To Hint or Not: Exploring the Effectiveness of Search Hints for Complex Informational Tasks

Denis Savenkov
Emory University
dsavenk@emory.edu

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

ABSTRACT

Extensive previous research has shown that searchers often require assistance with query formulation and refinement. Yet, it is not clear what kind of assistance is most useful, and how effective it is both objectively (e.g., in terms of task success) and subjectively (e.g., in terms of searcher perception of the search difficulty). This work describes the results of a controlled user study comparing the effects of providing specific vs. generic search hints on search success and satisfaction. Our results indicate that specific search hints tend to effectively improve searcher success rates and reduce perceived effort, while generic ones can be detrimental in both search effectiveness and user satisfaction. The results of this study are an important step towards the design of future search systems that could effectively assist and guide the user in accomplishing complex search tasks.

Categories and Subject Descriptors

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

Keywords

User studies, query reformulation, search suggestions and assistance.

1. INTRODUCTION

Search engines are ubiquitous, and millions of people of varying experience use them on daily basis. Unfortunately, not all searches are successful. Bilal and Kirby [3] reported that about half of the participants of their user study felt frustration when searching. Xie and Cool [10] demonstrated that most of the time users have problems with formulating and refining search queries. Besides good retrieval performance, a successful search requires users to possess certain skills. Search skills can be trained, e.g. Google offers a course[1] on improving search efficiency. Although very useful, such courses are time consuming and detached from real

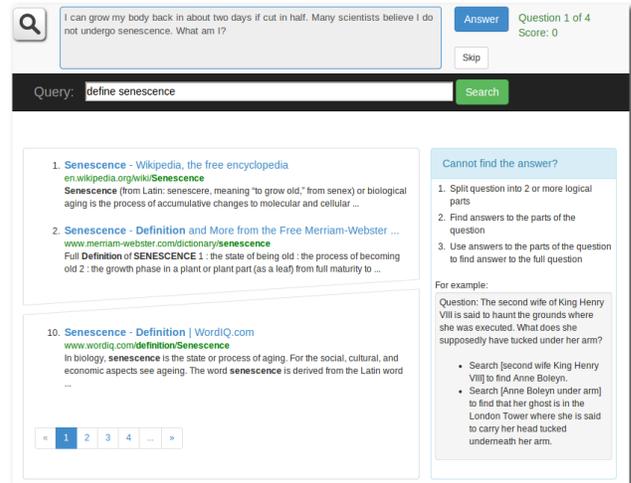[1]http://www.powersearchingwithgoogle.com

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

search problems of these particular users. Displaying search hints is another technique that has both learning effect, and offers immediate assistance to the user in solving her current search task. Moraveji et al. [8] demonstrated that hints, suggesting certain search engine functionality, help people find answers more quickly, and the effect is retained after a week without hints.

In this paper we focus on *strategic* search hints, that are designed to guide a user in solving her search problem. More specifically, we chose the divide-and-conquer strategy, i.e., splitting an original difficult question into smaller problems, searching answers to the subtasks and combining them together. Two sets of strategic hints were manually designed: *generic* hints describing the divide-and-conquer strategy in general and *task-specific* hints providing a concrete strategy to solve the current search task. To evaluate the effect of the hints on behavior and search success we conducted a user study with 90 participants. The results of the user study, described in this paper, demonstrate that well-designed task-specific hints can improve search success rate. In contrast, generic search hints, which were too general and harder to follow, had negative effect on user performance and satisfaction.

2. RELATED WORK

There has been considerable amount of work on search assistance and improving user experience with feedback, suggestions and hints. Results of the study in [10] demonstrate

Table 1: Search tasks used for the study, and specific search hints shown to one of the user groups

	Question	Correct Answer	Specific hints
Task 1	I can grow body back in about two days if cut in half. Many scientists think I don't undergo senescence. What am I?	Senescence means "biological aging". Hydra is considered biologically immortal and regenerates fast.	1. Find what is senescence 2. Find who does not undergo senescence 3. Find who can also regenerate body and choose the one that satisfies both conditions
Task 2	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.	1. Find the names of the gods from the Archaic triad 2. For each of the gods find a Greek counterpart
Task 3	As George surveyed the "waterless place", he unearthed some very important eggs of what animal?	"Gobi" in Mongolian means "Waterless place". The first whole dinosaur eggs were discovered there in 1923.	1. Find what is the "waterless place" mentioned in the question? 2. Search for important eggs discovery in this "waterless place"
Task 4	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.	1. Find the name of the battle mentioned in the questions 2. Search for which coded communications language was used in this battle

that in 59.5% of the cases users need help to refine their searches or to construct search statements. Individual term ([9]) or query suggestion ([2, 4, 5]) are among the most popular techniques for helping users to augment their queries. The study in [6] demonstrated that users prefer query suggestions over term relevance feedback, and that good manually designed suggestions improve retrieval performance. Query suggestion methods usually use search logs to extract 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 suggestions was presented in [7] and users reported overall satisfaction with its automatic non-intrusive advices. Modern search engines have many features that are not typically used by an average user, but can be very useful in particular situations as shown in [8]. The study demonstrated the potential effectiveness and teaching effect of hints. The major difference of our work from [8] is the type of search hints used. Rather than suggesting to users the available search functionality, this work focuses on *strategic* search hints, designed to solve difficult informational questions.

3. USER STUDY

To estimate the effect of strategic search hints on user behavior we conducted a study in a form of a web search game similar to "a Google a Day"[2] and uFindIt [1]. Participants were hired using Amazon Mechanical Turk[3].

The goal of the web search game used in the user study is to find answers to several questions with the provided web search interface (Figure 1). Players are instructed not to use any external tools. The questions are given one by one and 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. To an-

swer a question each player needs to provide a link to a page containing the answer as well as its text. The answer is automatically verified and a popup box notifies a player if the answer is incorrect (since the answer can be formulated differently, presence of a keyword was checked). A player can then continue searching or skip the question when she gives up. A bonus payment was made to players who answer all questions correctly. We used Bing Search API[4] as a back-end of the game search interface. All search results and clicked documents were cached so users asking the same query or clicking the same page got the same results. At the end of the game a questionnaire was presented asking for feedback on user satisfaction with the game, prior experience and other comments.

The tasks for the study were borrowed from the "A Google a Day" questions archive. Such questions are factual, not ambiguous and usually hard to find the answer with a single query, which makes them interesting for user assistance research. We filtered search results to exclude all pages that discuss solutions to "A Google a Day" puzzles. To do this we removed 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 limited the number of questions to 4. The tasks are described in Table 1 and were presented to all participants in the same order to ensure comparable learning effects.

The questions have multiple parts and to solve them it is helpful to search for answers to parts of the questions and then combine them. In one of the previous studies we observed, that most of the users didn't adopt the divide-and-conquer strategy, but kept trying to find the "right" query. We decided to estimate the effect of strategic search hints, suggesting users to adopt the new strategy.

We built 2 sets of strategic hints: *task specific* and *generic*. Task-specific hints described steps of one of the possible solutions to each question (Table 1). Second set contained a single hint, which was shown for all tasks. Generic hint described the divide-and-conquer strategy:

1. Split the 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, the 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.

To control for the learning effect demonstrated in [8], each user was assigned to one of the three groups:

1. users who didn't get any hints
2. users who got task-specific hints
3. users who got the generic hints

4. RESULTS

From 199 unique participants, who clicked the HIT on Amazon Mechanical Turk only 90 players finished the game. We further examined all games manually and filtered out 9 submissions for one of the following reasons: lack of effort (e.g. skipped several tasks after none or a single query) or usage of external resources (e.g. the answer was obtained without submitting any queries or results explored didn't contain the answer). Furthermore, 10 players from the group which received hints indicated in the survey that they didn't see them, so we filtered out those submissions and finally we had 71 completed games (29 for no hints, 20 for task-specific hints and 22 for generic hints groups).

4.1 Effects of Search Tips on Performance

In order to measure search success rate we looked at the number of questions answered correctly by different groups of users[5]. Figure 2 shows that success rate is higher for users who saw task-specific hints compared to users who didn't get such assistance. Surprisingly, having the generic hint decreased the success rate, although users could easily ignore a hint they didn't like. A possible explanation is: generic hints were harder to follow and users who tried and failed became frustrated and didn't restart their searches.

The plot of average time to answer a question on Figure 3 doesn't show an improvement for the task-specific hints group, except for the question 1. Our task-specific hints represent a possible way to solve a problem and there is no guarantee, that it is the fastest one. It is worth noting, that users from the generic search hint group had slightly higher variance in success time, which can probably be explained by the fact that some users were successful in finding the right way to follow the hint and some other users struggled with it much longer. Another insight comes from the number of incorrect attempts users made. Figure 4 demonstrates the average number of incorrect answer attempts for all groups of users. Although the variance is high, there is a tendency for users who saw task-specific hints to make less attempts than both other groups. This is not in direct correspondence

[5]Since users were allowed to skip a question we are counting the number of questions that were eventually solved correctly even if a player made some incorrect attempts

with time spent on the game. It seems that the users who saw a clear strategy to solve the question were less likely to notice plausible, but incorrect solution. Moreover, we analyzed texts of incorrect answers, and can conclude that a big part of incorrect submission are due to users trying all possible options they found on the way, even if these options are clearly wrong.
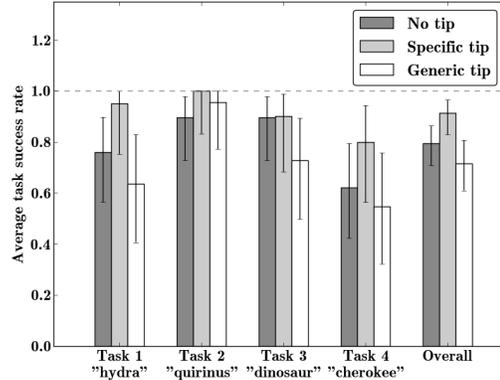

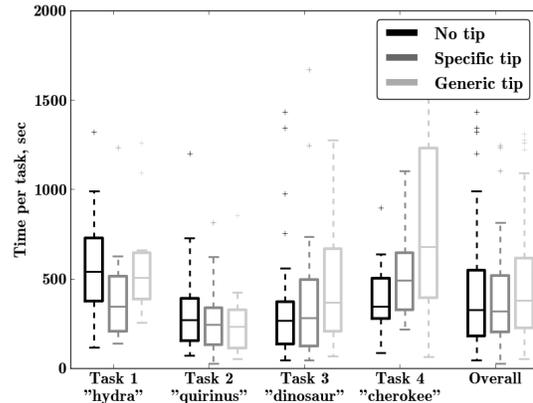Figure 2: Success rate per task for each group of participants

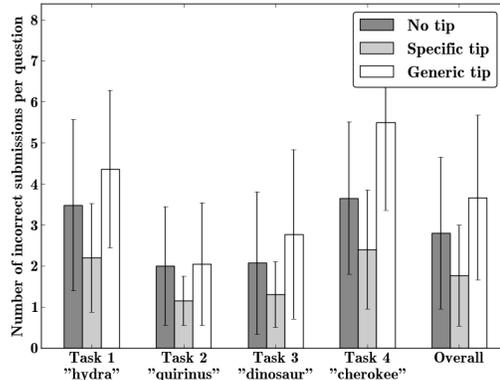Figure 3: Task completion time for each group of players

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

We also looked at other search behavior characteristics: number of queries submitted, number of clicks made, average length of the queries. The variance in these characteristics was too high to make any speculations regarding their meaning.
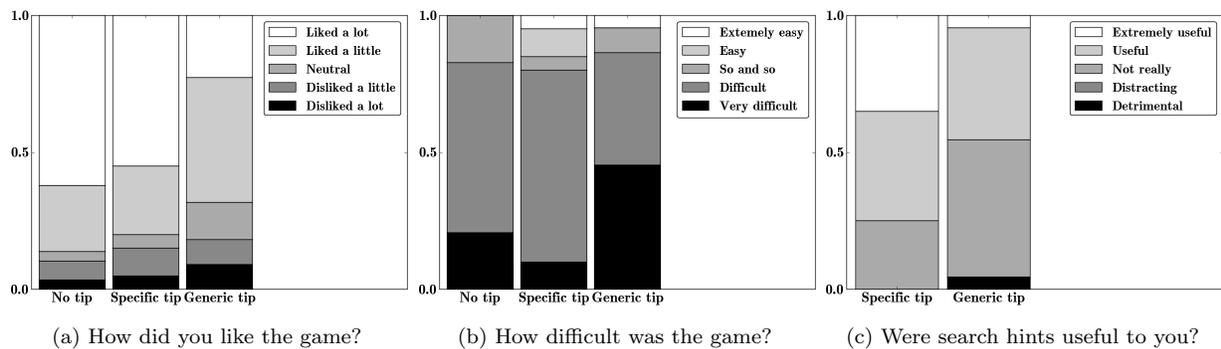
Figure 5: Proportions of replies to some of the survey question for each group of users

4.2 Effects of Search Tips on User Experience

Finally, we looked at the surveys filled out 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 hints useful to you?". Surprisingly, user satisfaction with the game was lower for users who saw hints during the game and users who didn't get any assistance enjoyed it more. The replies to the question about game difficulty are in agreement with the success rate: users who saw task-specific hints rated difficulty lower than participants who struggled to find the correct answers. The game was very difficult on average, however, some participants from the group who received task-specific hints surprisingly rated it as very easy, which suggests that our hints do help users. This is supported by the answers to the last question on whether hints were helpful (Figure 5c).

To summarize, the results of the conducted user study suggest that specific search hints can be helpful, which is indicated by higher success rate, lower number of incorrect attempts and positive feedback in the end of study survey. In contrast, generic hints can have negative effect on user experience, which is indicated by lower success rate, increased number of incorrect attempts and higher perceived tasks complexity according to the survey.

5. CONCLUSION

In this paper we studied the effect of strategic search hints on user behavior. The conducted user study in a form of a web search game demonstrated the potential of good hints in improving search success rate. However, to be useful, they should be designed carefully. Search hints that are too general can be detrimental to search success. We also find that even searchers who are more effective using specific search hints, feel subjectively less satisfied and engaged than the control group, indicating that search assistance has to be specific and timely if it is to improve the searcher experience.

We should note, that specific search hints used in this work were manually generated and an interesting question of future work is how to generate such useful hints automatically. It should be possible to learn strategies applied by the experienced search users and suggest them to the rest.

6. ACKNOWLEDGMENTS

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*, pages 345–354, New York, NY, USA, 2011. ACM.

[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*, pages 795–804, New York, NY, USA, 2011. ACM.

[3] D. Bilal and J. Kirby. Differences and similarities in information seeking: Children and adults as web users. *Inf. Process. Manage.*, 38(5):649–670, Sept. 2002.

[4] 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 International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 875–883, New York, NY, USA, 2008.

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

[6] 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*, pages 371–378, New York, NY, USA, 2009.

[7] S. Kriewel and N. Fuhr. Evaluation of an adaptive search suggestion system. In *Advances in Information Retrieval*, volume 5993 of *Lecture Notes in Computer Science*, pages 544–555. Springer Berlin Heidelberg, 2010.

[8] 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*, pages 355–364, New York, NY, USA, 2011.

[9] 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.

[10] 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.