On a Tip of Your Search: Evaluating Effect of Strategic Search Tips on User Success in Complex Informational Search Tasks

Denis Savenkov
Emory University
dsavenk@emory.edu

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

ABSTRACT

According to [13] the majority of cases when users require assistance comes from the query formulation and refinement stages. Query suggestion is an efficient technique that helps users find a popular query that is similar to theirs and potentially better. However, currently it works better for relatively popular topics and there are still many difficult tasks that search engines do not handle perfectly and users do not get any assistance and become frustrated [5]. Displaying search tips is another alternative for automatic user assistance and can have better coverage than query suggestion. ([11]) studied the effect of strategic tips that suggest users some features of search engines that might be optimal for a particular task. However, the tasks used in the study were specifically designed to benefit from some search features available and such tactical features tips might not help with solving difficult information questions. 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, searching for parts of the question and combining these answers together. We focus on 2 types of tips: task-specific tips, tailored to a particular search problem and generic tips, describing the general strategy of solving a search task. In the user study conducted using a web search game users were able to follow optimally designed task-specific tips and achieve higher success rate than users who didn't get such assistance. However, generic tips were shown to be rather distracting and detrimental.

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

SIGIR '14 Gold Coast, Australia

Keywords

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

1. INTRODUCTION

Search engines are ubiquitous tools used by millions of people on a daily basis. But certain skills are required in order to use it efficiently. 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 study of [13] revealed that most of the times users have problems formulating and refining queries. Modern search engines provide users with some automatic assistance during their searches and arguably one of the most successful technique is query suggestion. Although it is very effective for relatively popular topics, there are still many tasks that search engines do not handle perfectly and users do not get any assistance and become frustrated [5].

Search skills can be trained. For example, Google offers a course[1] focused on improving search efficiency. Although useful such courses have very limited coverage of users willing to spend their time learning how to use web search efficiently. Displaying tips on a search screen could be another technique that has some learning effect and offers immediate assistance to the user in solving her current search task. The user study from [11] demonstrated that tactical search, suggesting to use a certain feature of a search engine, helps users find answers more quickly and the effect is retained after a week with tips removed.

Besides the awareness about search tools available, general search strategies are extremely important when dealing with difficult search task. In this paper we focus on strategic search tips, that are designed to guide a user in solving her search problem. More specifically we conducted a user study to evaluate the effect of tips proposing a divide-and-conquer strategy for search. Search tasks used in the study can be solved by splitting the original problem into pieces, finding answers to subproblems and combining these answers together to solve the original question. We manually designed 2 sets of tips: task-specific, which present a way of correctly solving the questions and generic, which states the strategy and leaves the details to the user. Should I say this here? Or how to finish this section?. The results of the work

[1]http://www.powersearchingwithgoogle.com

demonstrate that users are able to follow carefully designed task-specific tips and this increases the task success rate. However, generic tips were probably too hard to follow and caused the success rate to drop.

2. RELATED WORK

There has been considerable amount of work on search assistance and improving user experience with feedback, suggestions and hints. Interactive information retrieval and human-computer information retrieval [10] focuses on interactions between users and search systems. Graphical techniques can be used to visually represent large-scale collections of information and help searchers in their tasks [4].

Results of the study in [13], which focused on identification of different categories of help-seeking situations, demonstrates that in 41.5% of the cases users were seeking for help to refine their searches followed by inability to construct search statements in 18% of the cases, which confirms findings of [6]. Individual terms ([12]) and queries suggestion ([7],[2],[3]) are among the most popular techniques for helping users to augment their queries. The study from [8] demonstrated that users prefer query suggestions over term relevance feedback and that good manually designed suggestions improve retrieval performance. Query suggestion methods usually provide previous queries that are similar to the query of interest and work better for popular information needs [2].

When query or term suggestions are not efficient, it is still possible to help users by providing potentially useful search hints. An adaptive tool providing tactical search suggestions was presented in [9] and users reported overall satisfaction with its automatic non-intrusive advices. Modern search engines have many features that are not typically used by a average user, but can be very useful in particular situations as shown in [11]. The study demonstrated the potential effectiveness of tactical search feature tips and their teaching effect.

The major differences of this work from [11] is the type of search tips used. Rather than suggesting users the available search features to use, this work focuses on strategic search tips, designed to help users solve difficult informational questions. Many informational questions cannot be answered by a single web query and require splitting the task into pieces and combining partial answers into new searches. From our studies we noticed that users do not actively use this tactic and usually keep trying to reformulate their query expecting to find the query that will give them the correct result. However, in these tasks it is hard to come up with a good hint which will help all the users. In this study we are evaluating the effect that task-specific manually designed strategic tips as well as generic task-independent tips might have on user experience and success rate.

3. USER STUDY DESCRIPTION

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 the task as a web search game similar to a Google a Day[3] and uFindIt [1].

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

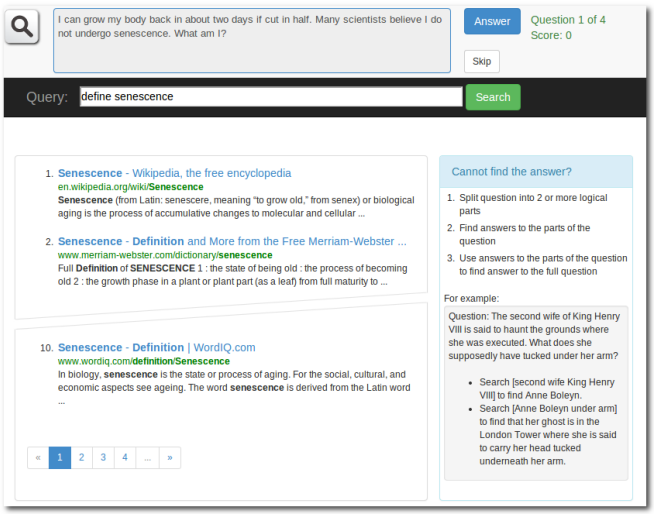


Figure 1: The interface of the search game used in the study

3.1 Web Search Game

The web search game used for the study asks 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 are instructed to use only the search interface provided. Answers should actually be found using web search. In a rare occurrence that a user might know the answer to a question she is instructed to ignore the prior knowledge and use the search anyway. Since tasks might be too 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 validation phase [1] and submit the first possible answer found when even shallow analysis reveals that it is incorrect. To overcome this problem the submitted answers were automatically checked for correctness (a 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 discussing solutions to these questions exist. So we had to filter search results and exclude all pages that mention a major part of the search question or "a google a day" phrase. To keep users focused throughout the whole game we decided to limit the number of questions to 4. Table 1 describes all 4 tasks.

3.3 Search Tips

The tasks used for the game are examples of complex

Table 1: Search tasks used for the study

Task ID	Task Text	Answer
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?	Senescence means “biological aging”. Hydra is considered biologically immortal and regenerates fast.
Task 2 (“quirinus”)	Of the Romans ”group of three” gods in the Archaic Triad, which one did not have a Greek counterpart?	Archaic Triad includes 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?	Waterless place is the translation of the Mongolian word ”Gobi” or “Gobi Desert”. George Olsen found 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?	Cherokee served as code talkers in the Second Battle of the Somme.

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.

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 avoid the learning effect demonstrated in [11] we split users into 3 groups: users who were shown no tips; users who were shown task-specific tips and users who were shown generic tip for all tasks. This tips for were displayed all the time in the panel to the right of search results as shown on Figure 1.

4. RESULTS

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. What is missing?

4.1 Analysis

The main characteristic of interest was the search success rate, measured in the number of questions answered correctly by users in different groups. Figure 2 plots the fraction of correct answers for players from different groups. As we can see, success rate tend to be higher for users who saw task-specific hint compared to users who didn’t see any hint. Somewhat surprising is the fact, that users who saw generic tip were slightly less successful. The difference is more significant in more difficult tasks 1 and 4. This can probably be explained by the fact that generic tip was more difficult to implement and unsuccessful attempts lead to frustration Sound reasonable?.

Similar to [11] we looked at the average time to answer the question (we removed games where a user didn’t find the answer and skipped the task), the plot is provided on Figure 3. Here we don’t see a clear pattern, for task 1 partic-

[4]http://mturk.com/

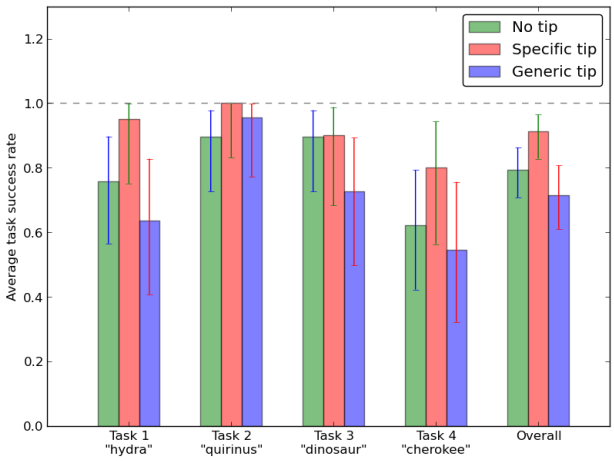


Figure 2: Success rate per task for each group of participants

ipants who saw task-specific tip were able to find the answer faster, however this is not the case for task 4. A possible explanation is that the provided tips were not the optimal ones and the problem allowed a faster solution, which was found by users who didn't see the tips Can I find an example of a search trail?]. It is worth noting, that users from the generic search tip group had slightly higher variance in success time, which is probably explained by the fact that some users were successful in finding the right way to follow the tip and some other users struggled with it more Reasonable?.

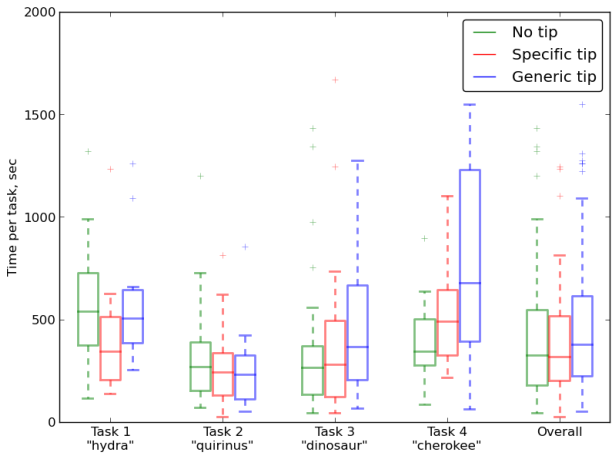


Figure 3: Task completion time for each group of players

Another interesting insight comes from the number of incorrect attempts users made. Figure 4 demonstrates the average number of incorrect submissions for all groups of users. Although the variance is high, but there is a tendency, that users who saw task-specific tips made less submission attempts and users who saw generic tip were incorrect slightly more often than users who didn't get any tips during the game. This is not in direct correspondence with time spent on the game. The effect can probably be explained by the difference in frustration in each of the user groups. Users who saw a clear strategy to solve the question were less likely to notice plausible, but incorrect solution. Moreover, by analyzing texts of incorrect answers we can conclude that users sometimes tried to guess the correct answer by typing everything that came in mind.

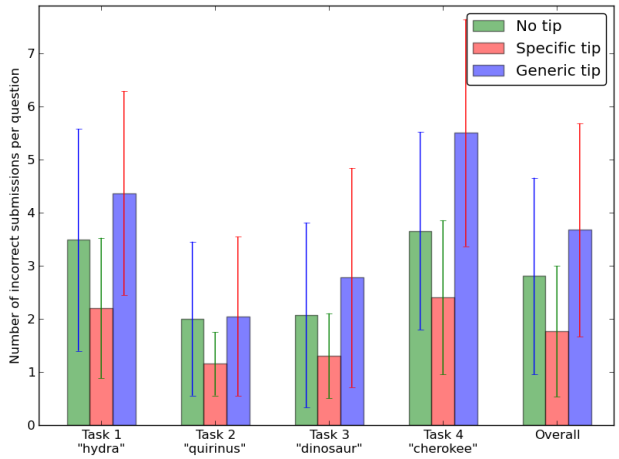


Figure 4: The number of incorrect submission attempts per question for all groups of users

Finally, we looked at the surveys left by each group of users. Figure 5 presents proportions of different answers to three of the questions: "How did you like the game?", "How difficult was the game?" and "Were search tips useful to you?". Surprisingly, results for the first question were lower for users who saw tips during the game and users who didn't saw tips liked the game more. It could be explained by the self-achievement effect, that is lower when users get help even if it wasn't helpful. The answers to the question about game difficulty are in agreement with the success rate: users who saw task-specific tips rates game as being easier than participants who struggled more to get the correct answers. The Figure 5c shows that users actually found our hints useful for their searches and indicated this in the final survey.

5. CONCLUSION

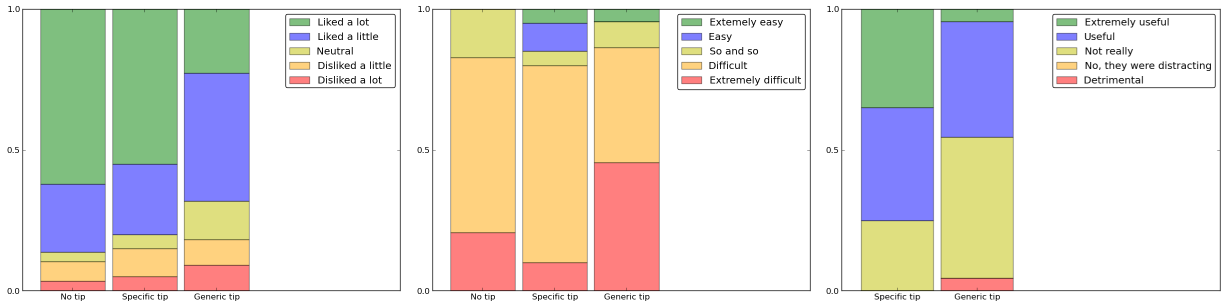
The results of the user study described in this work demonstrated the potential of good strategic search tips and their effect on search success. However, search tips that are too general and hard to implement could be detrimental. Users in our study were less successful when they saw generic strategic search tip.

6. ACKNOWLEDGMENTS

The authors would like to thank Daniel Russel for providing an archive of questions from "a Google a Day" search game.

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.



(a) How did you like the game? (b) How difficult was the game? (c) Were search tips useful to you?

Figure 5: Proportions of replies to some of the survey question for each group of users

[2] S. Bhatia, D. Majumdar, and P. Mitra. Query suggestions in the absence of query logs. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 795–804, New York, NY, USA, 2011. ACM.

[3] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 875–883, New York, NY, USA, 2008. ACM.

[4] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[5] H. A. Feild, J. Allan, and R. Jones. Predicting searcher frustration. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 34–41, New York, NY, USA, 2010. ACM.

[6] C. HÃűlscher and G. Strube. Web search behavior of internet experts and newbies. *Computer Networks*, 33(1âĂŞ6):337 – 346, 2000.

[7] R. Jones, B. Rey, O. Madani, and W. Greiner. Generating query substitutions. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, pages 387–396, New York, NY, USA, 2006. ACM.

[8] D. Kelly, K. Gyllstrom, and E. W. Bailey. A comparison of query and term suggestion features for interactive searching. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, pages 371–378, New York, NY, USA, 2009. ACM.

[9] S. Kriewel and N. Fuhr. Adaptive search suggestions for digital libraries. In D.-L. Goh, T. Cao, I. Solvberg, and E. Rasmussen, editors, *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers*, volume 4822 of *Lecture Notes in Computer Science*, pages 220–229. Springer Berlin Heidelberg, 2007.

[10] G. Marchionini. Toward human-computer information retrieval. *Bulletin of the American Society for Information Science and Technology*, 32(5):20–22, 2006.

[11] 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.

[12] I. Ruthven and M. Lalmas. A survey on the use of relevance feedback for information access systems. *The Knowledge Engineering Review*, 18(02):95–145, 2003.

[13] I. Xie and C. Cool. Understanding help seeking within the context of searching digital libraries. *Journal of the American Society for Information Science and Technology*, 60(3):477–494, 2009.