A Method to Evaluate the Synergic Effect in Collaborative Information Seeking

Roberto González-Ibáñez Chirag Shah Muge Haseki
School of Communication & Information
Rutgers, The State University of New Jersey
4 Huntington St., New Brunswick, NJ 08901, USA
{rgonzal, chirags, mhaseki}@rutgers.edu

ABSTRACT
One of the strongest appeals of collaboration is that it allows collaborators achieve something greater than the sum of their individual contributions. This notion, which we referred to as synergy, has been understudied in information seeking domain. To address this shortcoming, here we outline an approach to study and evaluate the synergic effect in collaborative information seeking (CIS). Our main contributions consist of a method of evaluation, a set of measures, and a system for experimentation. This approach has been applied to several of our recent works relating to CIS, and shown to be very promising for studying user-driven collaborative projects.

Categories and Subject Descriptors
H.3: INFORMATION STORAGE AND RETRIEVAL **H.3.3: Information Search and Retrieval**: *Search process*; **H.5.3** [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*Collaborative computing*, *Computer-supported cooperative work.*

General Terms
Experimentation; Human Factors; Measurement.

Keywords
Collaborative information seeking; Synergic effect; Evaluation.

1. INTRODUCTION
Collaboration is typically required when solving hard problems (e.g.,[1],[2]). Interestingly, though information seeking can be classified as a hard problem in several circumstances, IR researchers have focused mainly on improving the tools that supports the information search process of users as individual units. On the other hand, few have explored the implications of human collaboration in information seeking. This paper, describes a method to study the information search process of teams when working on an exploratory search task. The method consists of three main components: (1) a procedure to evaluate the synergic effect, (2) a set of measures, and (3) a system for experimentation. We have successfully applied this research approach to study different collaborative conditions in an exploratory search task [20]. Some of the results derived from this study are briefly discussed later.

The following section contains a brief review of the relevant literature on this matter. Section 3 describes the method of conducting a user study and computing various measures relating to synergy in CIS. In Section 4 we present some insights from our results so far of using this framework. We conclude in Section 5 with the implication of the framework, the results, and future work.

2. BACKGROUND
Synergy, from the Greek *synergos* (which means working together) is when two or more things, individuals, disciplines, and so on, interact producing a result that is greater than the sum of the individual contributions of the involved parts [1]; this is commonly explained through the expression *1+1>2*. Synergy has been widely studied in a variety of contexts, which includes chemistry, medicine, companies, and learning [1]. In the particular context of collaborative information seeking (CIS), little is known about the synergic effect and to what extent, if any, it affects users' behaviors, the processes, and the results produced by individuals searching information in collaboration, in comparison to those working individually.

During the past decade, several researchers have explored different aspects of CIS in both naturalistic and experimental settings. The focus, however, has been on describing users' behaviors as well as the information seeking processes of teams. For example, Hyldegard [9][10] and similarly, Shah and González-Ibáñez [19] studied the applicability of Kuhlthau's information search process [12] in the context of groups.

More importantly, it has been recognized that collaborative search is a phenomenon widely present in social context where collaboration - in a more general sense - is required. For example, Morris [13] conducted a survey that revealed that knowledge workers engage in a variety of activities in which they collaborate searching for information. Through this survey, the author also identified obstacles that impact the collaborative information seeking process, different tasks that motivate collaborative search, and common methods that are employed to share search results during the process.

Specific studies have been conducted with the aim of comparing individual and collaborative search. For instance, Joho *et al.* [11] compared the search process of single users with collaborative-concurrent search. Through this study, the authors found that those working collaboratively were able to reduce overlapping in terms of the webpages covered during a recall-oriented search task; however, as the authors pointed out, this redundancy

CIS'11, October 12, 2011, Mountain View, California, USA.

reduction did not improve the retrieval effectiveness. In a related study, Pickens *et al.* [15] and Shah *et al.* [21] found that collaborative search, with algorithmic mediation to enhance the collaboration process among participants, end up with better results than those obtained by merging of single users' results. Similarly, Foley [4] demonstrated that in order to enhance the performance in synchronous collaborative information retrieval, it is necessary to have an appropriated division of labor and also a mediated support for sharing knowledge. A comprehensive framework for evaluating the effectiveness of collaboration in information seeking, especially the synergic effect, is missing. Here, we provide a summary of our investigations in this direction that includes a procedure and a system for conducting a suitable lab study, as well as a set of evaluation measures.

3. METHOD

3.1 General Procedure

In this section we describe a procedure to experimentally evaluate the synergic effect in collaborative information seeking settings. The method is inspired by the expression *1+1>2*, in which each unit is represented by an individual and the addition operator expresses collaboration.

The method here proposed considers a controlled experimental setting in which both single users and teams perform the same search task for a limited period of time. Sample size, number of members per team, as well as the task that participants will be asked to perform will depend on the particular characteristic of the study. With regard to the experimental design, at least two experimental conditions must be considered: (1) single users, whose data is used as a baseline for comparisons and also to create what we call artificial teams, and (2) collaborative teams. For the case of teams, regardless their size, it is recommended that in the recruitment process participants are asked to sign up with someone they have previous experience collaborating (e.g. friends, spouses, colleagues, etc.).

To test the effect of *real collaboration* compared to *pseudo-collaboration*, one could also form artificial teams that constitute the main baseline to properly determine whether or not real teams were able to achieve synergy. Such artificial teams are created by generating all possible unique combinations of the results of single users' products after completing the task. The size of artificial teams must be the same established for real teams. Once single users' results are effectively combined, the set of measures described in the following section are computed for both real teams and artificial teams. Finally, the results are compared using appropriate quantitative analyses.

3.2 Evaluation

In this section we present a number of traditional and non-traditional evaluation measures. Here we also describe other useful constructs and definitions that can be used to interpret and discuss the results obtained through these measures.

3.2.1 Universe of webpages

In order to compute quantities such as coverage, it is necessary a universal set of webpages. Given that the search domain in experimental settings can go from very small sets to the open Web, it is needed a more confined set that can be used to compare with. The universe of webpages, as we define it here, is the union of all the webpages visited by all of participants in a particular study.

$$U = \bigcup_t Coverage(t) \qquad \text{... (1)}$$

Here, *Coverage(t)* is the coverage (webpages visited) by every participant/team *t*.

3.2.2 Relevant webpages

This corresponds to the webpages that participants either bookmark or from where one or more fragments of information are collected. Once again, we took the union of all such webpages by each participant/team to form a universe of relevant webpages.

$$U_r = \bigcup_t RelevantCoverage(t) \qquad \text{... (2)}$$

Here, *RelevantCoverage(t)* is the set of webpages that participant/team *t* visited and found as relevant.

3.2.3 Precision, recall, and F-measure

Two of the most common evaluation measures in IR are precision and recall, which for our purpose here, are defined as the following.

$$Recall(t) = \frac{RelevantCoverage(t)}{U_r} \qquad \text{... (4)}$$

$$Precision(t) = \frac{RelevantCoverage(t)}{Coverage(t)} \qquad \text{... (3)}$$

To combine precision and recall into one measure of effectiveness, we use the traditional formulation of F-measure as defined below.

$$F = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \qquad \text{... (5)}$$

3.2.4 Coverage

We define coverage of a given team/participant as the total number of distinct webpages visited within the universe of webpages.

$$Coverage(t) = \{wp_i : wp_i \text{ was visited by } t \wedge wp_i \in U\} \qquad \text{... (6)}$$

We also considered a particular region of the coverage of teams/participants that is unique within the universe. We call such region unique coverage, which consists of all webpages within the coverage of a given team/participant *t* that were visited only by *t*.

$$UniqueCoverage(t) = Coverage(t) \setminus \bigcup_{t_i \in (T \setminus \{t\})} Coverage(t_i) \quad \text{...(7)}$$

3.2.5 Relevant coverage

We define relevant coverage as the region of coverage of a given team/participant *t* that intersects with the universe of relevant webpages.

$$\mathrm{RelevantCoverage}(t) = \mathrm{Coverage(t)} \cap \mathrm{U_r} \qquad \text{... (8)}$$

In a similar way, we call unique relevant coverage to the set of webpages within the unique coverage of a given team/participant *t* that intersects with the universe of all relevant webpages.

$$\mathrm{UniqueRelevantCoverage}(t) = \mathrm{U}\mathit{nique}\mathrm{Coverage(t)} \cap \mathrm{U_r} \qquad \text{... (9)}$$

3.2.6 Useful webpages

As part of the evaluation measures, we also consider an implicit measure based on the dwell time on a webpage as described in [5] and [25]. As reported in these prior works, we considered a webpage to be useful if a team/participant *t* spends at least 30 seconds on it. Note that we only consider content pages, discounting any search engine homepage or search engine results pages (SERPs).

3.2.7 Likelihood of discovery

To evaluate effectiveness of a team/participant *t* in discovering hard to find information, we devised a new measure called *likelihood of discovery*. We assume that webpages with a high likelihood are easier to find and are common among the majority of the users. On the other hand, those webpages with a low likelihood are difficult to reach and probably beyond the first results page of search engines. A team/participant *t* finding these webpages is being more effective in discovering information that is not just relevant, but also diverse.

In order to operationalize this idea, we use a formulation similar to that of inverse document frequency (IDF). Using the frequency of each webpage in the universal set, we compute its likelihood to be visited; in addition, each webpage's likelihood is multiplied by -1 in order to denote the IDF. As a result, each webpage is assigned with a normalized value between -1 and 0. In this sense, those webpages with a value close to 0 are rare (and even unique) to be reached by teams/participants, while those close to -1 are more likely to be visited.

3.2.8 Query diversity

In addition to the webpages that teams/participants visit during a given task, it is also of interest studying how they approached the task in terms of the queries they issued to find information. The purpose of this measure is to understand how similar or different are the queries formulated by participants during their search process.

In order to evaluate query diversity, one possible measure is the Lavenshtein distance between pairs of queries for each team/participant. Based on the results of this computation, for a given pair of queries; the closer the distance to 0 the higher the similarity between them. On the other hand, the higher the distance between queries, more different (therefore diverse) were the queries formulated within a team.

3.2.9 Cognitive Load

To study if collaboration has some negative implications for users in terms of cognitive load; participants are asked to respond a questionnaire after finishing their task. This questionnaire is a simplified version of NASA's Task Load indeX (TLX).[1] This instrument had the following questions.

1. *How mentally demanding was this task?*
2. *How physically demanding was this task?*
3. *How hurried or rushed was the pace of the task?*
4. *How hard did you have to work to accomplish your level of performance?*
5. *How insecure, discouraged, irritated, stressed, and annoyed were you?*

All these questions are responded using a 5-point Likert scale, where higher values in the responses indicate a more negative perception of the user with respect to of the areas considered in the above questions.

[1] Taken from http://www.cc.gatech.edu/classes/AY2005/cs7470 fall/papers/manual.pdf

3.2.10 Emotional Experience

Similar to the purpose of the questionnaire above, another source to study the positive or negative implications of working in collaboration is to ask participants how they feel. We suggest using PANAS [24] right before and right after teams/participants performed the task.

3.3 Experimental System

We developed *Coagmento*, a plugin for Firefox Web browser and a Web-based infrastructure that serves as an experimentation platform [7][17]. *Coagmento* is a flexible system that supports the search process of users either working individually or in collaboration with others (Figure 2). The system on the client side consists of a toolbar and a sidebar that provide users with a variety of tools for collecting, sharing, rating, and annotating information as a well as a communication system.

Coagmento is capable to keep track of browsing activity, users' actions, chat logs, and questionnaires, among other useful data. This log data is used later to compute the measures described in the above section.

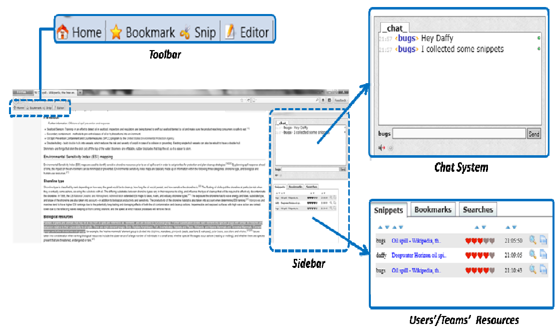

Figure 1: A snapshot of the experimental system with parts of it shown in details.

4. APPLICATION

We have successfully applied the method described above in a large user study involving 10 single users and 80 pairs working in eight collaborative conditions. From the 10 single users, we generated 245 artificial teams, whose results were compared to those from real teams. Results from a recently published study [20] with four conditions: single users, co-located at the same computer, co-located at different computers, and remotely located. The results revealed that two people collaborating at different computers is not the same as having the outcomes of two completely independent individuals combined. Indeed, the former do better in terms of discovering more and diverse information for an information-seeking task. Not only that, but the cognitive load in a real collaborative situation was found to be no more than what was perceived by those working individually. Thus, the synergic effect of the whole being greater than the sum of all was demonstrated and evaluated.

5. CONCLUSION

Collaboration is often a useful approach for solving a complex problem, but it has its costs and overheads [3]. One typically gets involved in collaboration if it has good benefit-to-cost ration, or if the given problem is too difficult to be solved without collaboration [18]. It is, however, difficult to measure if and how a collaborative endeavor would pay off. Traditional objective or quantitative approaches uses in IR are insufficient, and subjective or qualitative approaches may be expensive or difficult to employ.

In this paper we proposed and demonstrated a unique framework for evaluating various aspects of collaborative information seeking (CIS), especially the synergic effect.

We argue that the traditional measures of evaluating relevance are not appropriate for such situations, and proposed either modified or new kinds of evaluations. These included coverage, usefulness, likelihood of discovery, and query diversity. We believe this itself is an important contribution to the community, helping the researchers evaluate and design CIS systems and interfaces.

We have successfully applied the method in a large user study and so far results inform that pairs of users collaborating in an exploratory search task are able to find more and diverse information than simply combining the results generated by single users. These findings demonstrate, at least in a particular context, the synergic effect in CIS.

Data collected from this study have also served to study additional aspect that may help to explain the synergic effect in CIS. An example of such studies can be found in [6], in which the smile of users was explored as one potential contributing factor of the synergic effect.

We believe the method and the measures proposed and demonstrated here could help us further investigations of CIS with different setups, including asynchronous, non-time bound, multi-session, and non-dividable tasks, as well as collaborations that involve more than two participants.

6. REFERENCES

[1] Buckminster Fuller, R. (1975). SYNERGETICS Explorations in the Geometry of Thinking. http://www.rwgrayprojects.com/synergetics/synergetics.html. Retrieved January 15, 2011.

[2] Corning, P. A. (1995). Synergy and self-organization in the evolution of complex systems. *Systems Research 12*(2):89-121.

[3] Fidel, R., Mark Pejtersen, A., 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. doi: 10.1002/asi.20041.

[4] Foley, C. (2008). Division of Labour and Sharing of Knowledge for Synchronous Collaborative Information Retrieval. PhD Thesis. School of Computing, Dublin City University.

[5] Fox, S., Karnawat, K., Mydland, M., Dumais, S., White, T. (2005). Evaluating implicit measures to improve web search. ACM Transactions on Information Systems (TOIS). 23(2): 147-168.

[6] González-Ibáñez, R., Shah, C., & Córdova-Rubio, N. (2011). Smile! Studying Expressivity of Happiness as a Synergic Factor in Collaborative Information Seeking. *Proceedings of American Society of Information Science & Technology (ASIST) Annual Meeting*. New Orleans, Lousiana.

[7] González-Ibáñez, R., & Shah, C. (2011). Coagmento: A System for Supporting Collaborative Information Seeking. *Proceedings of American Society of Information Science & Technology (ASIST) Annual Meeting*. New Orleans, Lousiana.

[8] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 239-250). North Holland Press.

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

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

[11] Joho, H., Hannah, D., & Joemon, J. M. (2008). Comparing collaborative and independent search in a recall-oriented task. *Proceedings of Information Interaction in Context*. London, UK.

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

[13] Morris, M. R., & Horvitz, E. (2007). SearchTogether: An Interface for Collaborative Web Search. *ACM Symposium on User Interface Software and Technology (UIST)* (pp. 3-12). Newport, RI.

[14] Morris, M. R. (2008). A survey of collaborative web search practices. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. 2008:1657.

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

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

[17] Shah, C. (2010). *Coagmento* - A Collaborative Information Seeking, Synthesis and Sense-Making Framework. Integrated demo at *Computer Supported Cooperative Work (CSCW)*. Savannah, GA: February 6-11, 2010.

[18] Shah, C. (2010). Working in Collaboration - What, Why, and How? *Proceedings of Collaborative Information Retrieval workshop at CSCW 2010*. Savannah, GA.

[19] Shah, C., & Gonzalez-Ibanez, R. (2010). Exploring Information Seeking Processes in Collaborative Search Tasks. *Annual Meeting of the American Society for Information Science*. Pittsburgh, PA.

[20] Shah, C., & González- Ibáñez, R. (2011). Evaluating the Synergic Effect of Collaboration in Information Seeking. SIGIR '11: Proceedings of the 34th annual international ACM SIGIR conference on Research and development in information retrieval, July 24–28, Beijing, China: July 24-28.

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

[22] 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.

[23] Twidale, M. B., & Nichols, D. M. (1996). Collaborative browsing and visualisation of the search process. *Proceedings of Aslib* (Vol. 48, pp. 177-182). Retrieved from http://www.comp.lancs.ac.uk/computing/research/cseg/projects/ariadne/docs/elvira96.html.

[24] Watson, D., Clark, Lee A.; Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.

[25] White, R. W., & Huang, J. (2010). Assessing the scenic route: Measuring the value of search trails in web logs. *Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval (SIGIR)*. Geneva, Switzerland.