Collaborative Information Seeking Lab Experiments Dataset
Chirag Shah and Roberto Gonzalez-Ibanez
chirags@rutgers.edu, roberto.gonzalez.i@usach.cl

INTRODUCTION

This is a brief documentation for the dataset being shared from a set of lab experiments conducted by Chirag Shah and Roberto Gonzalez-Ibanez at Rutgers University in 2010-2011. A reasonable care has been taken to clean up the data and remove any identifying information about the participants, but anyone using this data should assume it to be their responsibility to do due diligence regarding this. Any publications and presentations resulting from the usage of this data must cite and/or acknowledge this data. It took months to design the study and the tools used for the experiments, and several more months to collect the data.

How to cite this data:

Shah, C., & Gonzalez-Ibanez, R. (2017). Collaborative Information Seeking Lab Experiments Dataset. Available from http://infoseeking.org/data.php#cis2010

Relevant publications to learn more about these experiments and cite:

- Shah, C., and Gonzalez-Ibanez, R. (2011). *Evaluating the synergic effect of collaboration in information seeking*. Proceedings of ACM SIGIR, pp. 913-922. Beijing, China.
- Shah, C., Gonzalez-Ibanez, R. (2012). Spatial context in collaborative information seeking. *Journal of Information Science (JIS). 38*(4), 333-349.
- Gonzalez-Ibanez, R., Haseki, M., & Shah, C. (2012). *Time and space in collaborative information seeking: The clash of effectiveness and uniqueness*. Proceedings of Association for Information Science & Technology (ASIST) Annual Meeting, Baltimore, MD. 10 pp. Available from http://onlinelibrary.wiley.com/doi/10.1002/meet.14504901080/epdf.
- Gonzalez-Ibanez, R., Shah, C., & Cardova, N. R. (2011). *Smile! Studying expressivity of happiness as a synergic factor in collaborative information seeking*. Proceedings of Association for Information Science & Technology (ASIST) Annual Meeting , New Orleans, LA, 10 pp. Available from http://onlinelibrary.wiley.com/doi/10.1002/meet.2011.14504801171/epdf.
- Gonzalez-Ibanez, R., Haseki, M., and Shah, C. (2013). Let's search together, but not too close! An analysis of communication and performance in collaborative information seeking. *Information Processing & Management, 49*(5), 1165-1179.

Format

The data is available as an SQL dump. One should be able to load it up easily using 'source' command on an SQL prompt or import function available in most GUI-based SQL clients. There are five tables in this database:

- users – contains information about user ID, team ID, and the condition (1 to 8) in which a given team was randomly assigned to (see Table 1).
- actions – contains log data of various actions the users performed.

- pages – table containing page visits. If a page was a SERP, the query is also extracted. If the page was saved, 'bookmark' is set to 1.
- queries – table containing queries.
- snippets – table containing text snippets the participants collected.

Tools and instruments

The primary tool used for collecting this data was Coagmento, which is available as a free, open-source downloadable tool from http://coagmento.org. Information about other tools (e.g., emotion extraction) and instruments (e.g., questionnaires) can be obtained by contacting the authors.

METHOD

We conducted a laboratory study involving a total of 160 participants in 80 teams. This section describes the study procedure, the subjects, the system, the task, and the experimental conditions.

Participants

Participants in this study were asked to sign up in pairs with someone with whom they had previous experience collaborating. In addition, they were informed of their compensation for participating in the study, which consisted of $10 per person and the possibility to obtain additional prizes if they were among the six best performing teams (two teams each for $50, $25, and $15 per person additionally) at the end of the study. All of these were students from Rutgers University, recruited and randomly assigned to eight experimental scenarios (Table 1).

System

We developed a plugin for the Firefox web browser, called Coagmento,[1] which provided appropriate tools and support for the participants working in various conditions. A screenshot of this plugin within Firefox is shown in Figure 1. As shown, the plugin included a toolbar and a sidebar. The toolbar had the following buttons: (1) Home – for taking the participant to appropriate questionnaires, (2) Bookmark – for bookmarking a webpage, (3) Snip – for collecting a snippet using highlighted text from a webpage, and (4) Editor – for accessing a shared editor for writing the report (Figure 2).

The sidebar had two major components: a chat-box and a resources panel. The chat-box allowed the collaborators in a given team to communicate with each other. The researcher conducting the study also used it to provide instructions to the participants. The resources panel included tabs for bookmarks, saved snippets, and executed queries.
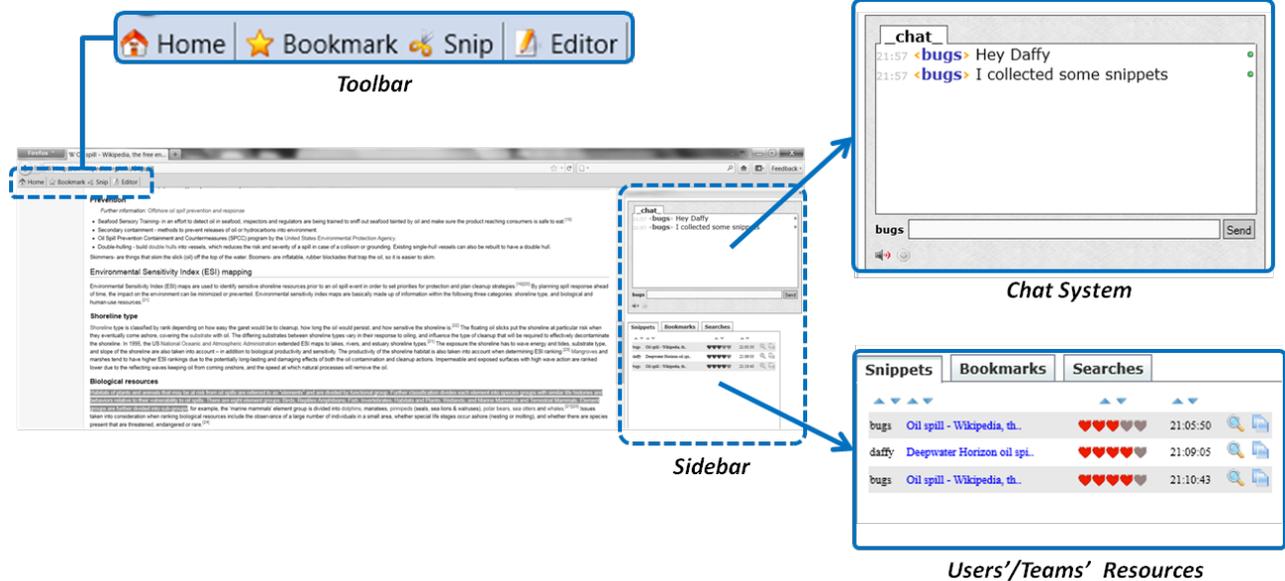
---

[1] Publicly available from http://coagmento.org.

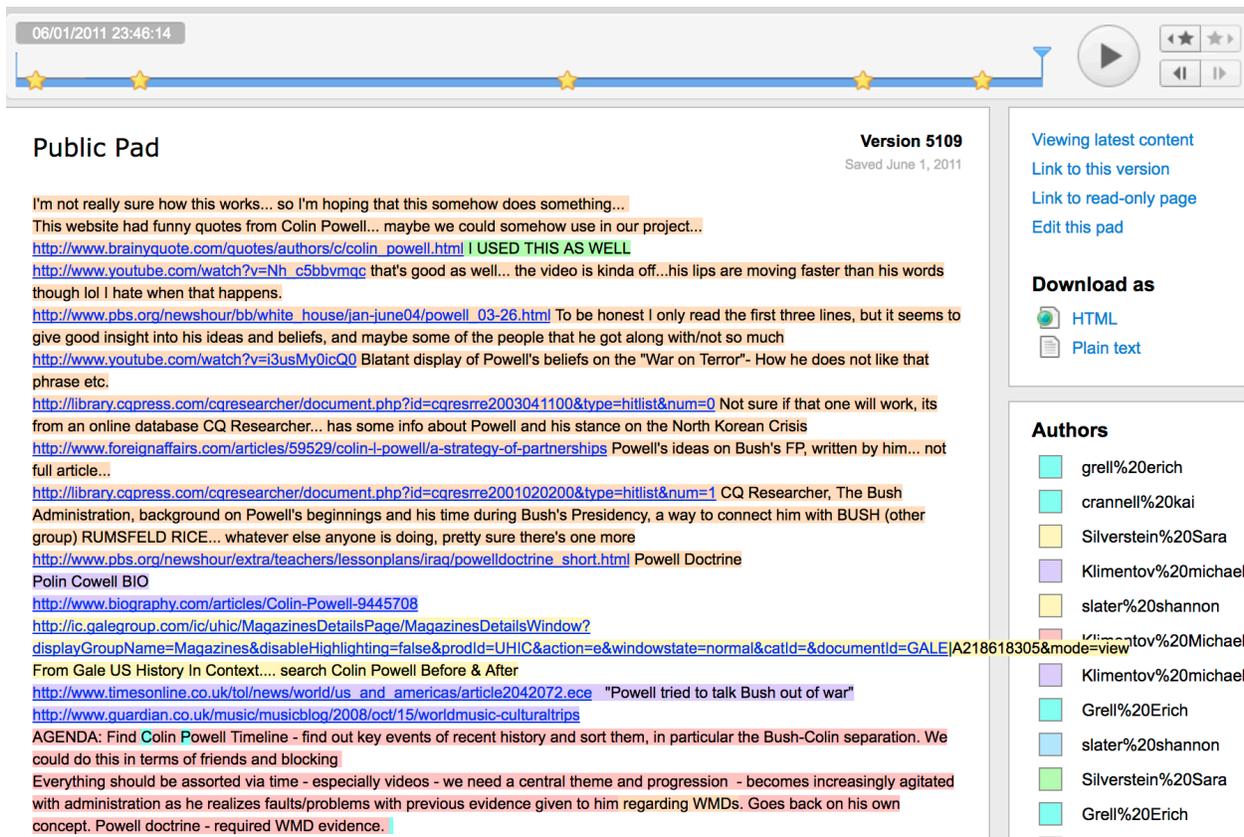**Figure 1: A snapshot of the experimental system with parts of it shown in details.**

Figure 2: Coagmento's collaborative editor.

Session workflow

Each experimental session lasted less than an hour and was structured in six parts as depicted in Figure 3 and described below.

1. Participants were introduced to the study and asked to sign a consent form.
2. Participants watched a brief tutorial in order to learn the basic functionalities required during the task.
3. Participants individually filled out a set of pre-task questionnaires. In the case of two participants working at the same computer (condition later explained), the participants were separated for this phase.
4. Participants read the task description (given later).
5. Each participant/team worked for approximately 25 minutes on the given task that included searching for relevant information, and using it to compose a report.
6. Participants filled out post-task questionnaires.

The researcher conducting the study communicated with the participants through the chat-box at different times during the study instructing them to start/stop the task or fill in a questionnaire.
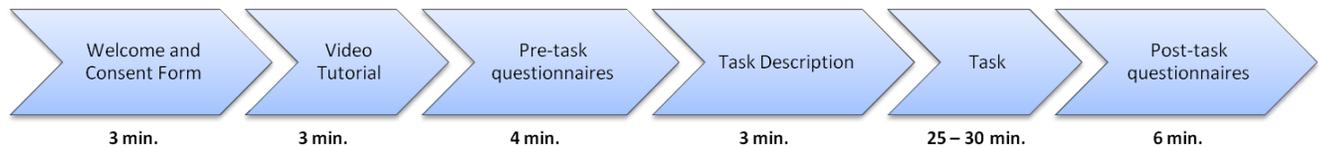
Welcome and Consent Form	Video Tutorial	Pre-task questionnaires	Task Description	Task	Post-task questionnaires
3 min.	3 min.	4 min.	3 min.	25 – 30 min.	6 min.

Figure 3: Study session workflow.

Conditions

Various conditions (C1 to C8) for these experiments were defined based on (1) space (C1, C2, C3); (2) time (C1-C7 for synchronous, C8 for asynchronous); (3) mode of communication (C1-C5); (4) tools available for information synthesis (C7 vs. rest of the conditions); and (5) availability of emotional feedback (C3 vs. C6).

All but one condition used the collaborative editor shown in Figure 2 for synthesizing information (writing a report). For C7, the participants were set up with Gmail accounts that they used for sending their pieces or drafts of the reports back-and-forth.

We developed a tool for extracting facial emotions using a webcam in real time. This information was used in C6. In other words, in C6, a participant was given a real-time emotional status from his/her collaborator.

For the asynchronous condition (C8), the collaborators were invited to the lab, and after explaining the task, they were split up. One of them worked on the task for about 30 minutes, and then the other one took over for another 30 minutes.

A summary of all these conditions can be found in Table 1. Please refer to the publications listed earlier for more details about the conditions and the experimental setups.

Table 1: Experimental conditions.

Cond.	Description	Communication	Information synthesis tool
C1	Co-located using the same computer	In person	Collaborative editor
C2	Co-located using different computers	In person	Collaborative editor
C3	Remotely located	Text chat	Collaborative editor
C4	Remotely located	Text chat + audio	Collaborative editor
C5	Remotely located	Text chat + video	Collaborative editor
C6	Remotely located	Text chat + emotional feedback	Collaborative editor
C7	Remotely located	Text chat	Email
C8	Asynchronous	In person (at the beginning and during the handoff)	Collaborative editor

Task

We chose "gulf oil spill" as the topic for this experimentation since it was quite popular and relevant at the time the study was being conducted. Our preliminary investigations, including a few pilot runs, indicated that there was a huge amount of material on this topic, and that the participants would find it interesting and challenging enough as an exploratory search task. Each participant was given the following task description.

"A leading newspaper has hired your team to create a comprehensive report on the causes, effects, and consequences of the recent gulf oil spill. As a part of your contract, you are required to collect all the relevant information from any available online sources that you can find.

To prepare this report, search and visit any website that you want and look for specific aspects as given in the guideline below. As you find useful information, highlight and save relevant snippets. Make sure you also rate a snippet to help you in ranking them based on their quality and usefulness. Later, you can use these snippets to compile your report, no longer than 200 lines, as instructed.

Your report on this topic should address the following issues: description of how the oil spill took place, reactions by BP as well as various government and other agencies, impact on economy and life (people and animals) in the gulf, attempts to fix the leaking well and to clean the waters, long-term implications and lessons learned."

The participants saw this description on the screen (part 4 in the study), and were also given a printed copy to refer to during their session.