A. M., Cleal, B. & Bruce, H. (2004) A multidimensional approach to the study of human-information interaction: A case study of collaborative information retrieval. *Journal of the American Society for Information Science and Technology*, 55(11):939-953.
- [2] Hyldegard, J. (2006) Collaborative Information Behaviour - Exploring Kuhlthau's Information Search Process Model in a Group-based Educational Setting. *Information Processing and Management*, 42(1):276-298.
- [3] Shah, C. (2008) Toward Collaborative Information Seeking (CIS). In: 1st International Collaborative Search Workshop, Pittsburgh, PA, USA.
- [4] Surowiecki, J. (2004) *Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. Doubleday Publishing.