On a Tip of Your Search: Evaluating Effect of Search Tips for Complex Informational Search Tasks

Denis Savenkov
Emory University
dsavenk@emory.edu

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

ABSTRACT

Search engine is a ubiquitous tool used by millions of people on a daily basis. However, as with every tool, certain skills are required in order to use it efficiently. Unfortunately users have different experience and not everybody is able to find answers to all questions she is interested in [find a paper to cite here]. Helping users to develop their search skills was included as one of the key research directions by [2]. However, the assistance offered by the modern search engines are limited to query suggestion and spelling correction [something else?]. A number of researches are available that study different ways of helping users be more successful with their searches [cite some reviews, or a couple of different papers]. One of such ways is search tips, designed to improve user experience and search success [3]. However, previous research focused on tips suggesting different functionality on search engines that better suit a particular problem. In this work we study the effect of displaying strategic search tips, which suggest a way of solving a difficult search problem by splitting it into pieces and searching for parts of the question. We focus on 2 types of tips: specific hints, tailored to a problem user is currently solving and generic hints, describing the general strategy of splitting problem into pieces. The results show that optimally designed specific tips can be implemented effieciently by the users and improve search success rate, however generic tips might be distracting and detrimental to user experience.

Categories and Subject Descriptors

H.3.3 [**Information storage and retrieval**]: Information Search and Retrieval—*query formulation, search process*

General Terms

Measurement, Design, Experimentation, Human Factors

Keywords

User studies, search interface, experimental design, effectiveness measures, query reformulation, expertise, tactics, tips, suggestions, assistance, efficiency.

1. INTRODUCTION

Solving a search problem involves 2 parties: a system that answers queries and a user who interacts with the system. Asking the right questions and interacting with the system in a right way is as important (or even more important) than the quality of system answers [can we support this with some citation]. The importance of user interactions is even more important for difficult queries which are still not answered correctly by the modern search engines [cite robustness paper].

The importance of helping users to interact with the system is well understood and various researches studies different ways of user assistance [a list of citations goes here]. But unfortunately, modern search engines are still limited in providing user feedback and suggestions. It is a standard to offer query syntax correction to help fixing typos and misspellings, query suggestions to guide users to "better" queries, rich search results representations to help users choose which results to click on or to give an answer Is there something else?.

However, these tools are limited in assisting users to solve difficult informational search task. The ability to formulate good search queries and use search tools as efficiently as possible becomes the crucial factor towards the overall success. People have different skills and abilities and the effect of these variables to successful search is sometimes more important than retrieval performance [need to support this with something].

[Need some transition to why we study these particular type of tips for these particular type of search problems. Possibly link to agoogleaday as example of difficult search problems].

In this paper we study the effect of strategic search hints, designed to guide user in splitting the original problem into pieces and combining results for subproblems together. More specifically we manually designed 2 type of tips: task-specific, which present a way of correctly solving the questions and generic, which explains "divide and conquer" strategy, i.e. splitting the original question into pieces and combining answers to these pieces together. The user study was conducted in a form of a search game and it reveiled that users were able to follow carefully designed task-specific hints and this lead to increase in success rate. However, generic tips were probably distracting to the user and the success rate dropped as compared to not seeing any search tips. This

suggests that [what?].

2. RELATED WORK

Based on related work review from Daniel Russel [3].

Search skills can be trained [3]. For example, Google offers courses[1] designed to improve ones efficiency in solving difficult search tasks and interacting with a modern search engine.

Summarize works on task level query suggestion Experiment Design State main difference from Dan Russel paper: we focus on informational tasks which are solved with web search and hints are less likely to be 100% helpful

Query terms selection

Other dialog systems?

Showing predicted retrieval performance

3. TIPS FOR DIFFICULT SEARCH TASKS

To estimate the effect of search tips on user behavior and success we conducted a user study using Amazon Mechanical Turk platform[2]. The motivation to find the correct answer is very important for the study, thus we decided to pose it as a web search game, similar to "a Google a Day"[3] and uFindIt [1].

3.1 Web Search Game

The web search game used for the study offers users to find answers to several questions using the provided web search interface. Figure 1 shows the interface of the game. At the beginning of the game users were instructed to use only the search interface provided. Answers should actually be found when using web search. In a rare occurrence that user might know the answer to a question she is instructed to ignore the prior knowledge and find the answer anyway. Since tasks might be difficult a chance to skip a question was provided, although users were instructed that effort put into solving a question will be evaluated.

In the previous experiments we've noticed that with difficult search tasks Mechanical Turk workers tend to skip the answer validation phase and submit the first hypothesis they found, although even shallow analysis reveals that it is incorrect. To motivate deeper analysis of the candidates the final version of the game featured automatic answer verification. Each answer submitted by a user was checked for the presence of the required keyword. If the answer was incorrect a dialog popped up and a player could continue search.

The game search interface was based on API of one of the major web search engines. All search results were cached so that users asking the same query get the same results. Moreover, all links to web pages were rewritten to use our caching HTTP proxy.

At the end of the game a questionnaire was presented asking for feedback on user satisfaction with the game, prior experience and other comments.

3.2 Search Tasks Description

The tasks for the study were borrowed from the a Google a Day questions archive. Unfortunately, a lot of web pages exists that discuss solutions to these questions. Thus we had to filter search results excluding all pages that mentions the major part of the search questions or phrase "a google a day". To keep users focused throughout the whole game we decided to limit the number of questions to 4. Table 1 describes all 4 tasks with the possible solutions. These tasks are quite difficult and usually involved more than one query to solve them with the game search interface.

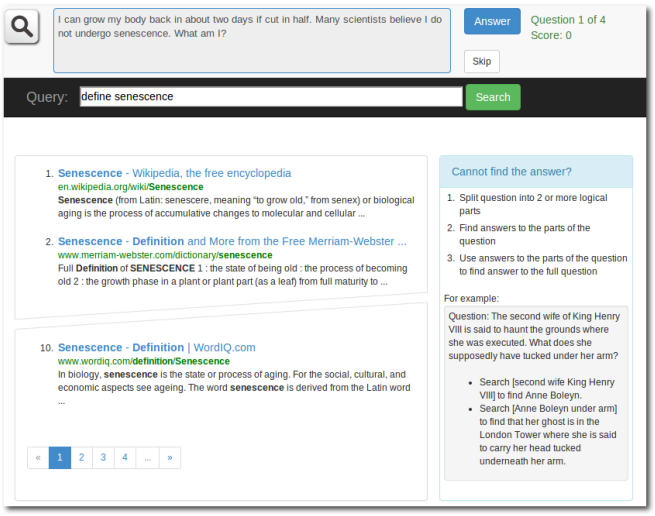

Figure 1: The interface of the search game used in the study

3.3 Search Tips

The tasks used for the game are examples of complex informational search problems and usually require several searches. Questions have multiple parts and to solve them it is helpful to search for answers to parts of the questions and then combine them.

We studies 2 types of search tips: task specific and generic. Task specific hints were constructed from one of the possible solutions to the question and described one way to search and find the answer. Generic hint described the general strategy that can be applied to many difficult informational search task. The actual tips shown to the players are described below.

Generic hint was the same for all tasks and looked the following way:

1. Split question into 2 or more logical parts
2. Find answers to the parts of the question
3. Use answers to the parts of the question to find answer to the full question

For example:
Question: The second wife of King Henry VIII is said to haunt the grounds where she was executed. What does she supposedly have tucked under her arm?

1. Search [second wife King Henry VIII] to find Anne Boleyn.
2. Search [Anne Boleyn under arm] to find that her ghost is in the London Tower where she is said to carry her head tucked underneath her arm.

[1]http://www.powersearchingwithgoogle.com
[2]http://www.mturk.com/
[3]http://www.agoogleaday.com/

Table 1: Search tasks used for the study

Task ID	Task Text	Possible
Task 1 (“hydra”)	I can grow my body back in about two days if cut in half. Many scientists believe I do not undergo senescence. What am I?	Search [regenerative animal]. This will yield a number of possibilities including starfish, salamanders and several others. Search [senescence]. Find out that it means “biological aging”. Search [animal biologically immortal], and learn that the hydra supposedly fulfill both of the above criteria.
Task 2 (“quirinus”)	Of the Romans ”group of three” gods in the Archaic Triad, which one did not have a Greek counterpart?	Search for [Archaic Triad] to find Jupiter, Mars and Quirinus. Among those Quirinus didn’t have a Greek counterpart.
Task 3 (“dinosaur”)	As George surveyed the “waterless place”, he unearthed some very important eggs of what animal?	Search for [waterless place] to find out that it is the translation of the Mongolian word ”Gobi” or “Gobi Desert”. George Olsen unearthing the first whole dinosaur eggs in 1923.
Task 4 (“cherokee”)	If you were in the basin of the Somme River at summers end in 1918, what language would you have had to speak to understand coded British communications?	Search [somme river basin 1918]. Find out that’s when the Second Battle of the Somme (a WWI battle) took place. Searching [second battle of somme code language] reveals the Cherokee served as code talkers in battle. They relayed messages in the Cherokee language that Germans couldn’t decipher.

Specific hints were designed for each question separately and presented a way to solve the problem by searching for a specific parts of the question. Specific tip for Task 1:

1. Find what is senescence
2. Find who do not undergo senescence
3. Find animals who can regenerate body and choose the one that satisfy both conditions

Tip for Task 2:

1. Find the names of the gods from the Archaic triad
2. For each of the gods find a Greek counterpart

Tip for Task 3:

1. Find what is the “waterless place” mentioned in the question?
2. Search for important eggs discovery in this “waterless place”

To estimate the effect of search tips on users behavior and success we split users into 3 groups: users who were shown no hints; users who were shown task specific hints and users who were shown generic hint for all tasks. To avoid the learning effect demonstrated in [3] we assign a user to a single group only.

4. RESULTS

Provide raw results of the experiments, e.g. how many participants, how many HITs accepted, rejected, problems. Show examples of search trails for successful and unsuccessful searches.

From 199 participants, who accepted the HIT on Amazon Mechanical Turk[4], only 169 moved further than the rules of the game and got to the first questions. The search tasks were difficult and some players decided to quit the game, thus only 90 players finished the game, from those there were 9 submissions which we filtered out from the future analysis. The only 2 reasons for that were: lack of effort, e.g. some players skipped several tasks after only a single query; some other submissions indicated usage of external resources such as outside of the game search engine, e.g. the only query asked was the correct answer to the task. From 81 submissions 10 players indicated in the survey that they didn’t see the tips which were shown to them, so we further filtered those submissions and finally we had 71 completed games, which split into the groups of 29, 20, 22 for players who didn’t have tips, who had task-specific hints and who had generic tips correspondingly.

[4]http://mturk.com/

4.1 Analysis

Give a table and a couple of pictures with main quantitative results

Figure 2 shows plots of the correct answers rate per task for groups of users who saw no hint, specific task-oriented hint or generic hint. As we can see, success rate is higher for users who saw specific hint than no hint at all. Somewhat surprising is the fact, that users who saw generic search hint were slightly less successful. The difference is more significant in more difficult tasks 1 and 4.

The main findings can be summarized as follows:

- “Correct” search tips allows users to find correct answer more often and do this faster than without search tips
- “General” search hints can have detrimental effect on search success, reducing the success rate and increasing the time spent on task

5. DISCUSSIONS AND FUTURE WORK

Make a conclusion by summarizing the findings one more time

Speculate on the negative effect of general search hints - distracting? hard to follow?, less satisfaction when tips were shown - self satisfaction?.

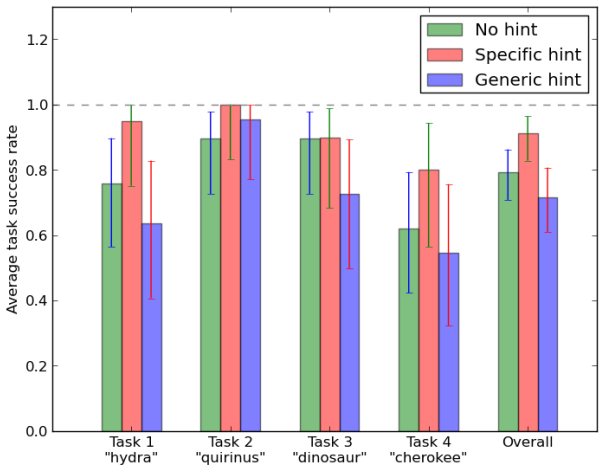

Figure 2: Success rate per task for each group of participants

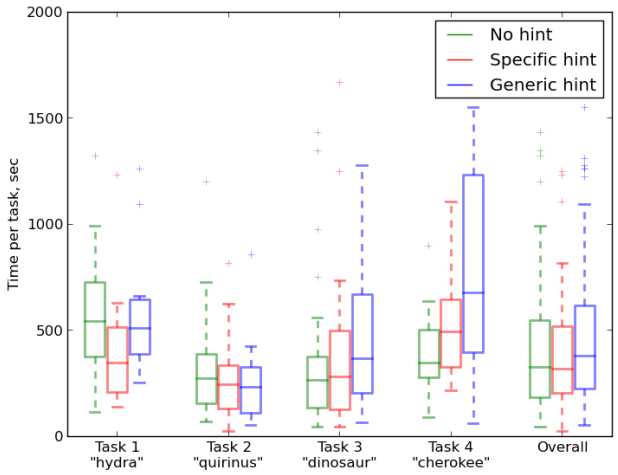

Figure 3: Task completion time for each group of players

6. ACKNOWLEDGMENTS

Thanks to Dan Russel for sharing questions.

7. REFERENCES

[1] M. Ageev, Q. Guo, D. Lagun, and E. Agichtein. Find it if you can: A game for modeling different types of web search success using interaction data. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 345–354, New York, NY, USA, 2011. ACM.

[2] J. Allan, B. Croft, A. Moffat, and M. Sanderson. Frontiers, challenges, and opportunities for information retrieval: Report from swirl 2012 the second strategic workshop on information retrieval in lorne. *SIGIR Forum*, 46(1):2–32, May 2012.

[3] N. Moraveji, D. Russell, J. Bien, and D. Mease. Measuring improvement in user search performance resulting from optimal search tips. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 355–364, New York, NY, USA, 2011. ACM.