Discovering Ellis' Phases of Information Seeking Behavior in Collaborative Search Processes

Tanja Kußmann Stefanie Elbeshausen Thomas Mandl Christa Womser-Hacker
Department of Information Science &
Natural Language Processing
University of Hildesheim
Marienburger Platz 22,
31141 Hildesheim,
Germany
{kussma, elbesh, mandl, womser}@uni-hildesheim.de

ABSTRACT
In this paper, we describe the analysis of collaborative information seeking behavior of university students through a user study based on SearchTeam. The analysis of system interactions applies the behavioral model of information seeking strategies by David Ellis (1989) for collaborative processes. The results show that single phases from the model are applicable on collaborative processes, but the characteristics of collaborative information seeking call for an extension of the model. Especially social components must be considered. Further we could identify two different search strategies, the scanning-strategy and the reading-strategy. Based on our findings we can state that the scanning-strategy led to better results regarding search effectiveness.

Categories and Subject Descriptors
H.3: INFORMATION STORAGE AND RETRIEVAL **H3.3: Information Search and Retrieval**: *Search process, Information Filtering* **K.4.3 Organizational Impacts**: *Computer-supported collaborative work.*

General Terms
Measurement, Performance, Experimentation, Human Factors, Verification

Keywords
Collaborative Information Seeking, Evaluation

1. INTRODUCTION
It is often argued that collaboration is a helpful method for solving complex problems [3]. In the area of information seeking however, collaboration is still understudied [11].Unlike the study of information seeking performed by individual users, little is known about the way users seek information collaboratively. There exists only a few information seeking models which accommodate collaboration. Nevertheless, collaborative information seeking (CIS) is an everyday task especially in the area of teaching and learning [8]. In this paper we relate the phases of information seeking behavior from Ellis' model [4] to collaborative search processes. We show that some of the phases described by Ellis for individual information seeking behavior are valid for CIS-processes while others do not fit in social search processes. We also show that the characteristics of CIS call for an extension of the model especially regarding social components. Furthermore we provide an analysis of the search performance and the acceptance of the used system.

In section 2 we review some of the relevant literature on CIS. The focus of this review lies on search strategies in CIS and collaboration styles. Another emphasis lies on Ellis' model of information seeking behavior (ISB = Information Seeking Behavior) to provide the foundation for the conducted study. To survey the fitting of Ellis' model for CIS processes we conducted a laboratory study with 15 users who were organized in teams of three. The details are given in section 3, in section 4 we present the results of the study. We describe how Ellis' ISB-phases match CIS-processes when teams search collaboratively and with system support. Further we provide an overview of the role allocation during the collaboration process and the effectiveness and efficiency of search strategies used. In the last section a conclusion is given together with implications and directions for future research.

2. BACKGROUND
In the following section a brief overview of related research is presented to situate the present study in the context of Information Seeking and especially CIS. We also present a more detailed description of Ellis' behavioral model of information seeking.

2.1 Related Work
So far there exists no clear conceptual definition of collaboration and no clear dissociation from related concepts, i.e. cooperation [15]. In order to reach a suitable definition several authors like Shah [15] or Fidel et al. [5] defined conceptual implications which can be used as base for developing definitions of collaboration. A possible definition of collaborative information search is

"[…] activities that a group or team of people undertakes to identify and resolve a shared information need." [13]

Accordingly a shared or similar information need is a necessity for collaboration and people need to explicitly decide in favor for collaboration. The result of collaboration is a connection of individual contribution of team participants. Knowledge is collectively created in this context and therefore different from knowledge a single actor would create [8]. Especially in situations where problems are too complex to be solved by an individual it is beneficial to collaborate [14]. That also counts for collaborative search. For instance, Pickens et al. [12] and Shah et al. [18] discovered that collaborative search with algorithmic mediation to enhance the collaboration process, lead to better results than those of individual search.

Capra et al. [1] identified among other things two different styles of collaboration during a study examining the user behavior in IR-processes. *Directed Collaboration* was especially common in academic or corporate groups. Usually one person leads the search while the team members conducting it. In *Tightly Coordinated Collaboration* team members split the search task. That is especially the case in symmetric collaboration scenarios where all team members have the same power and responsibilities. Directed Collaboration on the other hand refers to an asymmetric setting with different power, responsibilities and skills of the participants [14]

In our preliminary study we examined the search behavior of four teams working collaboratively with two participants within each team [10]. Two of the teams split the search task and each member of the team worked individually on its own task. That indicates a *Tightly Coordinated Collaboration* which implies a symmetric role allocation. The other two teams did not split the work task and searched for information simultaneously. In the current study we checked among other things, the role allocation in teams with three participants.

Hyldegard [6, 7] studied the applicability of Kuhlthau's information search process [9] in the context of group searching. The study exposed that some of the stages of Kuhlthau's model apply to CIS but they did not cover the social dimensions of collaboration. Also Shah & González-Ibañez [17] tried to map the stages identified by Kuhlthau to CIS with similar results as in Hyldegard's study.

2.2 Ellis' behavioral model of information seeking

As in Kuhlthau's model and most research in IR Ellis' behavioral model of information seeking concentrates on single users [4]. In principle it is possible to transfer Ellis' model on different information environments [4], i.e. hypertext-based environments as the internet [2]. The results are based on a study with scientists and researchers but it is possible to map the phases to other user groups, i.e. attorneys [2]. A modification of the model was developed by Wilson [20] who linked the phases in the following order:

- *Starting*: all activities connected with the beginning of a search process
- *Chaining*: continuation of the search via tracing hints and links
- *Browsing*: semi-targeted search in promising areas
- *Differentiating*: qualitative filtering of the material
- *Monitoring*: observing a collection of sources to detect alterations
- *Extracting*: systematic examination of one specific information source
- *Verifying*: examination of information regarding correctness
- *Ending*: ending of the search and linking of collected information

It should be recognized that according to Ellis the order of appearance of the phases depends on the search context.

3. METHOD

We accomplished a laboratory study involving a total of 15 participants that were split up in collaborative teams with three participants each team. The following section describes the subjects, the system used, the study procedure and the task.

3.1 Subjects

15 students from the University of Hildesheim participated in the study. 14 of them were enrolled in the degree course International Information Management (IIM). The participants were randomly assigned to groups with 3 students in each group. As compensation for participating in the study they had the possibility to win prize money, 40€, 30€ or 20€ or one of three gift coupons from a local book store worth about 10€.

3.2 System

For our study we used the collaborative search engine SearchTeam.[1] It was developed by Zakta, an US-american company specialized on social applications, in 2010.[2] SearchTeam is a publicly available and cost free tool which provides support for collaborative information search. Users can save, manage and share information collaboratively. All user specific search-activity is saved in a personal account. All search-activities take place in so called SearchSpaces which are the starting points for all activities in the system. A SearchSpace constructed by a user team always represents one topic which is the actual area of interest. At the same time it supports asynchronous collaborative search because when a user reconnects to the system he or she always sees which activities took place in the SearchSpace during his or her absence first. Since the system is web-based the users are not only time- but also location-independent.

The application provides different functionalities. Within the SearchSpace users have the possibility to enter a query which is answered by getting a result list, similar to common web-search engines. If the user types a misspelled query, the system automatically provides proposals for improvements. If possible SearchTeam also gives recommendations for connected topics which can be chosen by the user. Search results which are marked as relevant by the collaborators can be saved in a file. The users may create new files for every topic or sub-topic and if one user saves a search result the other team members have the possibility to comment the results or rate them. All files and results can be modified, renamed or rearranged by the collaborators and they can save additional items which were not found via SearchTeam.

[1] Publicly available from http://searchteam.com/

[2] http://blog.searchteam.com/

Another functionality of SearchTeam is the chat which can be used for synchronous communication. Saved chat data includes the exact time and the participants so that team members who were not online can comprehend the chat activities. To follow only the activities of team members the system also provides an activity-history. This is implemented in a permanent bar on the right hand of the screen which shows all activities of all team members in the specific SearchSpace.

3.3 Session workflow

To study the collaboration supported by SearchTeam and to focus on the effectiveness of the search strategies as well as the applicability of CIS on Ellis' ISB-model, we conducted a study with the following session workflow. Conditions were the same for each participant and each collaboration team.

Each test was conducted in one of the computer labs at the University of Hildesheim. The participants worked in groups of three people while every participant used one personal computer for the study. Since it should at least be simulated that the participants are not co-located, they were placed on distant computers. The participants were asked to only communicate via the system. The interaction with the system was recorded via Camtasia Studio 7[3] for further analyzes of the search performance and user behavior.

Before the start of the actual test SearchTeam was opened on each computer with an own account for each participant. The accounts were configured with special email addresses so the test participants were not forced to use their private email addresses for the test.

Participants were asked to fill in a pretest questionnaire to collect demographic as well as test-related data. Test-related data was especially collected to get information about former experience with the system and with group work in general.

After the participants filled in the pretest questionnaire they got an oral introduction of SearchTeam. Afterwards the participants were asked to work on the task which is described in section 3.4. The task was to be completed within 40 minutes. After that the participants had to fill in a post-task questionnaire on subjective statements regarding their experience with SearchTeam.

3.4 Task

We decided to choose a task which is possible to finish within a maximum of 40 minutes to not overstrain the test participants. But to sustain the artificial laboratory situation as much as possible we chose a task which is not too trivial. Further it should involve a complex problem with an inherent reason for collaboration and to allow monitoring of the way people solve complex problems in teams. For students it is kind of natural to work in teams. All of the participants pointed out that they already had experience with group work. According to this we chose a scenario in which the participants had to make a power-point presentation for a university course called "Computer mediated communication". Computer mediated communication is part of the study program for IIM-students. This scenario concerns the reality of the test participants so that it was kind of natural for them to fill in the role of a student searching and working collaboratively. The scenario was divided in two subtasks:

1. You and your Teammates decided to make a presentation about social networks. To find the relevant information you decided to just use SearchTeam. You try to find and save as much relevant information as possible. The presentation should contain information about the following aspects:
    - Historical development and spreading of social networks
    - Statistics regarding the acceptance of social networks
    - Funding of social networks
    - Areas in which social networks are used
2. Make a presentation out of the information you collected which contains at least one slide for every aspect.

4. RESULTS AND DISCUSSION

In this section we present the results of our study and the related discussion. First of all we analyze and discuss the possibility to transfer the phases of Ellis' model on the information processes with SearchTeam. Then we present a short overview about role allocations found by our study. The next section deals with the effectiveness of the different search strategies used.

4.1 Transfer of Ellis' phases on the information processes with SearchTeam

Some but not all phases of Ellis ISB-model could be identified in the collaborative search. *Chaining* and *Verifying* played just a secondary role. Few participants followed the links they found on the result pages (*Chaining*). Most of them just used the pages directly found by SearchTeam. Only one participant pointed out that she would like to get further information regarding the year of one of the statistics found (*Verifying*).

Activities of the test participants focused on the phases *Starting*, *Browsing*, *Differentiating*, *Extracting* and *Ending*.

Starting refers to all activities at the beginning of the search. In this phase we could monitor active communication i.e. initial hellos between the team members. Also the team members verified that all members are online before they began searching which indicates group awareness.

During the *Browsing* phase we could monitor the highest amount of activities. In all groups participants browsed the SearchTeam result lists or followed headings or indexes on the result pages. The phases *Browsing* and *Extracting* were sometimes interwoven. This was especially true for participants who followed the reading-strategy (4.3). They switched frequently from *Browsing* to *Extracting* and/or *Differentiating*. A clear cut into different phases is therefore not present in CIS.

Differentiating played a major role during the creation of the presentation because at that stage participants had to examine and rate the information found by their team members. Only results which were rated as relevant were linked and adopted in the presentation which can be seen as part of the phase *Ending*.

Monitoring as described by Ellis could not be identified. That is actually not surprising since *Monitoring* is an activity which usually takes place in long term projects. In our study we could identify that team members instead monitor the search activities of their team members. Each team sent an average of 45 messages

[3] http://www.techsmith.de/camtasia.asp

via chat to monitor search activities of collaborators. Social activities are not considered in Ellis' model because it was constructed for individual search. Even if it is possible to adopt some of Ellis' phases it is necessary to consider that they do not exactly match with the search stages in CIS. Sometimes the phases are interwoven or they do not appear. Especially for implicating the social activities it is necessary to build a model which actually matches the phases in CIS.

4.2 Role allocation

A division of the tasks between the team members could be identified in every group. The given topics were adopted and split up between the team members which falls into the realm of tightly coordinated collaboration [1]. This type of collaboration implies a symmetric role allocation which means same responsibilities and equal power balance between the team members. An explanation for this is that all participants had more or less similar experience and skills regarding the task topics.

The team members worked individually on their subtask but a coordination of team activities mostly via chat was observed during the whole session. They mainly communicated strategies and progress of their own and the team's progress.

The strongest reference to activities of group members took place during the second part of the scenario. For the development of the power-point presentation it was inevitable to discuss the findings of the group members because every participant had to build at least one slide to every given topic.

4.3 Effectiveness of search strategies

We could identify two different search strategies: scanning strategy and reading strategy.

The scanning strategy refers to a more cursory examination of the result documents, i.e. participants read the headings and single words on the page but they did not read the whole document. Participants further scrolled through the whole result page, saved it and left for another result page quickly.

The reading strategy on the other hand refers to a more thorough reading of the result documents. We could monitor that by the Camtasia screencast, i.e. the mouse-cursor dwelled on single text passages for a while which implies a more detailed reading. Furthermore the result pages stayed open for a longer time and the participants tended to copy single text phrases and save them in the files rather than simply saving the whole page. This text phrases were also often commented and rated by the team members.

The effectiveness of the search strategies was measured on the basis of recall and precision. The task for the study contained a complex problem with four subtasks. Complex problems lead to information search where searchers need to verify more than one relevant document, i.e. there is no predefined quantity of relevant documents. For our calculation we implied that the documents saved by all teams are the relevant documents. Accordingly our formula for the recall is:

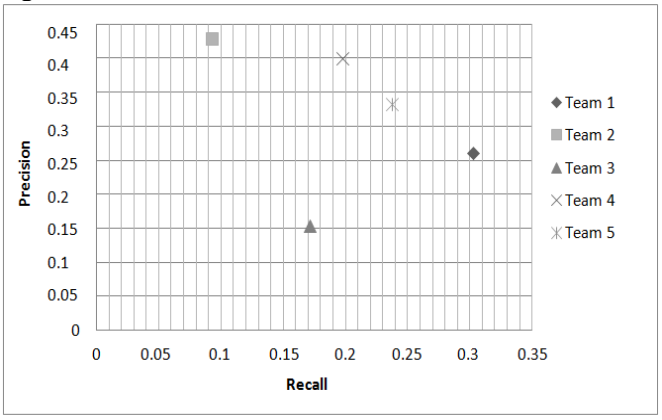
$$Recall = \frac{Number\ of\ saved\ documents\ of\ the\ particular\ team}{Number\ of\ saved\ documents\ of\ all\ teams}$$

For the precision we supposed that a document is just relevant when at least one other team saved the same document. Accordingly we used the following formula for the precision:

$$Precision = \frac{Number\ of\ matching\ documents\ \text{with at least one team}}{Number\ of\ saved\ documents\ of\ the\ particular\ \text{team}}$$

We used the data from the SearchTeam activity history for our calculation. The recall and precision values for each team are shown in figure 1.

Figure 1: Recall and Precision for the collaboration teams



For calculating the effectiveness we used Van Rijsbergen's effectiveness measurement (E) [19]. We assumed that for the given scenario α should be 0.5 because the effectiveness is balanced between recall and precision. Accordingly the effectiveness is measured as:

$$E = 1 - \frac{1}{\alpha\left(\frac{1}{Precision}\right) + (1-\alpha)\frac{1}{Recall}},$$

while the highest effectiveness is reached with $E = 0$ and the lowest with $E = 1$.

The results of the effectiveness measurement are shown in table 2.

Table 2: Results of the measurement of *E*

	***E*-Measure**
Team 1	0.720
Team 2	0.848
Team 3	0.838
Team 4	0.736
Team 5	0.723

The results imply that Team1 got the most effective search strategy followed by Team 5 and 4. These results are especially interestingly correlated with our findings regarding the used search strategies. Every group member of Team 1 followed the scanning strategy. In Team 5 and 4 it was always one participant who used the reading strategy and two who used the scanning strategy. Team 2 and 3 used mainly the reading strategy.

It should be recognized that Team 2 even if it had the less effective search strategy, got the highest precision (Table1). The low effectiveness in this group can be explained by the low recall. Team 2 saved a total of just 7 links while Team 1 saved a total of 23 links. This is explainable by a lack of group coordination: Team 2 split subtask 1, 2 and 3 between the group members and decided to solve the last one together. For this topic they compiled a file but they did not make further arrangements about

when to start with the last topic or how to exactly coordinate group activity.

5. CONCLUSION

Explicit collaborative information seeking was neglected for a long time in information retrieval and information seeking research. Current studies show that collaborative work is often requested and useful as well as present in new technologies and tools for supporting information search.

Our study mirrors these indications. 80% of the test participants attest great usefulness to teamwork for study purposes. 67% of them state that it was easier for them to finish the task collaboratively than it would have been individually.

Especially the split of the work task can be beneficial for collaborative search since the complexity of the task is usually reduced. All groups followed this splitting-method and it is assumed that they could cover more aspects in the specified time span than individual users in the same time.

We could identify two different search strategies: scanning and reading. The groups that followed the scanning strategy were characterized by a high recall, i.e. they found more documents which were less relevant. The reading groups had a higher precision, i.e. they got less but more relevant documents.

Further we could identify some but not all of the phases from Ellis' Model of information seeking behavior in the CIS-process. *Monitoring*, *Chaining* and *Verifying* could not be observed. One reason could be that the duration of the search process was too short. For following studies we plan to monitor collaborative search activities over a longer period of time to recheck if *Monitoring*, *Chaining* and *Verifying* may be detected clearly then. Also *Browsing* seems to have a particular importance because that was the phase with the strongest activity of all participants.

The study reported here was conducted with a synchronous task. For further studies it would be interesting if our findings are also applicable for asynchronous tasks. Likewise, as mentioned above, monitoring CIS activities over a longer period of time could bring other interesting findings.

A limitation of the study was the time restriction. Not all groups were able to solve the task in time. One reason was that collaborators often waited for chat reactions of their team mates before they carried on searching so that they needed a lot of time to check the findings of other group members. This "idle time" should be considered when designing scenarios for synchronous collaborative search.

6. ACKNOWLEDGMENTS

We thank the students who participated in our study for their time and effort.

7. REFERENCES

[1] Capra, R., Marchionini, G., Valesco-Martin, J., Muller, K. 2001. Tools-at-hand and learning in multi session, collaborative search. In *Proceedings of the 28th international conference on Human factors in computing systems (CHI '10),* ACM, New York, 951-960.

[2] Choo, C.W., Detlor, B., Turnbull, D. 2000. Web Work: Information Seeking and Knowledge Work on the World Wide Web. Dodrecht u.a.: Kluwer Acad. Publ. Information science and knowledge management: ISKM.

[3] Denning, P. J. & Yaholkovsky P. 2007. Getting to "We". *Communications of the ACM, 51 (4),* 19-24.

[4] Ellis, D. 1989. A behavioural approach to information retrieval design. In *Journal of Documentation*, 46, 318-338.

[5] Fidel, R., Bruce, H., Pejtersen, A. M., Dumais, S. T., Grudin, J., & Poltrock, S. 2000. Collaborative Information Retrieval (CIR). In *The New Review of Information Behaviour Research, 1 (1)*, 235-247.

[6] Hyldegard, J. 2006. Collaborative information behaviour - exploring Kuhlthau's Information Search Process model in a group-based educational setting. In *Information Processing and Management*, *42*, 276-298.

[7] Hyldegard, J. 2009. Beyond the search process - exploring group members' information behavior in context. In *Information Processing and Management*, *45*, 142-158.

[8] Kölle, R., Langemeier, G., Semar, W. 2006. Programmieren lernen in kollaborativen Lernumgebungen. In *Effektive Information-Retrieval-Verfahren in Theorie und Praxis: Ausgewählte und erweiterte Beiträge des Vierten Hildesheimer Evaluierungs- und Retrievalworkshop*, Mandl, T., Womser-Hacker, C., Ed., UVK, Universitätsverlag, Konstanz, 147-170.

[9] Kuhlthau, C. C. 1991. Inside the Search Process: Information Seeking from the User's Perspective. In *Journal of the American Society for Information Science and Technology*, 42(5), 361-371.

[10] Kußmann, T. 2012. Vorstudie zur Analyse der Studentischen Zusammenarbeit unter Einsatz von SearchTeam. Technical Report. University of Hildesheim.

[11] Paul, S.A. 2010. Understanding together: Sensemaking in collaborative information seeking. Doctoral Thesis, The Pennsylvania State University.

[12] Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P. & Back, M. 2008. Algorithmic Mediation for Collaborative Exploratory Search. In *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*, 315 – 322.

[13] Poltrock, S., Grudin, J., Dumais, S., Fidel, R., Bruce, H., Pejtersen, A.M. 2003. Information seeking and sharing in design teams. In *Proceedings of the ACM Conference on Supporting Group Work (GROUP '03),* ACM Press, 239-247.

[14] Shah, C. 2009. Lessons and Challenges for Collaborative Information Seeking (CIS) Systems Developers. In *GROUP 2009 Workshop on Collaborative Information Behavior*. Sanibel Island, Florida.

[15] Shah, C. 2010. A Framework for Supporting User-Centric Collaborative Information Seeking. Doctoral Thesis, University of North Carolina, School of Information &Library Science.

[16] Shah, C. 2010. Working in Collaboration – What, Why and How? In *Proceedings of Collaborative Information Retrieval workshop at CSCW 2010.* Savannah, GA.

[17] Shah, C., & González-Ibañez, R. 2010. Exploring Information Seeking Processes in Collaborative Search Tasks. In *Annual Meeting of the American Society for Information Science*. Pittsburgh, PA.

[18] Shah, C., Pickens, J. Golovchinsky, G. 2010. Role-based results redistribution for collaborative information retrieval. In *Information Processing and Management, 46 (6)*, 773 – 781. doi: 10.1016/j.ipm.2009.10.002

[19] Van Rijsbergen, C.J. 1979. Information Retrieval. Butterworth-Heinemann, Massachusetts.

[20] Wilson, T.D. 1999. Models in Information Behaviour Research. In The Journal of Documentation, 55(3), 249-270.