

Adam T. Mack. The Power of Suggestion: An Evaluation of the Effects of Source and Position on the Use of Query Suggestions. A Master's paper for the M.S. in I.S. degree. April, 2011. 35 pages. Advisor: Diane Kelly

This study examines whether the perceived source of query suggestions and their location within the search interface affects users' likelihood of clicking on the suggested queries. Participants were asked to complete three search tasks using the search system provided. The query suggestions were placed on either left or right side of the page, and were labeled as either system-generated or as having come from other users of the system. Participants also evaluated their engagement with the search tasks and the quality and usefulness of the query suggestions. Results indicated that users presented with query suggestions on the left scored significantly higher on two measures of engagement. While the effects of source did not meet tests of statistical significance, participants who believed the query suggestions came from other users had higher mean scores for three of the four search engagement scales and the query suggestion rating scale.

Headings:

Information Retrieval / Online Searching

Information Retrieval / Evaluation

End-user Searching

THE POWER OF SUGGESTION: AN EVALUATION OF THE EFFECTS OF
SOURCE AND POSITION ON THE USE OF QUERY SUGGESTIONS

by
Adam T. Mack

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April, 2011

Approved by:

Diane Kelly

Introduction

With the advent of the digital age, the amount of information that can be easily accessed has increased rapidly. This proliferation has necessitated the creation of new or refined methods for aiding users in their search process. In large measure, the ability of an information system to return results that are relevant is dependent on the construction of a search query (White and Marchionini, 2007). A variety of approaches have been used, including assisting with the initial query construction process and providing alternative query formulations once a user's search has been submitted. Even the most well-conceived methods for assisting in information retrieval, however, cannot be effective if they are disregarded by users. Hence, it is important to investigate factors that may encourage searchers to utilize such methods or that may cause users to avoid them. The knowledge gleaned from this inquiry will assist in the development of information retrieval aids that people will use and that make sense for users of the particular information system.

A key aspect in assisting users' efforts to meet their information need is the ability to match search terms with documents that possess relevant information, even if the search terms may not correspond exactly to those used in the document. The information system may make recommendations based on multiple factors, including the use of similar search terms, similar queries previously submitted to the system, or documents retrieved by users with similar profiles (Terveen & Hill, 2001). Web-based search engines are one example of an information system that attempts to assist the user in

constructing more effective queries, using methods such as query suggestions, expansion, and refinement. Features such as AltaVista's Prisma and Google Suggest use these methods to present users with lists of potentially relevant queries (Rose, 2006). In order to maximize the effectiveness of such features, it is important to understand the reasons why users choose to use or ignore them, either consciously or subconsciously. Users' tendencies to make human attributions to computers (Moon & Nass, 1998) may negate the possibility that they would prefer socially generated suggestions (i.e. suggestions from other users). Conversely, findings from researchers such as Aharoni and Fridlund (2007) indicate that users do treat interaction with computers differently than interaction with humans, suggesting that the source of a query suggestion may affect users' likelihood of utilization. The location of query suggestions may also potentially play a role in users' likelihood of clicking on them, as prior experience may lead users to ignore content in certain regions of the page (Pandey et al., 2010).

This study examined two factors' effects on the use of query suggestions presented by an information retrieval (IR) system. The first factor examined was the perceived source of the query suggestion. Specifically, the study compared the use of query suggestions labeled as having come from other users against query suggestions labeled as having been generated by the system itself. The second factor examined was the positioning of the query suggestions within the interface. For this factor, the study compared the use of query suggestions placed on the left side of the page against query suggestions located on the right side of the page. It is hypothesized that users may be more likely to utilize query suggestions thought to have come from other users of the IR

system. Further, it is hypothesized that positioning query suggestions on the left of the page will result in increased utilization of these suggestions.

Literature Review

This literature review will examine the use of recommender systems to help users find information and the perception of recommendations by users. Methods of query expansion and refinement will be discussed, as well as the use of query suggestions. Factors potentially affecting users' perceptions of these technologies will be identified. Finally, to explain how users may interpret their interactions with an information retrieval system, research from the fields of psychology and human-computer interaction will be considered.

Recommender systems

Recommender systems are one way in which computers can help users overcome the sense of information overload that may result from the staggering amounts of available information. Such systems may operate in numerous ways (Carroll, 2001). They may consider only the preferences of the current user based on prior actions, may perform data mining on records of social activity, or may compare a user's preferences to other users with similar preferences. While each of these implementations may have its own positive and negative aspects, each must perform its task of connecting users with information in an effective manner in order for it to become widely used. Krishnan and colleagues (2008) investigated the ability of recommender systems to predict items that a user will like in comparison to the ability of humans to perform the same task. Participants were presented with profiles of movie ratings from the MovieLens system maintained by GroupLens research. Based on the ratings of 30 movies in the user profiles presented, participants were asked to predict ratings for 10 movies using a five-star scoring system. On average, the MovieLens algorithm performed better than the human

predictors, although a number of participants were able to outperform the system. The results of this study suggest that recommender systems can effectively predict the preferences of users in some situations.

Users who are knowledgeable about the types of recommendations provided by information retrieval systems may interpret the validity of the suggested searches differently based on the label used. Good and his colleagues (1999) discuss two approaches to recommendation systems—information filtering and collaborative filtering. The information filtering (IF) approach is based on the content of the item; a hypothetical music recommendation system using this method might suggest songs labeled with the same genre as a particular song. The content filtering (CF) approach instead examines the behavior of similar groups of users; a system using this method might suggest songs purchased by users who also purchased a particular item. Search suggestions labeled as coming from other users might be more likely to be interpreted as using the CF method, while “system-generated” would be more associated with the IF approach. If users make these types of assumptions, they might make erroneous judgments regarding the usefulness of the search suggestions provided.

It may also be the case that recommender systems influence users’ opinions of the items they recommend. Cosley and colleagues (2003) examined this phenomenon with regard to aspects of the recommender system’s user interface. This study also made use of the MovieLens recommender system. The study included three experiments; in one, users were asked to re-rate movies they had previously rated and were presented with a predicted rating slightly lower, the same, or slightly higher than their previous rating. The second experiment was similar to the first, but involved movies not previously rated.

Finally, the third experiment involved rating previously rated movies on a new scale. The researchers found that showing predictions can influence users to rate items differently. While the proposed study may not provide users with predicted ratings for suggested searches, it may prove to be the case that the act of making a recommendation leads users to rate the effectiveness of these search results differently.

Query Expansion and Refinement

A Web search engine interface can have significant effects on the behavior of its users. Rose (2006) points out the fact that early search engines gave the user only a very limited amount of space in which to enter a query, resulting in a greater tendency for users to enter short queries. Search also tends to be an iterative process, with users refining the query based on results returned for an initial query (Rose, 2006). Rose highlights how a search engine's user interface can help to guide the user through this process with the example of an AltaVista feature (Prisma) that suggests related terms. Users have the option of either replacing their query with the suggested terms or adding the terms to the existing query. While this feature was removed from AltaVista prior to the publication of Rose's (2006) paper, it serves as an example of how the user interface can help users to formulate better queries.

Unfortunately, adding features such as Prisma (an example of interactive query expansion) to a user interface in order to assist in query formulation is fruitful only if users are willing to utilize the new feature. Anick (2003) investigated the uptake of AltaVista's Prisma feature in a study of anonymous search engine logs. Only 16 percent of users used the query feedback in order to revise their original search query. Usability tests conducted by AltaVista prior to the feature's public release suggest multiple reasons

why users may have avoided using Prisma. Users may not have noticed the presence of the query expansion terms on the page, or they may have assumed that they were present for advertising purposes. However, nearly half (47 percent) of those who used Prisma once within a two-week period of observation used it again during the same two-week period (Anick, 2003). When users chose to include feedback terms in their query, they experienced equal success in finding relevant documents as users who manually revised their queries.

Use of Query Suggestions

As a method of query expansion, presenting users with query suggestions can assist the search process by providing different terms or making connections to related topics that a searcher might not consider on their own. By generating new ways of approaching a topic, query suggestions can ultimately lead the user to a broader understanding of a subject of interest (Kelly et al., 2010). They can also improve the user's search efficiency when performing multiple queries on a given topic, as clicking on a link requires less time and effort than entering search terms into a query field (Kelly, Gyllstrom, & Bailey, 2009).

Query suggestions may be generated in multiple ways. Terms related to those used in the original query may be added or substituted. Query popularity may also be considered, although this may result in a loop whereby a query becomes more popular because it is displayed as a suggestion (Kelly et al., 2010). Given the vast number of searches performed by users of Web search engines, these systems can inspect the search sequences of previous users that issued a similar query to find potential suggestions (White, Bilenko, & Cucerzan, 2007). Regardless of the method used to generate query

suggestions, rarely if ever are users made aware of the process by which suggested queries are selected (Kelly et al., 2010). Barring technical knowledge of the particular information system being used, a searcher is left to form their own ideas about the query suggestions' origins; these ideas may then color their evaluation of the potential value of the suggestions.

The ways in which information retrieval mechanisms present users with finding tools such as query suggestions and query expansions can affect users' likelihood of utilizing such tools. Studies such as that performed by Anick (2003) have investigated aspects of the interface which may influence users' search behavior. One area that has not been as thoroughly investigated, however, is how users interpret suggested search terms based on their perceived source. Results may be framed as having been generated by the information retrieval system, or they may be presented as queries submitted by other users. Though the literature does not currently address the differences in how users respond to these specific presentations, there is an abundance of work in the area of human-computer interaction that is potentially relevant.

Psychology of Human-Computer Interaction

Youngme Moon and Clifford Nass at Stanford University have conducted numerous studies investigating the interactions of humans with computers. Nass, Steuer, and Tauber (1994) examined the social nature of human interactions with computers to determine what types of human social rules people apply to computers. They studied an array of behaviors, including politeness, the concept of self vs. other, gender stereotypes, and whether the social tendencies exhibited by users were attributed to the computer itself or an outside agent (for example, a programmer). Their findings suggest that social

norms such as politeness are observed when interacting with computers, as is the self/other concept. Furthermore, the social behavior was found to represent the user's interaction with the computer rather than with an outside (human) agent. In the context of the current study, the findings of Nass and his colleagues suggest that users may make human attributions to the information retrieval system; this may occur regardless of how the search suggestions are labeled.

Moon and Nass (1998) looked at users' attributions of responsibility when interacting with computers; specifically, they examined what circumstances resulted in the user blaming the computer for negative outcomes and what circumstances resulted in the user giving the computer credit for positive outcomes. They examined this behavior in the context of both personality similarity and user control. When participants perceived themselves as dissimilar to the computer, they tended to exhibit a self-serving bias that resulted in blaming the computer for failures and taking credit for successes themselves. A perception of similarity resulted in the opposite pattern. An increased sense of user control also led participants to take greater responsibility for both successes and failures. It may be that these findings extend to the way in which an information retrieval system interface is presented. Search suggestions labeled as having come from other users may make a user less likely to view the retrieval system negatively if it fails to find a relevant document. This may make a user more likely to continue using an information retrieval system despite some less-than-successful search experiences.

User interfaces and human-computer interaction

Given users' tendency to make human attributions toward computers, why should it make any difference how the search results are labeled in the search suggestion

interface? Research suggests that people approach interactions with other people differently than they approach interactions with a computer system. Shechtman and Horowitz (2003) examined the discrepancies using an experimental design whereby participants had text-based interactions with a “partner” who they were told was either a computer program or another human. Instead, participants in both conditions received scripted responses. After checking to ensure that the manipulations had been effective (that is, that participants believed they were interacting with the appropriate type of partner), the data were analyzed to identify any potential differences. More words were used when corresponding with “human” partners, and more time was spent composing comments to these partners. Participants in the “human” condition used a substantially greater number of relationship statements, which were identified as connecting, influencing, yielding, or hostile (Shechtman & Horowitz, 2003). If users form a stronger connection with the partners labeled as human in this scenario, it may also be the case that users have a similar reaction when search suggestions are presented as being human generated.

Aharoni and Fridlund (2007) also examined the differences between users’ interactions with systems based on whether they had been identified as human or computer. Participants were informed that they were going to interview for a mock job by either a computer or a human. Verbal and nonverbal behavior of participants was recorded during the interview as well as afterward, when the participant was informed of acceptance or rejection. Self-reports of emotion and interviewer impressions were also collected. Participants did not feel any happier toward the “human” interviewer; nor did they describe the “human” as more likable or more sociable. However, they smiled more

when interacting with the “human” interviewer, and spoke more to him/her (Aharoni & Fridlund, 2007). Despite the fact that both human and computer interviewers behaved identically, users responded differently if they thought they were interacting with another human. Since a simple text label can demonstrably affect a user’s actions in this domain, it is reasonable to hypothesize that a similar phenomenon may occur in the context of an information retrieval interface.

HCI and Search

The significance of source is highlighted in a study by Sundar and Nass (2001). In this study, participants were given six news articles to read through an online news service. Though articles remained the same for all participants, the attributed source of the article was varied. One-fourth of the participants were informed that the articles were selected by news editors; another fourth were told their computer terminal selected the stories; another fourth were advised that the articles were selected by other users of the online news service; and the final group were led to believe they had selected the articles themselves (via a pseudoselection task). Participants were asked about the perceived credibility of the article, how much they liked the article, its quality, and its representativeness of the stories they read. With regards to the current study, the results were mixed. Stories were liked more and perceived to be of higher quality when selected by other users than when selected by the user or by news editors. However, when a computer is identified as the source, articles are also rated higher in quality than when selected by either news editors or the user. While there were no statistically significant differences between the ratings for news items selected by other users or the computer, they did differ slightly in their relationships to the other two source types. Additionally, a

study which narrowed the number of source variations to two (the computer and “other users”) would be better able to identify differences.

Sundar and colleagues (2006) further examined the significance of source with regard to the various elements provided by news aggregation sites such as Google News in users' assessments of a news item's relevance. News aggregators bring together an array of news coverage far greater in scope and variety than conventional news sources, potentially creating the possibility of information overload. To provide users with some degree of context, a news aggregator frequently displays information such as the name of the primary source for the news item, how long ago the item was published, and the number of articles found for the same topic. Sundar and colleagues found that source was the prime characteristic that affected the credibility of a news item. Stories with a high credibility source exhibited little effect from variations in the recency or number of related articles; these characteristics only came into play with low-credibility sources. This study again underscores the vital importance of source in a user's perception of information credibility. Manipulating the perceived source of search suggestions may in fact produce a significant shift in likelihood that users will take advantage of these suggestions, as well as their satisfaction with the results of suggested searches.

In addition to source, positioning of interface elements such as query suggestions can affect how they are perceived by users. In particular, users may learn to ignore items placed in a certain part of the page if they frequently contain irrelevant or unwanted results (Pandey et al., 2010). As users become habituated to this content from repeated exposure, it may no longer draw their attention (Portnoy & Marchionini, 2010). This results in a phenomenon known as “banner blindness.” A study conducted by Chatterjee,

Hoffman, and Novak (2003) found that users of an ad-supported Web site were less likely to click on banner advertisements later within a single session as well as after repeated sessions. Another potential example of this behavior pattern involves the typical search engine placement of sponsored links on the right side of the search results page. Users who are accustomed to ignoring these links on a search engine may have a tendency to avoid query suggestions placed in this region of the page by another information retrieval system.

An eye tracking study conducted by User Centric, a consulting firm, found that only 28 percent of participants looked at sponsored results on the right side of the Google results page (User Centric, 2011). The proportion of participants who looked at Bing's sponsored results section on the right side of the page was even lower (21 percent). This demonstrates a clear tendency to ignore links placed on this side of the page, which might extend to query suggestions that are placed to the right of the search results. On the other hand, it is far from clear that placing query suggestions to the left of the page would increase utilization. The same User Centric study found that the left pane of Bing's search result page, which contains a list of related searches, was viewed by only 18 percent of study participants. Additional examination of the positioning of query suggestions could help to clarify whether users possess a distinct preference for either location.

Method

Study Design

This study utilized a 2x2 factorial design; both independent variables were manipulated between subjects. The first independent variable was the position of the query suggestions, which were displayed on either on the left or right side of the search page. The second independent variable was the heading shown above the query suggestions. In one condition, the search system labeled the suggestions to indicate that they were drawn from queries submitted by previous users of the system (“Other users suggest these queries”). In the other condition, query suggestions were labeled to indicate that the system itself had generated them (“The system suggests these queries”). In both instances, however, users received identical query suggestions.

Participants

Twelve participants were recruited from the undergraduate population of the University of North Carolina at Chapel Hill. Respondents ranged from 19 to 21 years of age ($M=20.0$, $SD=0.60$), and 67 percent of the participants were female. Academic majors represented (including students with multiple majors) were Business (2), Political Science (2), American Indian Studies (1), Anthropology (1), Chemistry (1), English (1), Environmental Health Science (1), Exercise and Sports Science (1), History (1), Mathematics (1), Pharmacy (1), Psychology (1), and Sociology (1). Students who responded to the initial recruitment email were randomly assigned to one of the four possible conditions using a random number generator available on the Internet (Random.org). The participant was then provided with credentials for logging into the study that corresponded with the assigned condition.

	System	Other Users	Total
Left	3	5	8
Right	1	3	4
Total	4	8	12

Table 1. Participants assigned to each condition

A mass emailing list of all students was used in order to advertise the study. The recruitment email contained a brief description of the study and informed students that completion would enter them into a prize drawing for one of five \$30 online gift cards. Initially the recruitment message was only sent to first-year and sophomore students, due to concerns about overlap between participants in this study and a previous study which utilized the same search system. Ultimately, it was necessary to broaden the recruitment in order to increase the number of responses.

Corpus and Search Topics

The text documents to be searched by the information retrieval system came from a test collection from Text Retrieval Conference (TREC) (Voorhees, 2006). The collection includes over one million documents, with three gigabytes of newswire text. Fifty search topics are part of the original collection; of the fifty, three were selected for this study.

Topic Title	Description as presented to participants
Mercy killings (393)	As a new reporter for the Daily Tarheel, your first assignment is to prepare a story about a controversial topic. You select mercy killings. All individual cases of mercy killing are relevant, except that "letters to the editor" mentioning cases are not relevant. The removal of life support systems is relevant. A general mention or description of a case without specifics, such as victim's name are not relevant. Cases determined to be a murder-suicide are not relevant.
Law enforcement, dogs (426)	This semester you are volunteering at the Chapel Hill Police Department in the canine unit. Your first assignment is to find out as much as possible about the use of dogs worldwide for law enforcement purposes. Relevant items include specific information on the use of dogs during an operation. Training of dogs and their handlers are also relevant.
Wrongful convictions (638)	You are enrolled in a criminal justice class and your professor has asked you to prepare a paper about wrongful conviction. Specifically, your paper should discuss freed prisoners who have been wrongfully convicted based on faulty forensic evidence, poor police work, or false testimony. Documents about political prisoners who were freed because of incompetent prosecutions are relevant. However, documents that discuss prisoners who are pardoned or released on bond when their convictions are overturned are not relevant, nor are documents about prisoners freed to make a political statement or prisoners freed for an exchange.

Figure 1. Search Topics

These topics were chosen because of their potential to interest undergraduate students.

Consideration was also given to avoiding topics which matched closely with high-profile current events, since news items for these events would not be included in the corpus.

Procedure

Participants completed the study on their own computer in a location of their choosing, rather than in a lab setting. After logging into the system using the provided credentials, consenting to participate in the study, and completing a brief demographic questionnaire, participants were asked to complete three search tasks using the search system provided.² For each of the three tasks, participants were presented with a search topic, and then asked to conduct a search for relevant documents using the information retrieval (IR) system provided. After entering an initial search query for the topic, the IR system presents the results of the search, along with a list of six suggested queries for the search topic.

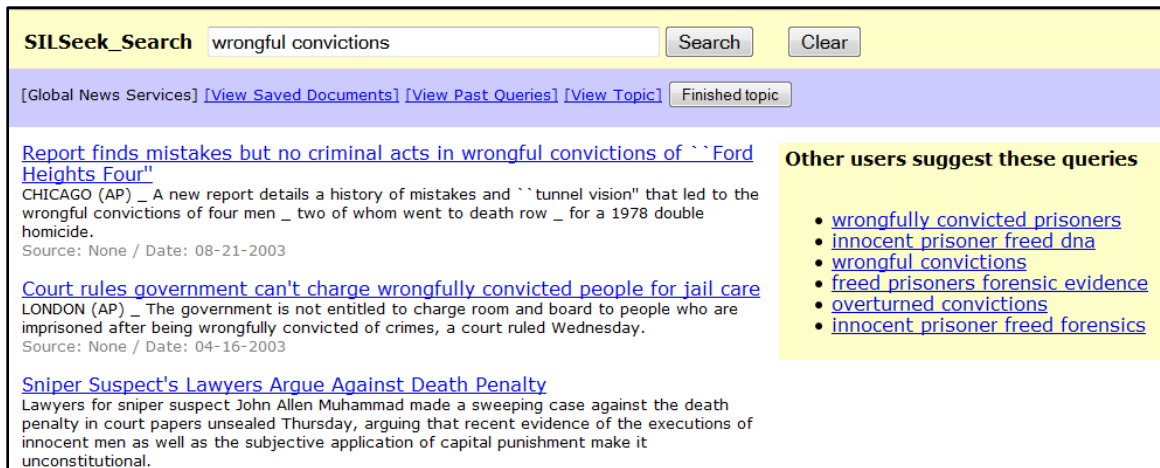


Figure 2. Interface of the search system

From this screen, participants could either enter another query by typing in the query box or clicking on a query suggestion, or click on a search result to display a document examine its relevance. Upon viewing a document, participants could save it if it was relevant to the topic. Participants were given up to ten minutes to search on a given topic, although they could move on to the next task at their own discretion. The sequence of the three tasks was varied across participants to prevent order bias. Upon completion of the three search tasks, participants were presented with an exit questionnaire which asked about their level of engagement with the search tasks and their satisfaction with the query suggestions presented. Finally, participants were presented with a debriefing message thanking them for their participation.

Measures

A brief demographic questionnaire consisting of four questions was used to gather information on the study sample prior to engaging in the three search tasks. The data from the questionnaire was used only to describe the sample and not for purposes of analysis.

Demographic Questionnaire

What is your age?

What is your sex?
 Female
 Male

What is your academic standing at UNC?
 First-year student
 Sophomore
 Junior
 Senior
 Graduate Student

What is your major / course of study?

Figure 3. Demographic Questionnaire

During the search tasks, the IR system recorded each action taken by a participant and stored it in an HTML log file. For each action, the log file indicates the time the action was taken, the ID number of the user, and a description of the action taken. Actions recorded include the issuing of a query by the user, clicking on a query suggestion, viewing a document that was returned as a search result, saving a relevant document, transitioning between topics, and the completion of the search tasks.

```

At 2010-12-02 15:00:45, 273 issued the query, "faulty forensic evidence in cases", for topic, "638"
At 2010-12-02 15:00:48, 273 triggered the results function call.
At 2010-12-02 15:00:49, 273 went to url http://ruby.ils.unc.edu:9191/results?recommended=1&query=innocent%20prisoner%20freed%20dna&ref=61863.
At 2010-12-02 15:00:49, 273 issued the query, "innocent prisoner freed dna", for topic, "638"
This query was recommended based on query "faulty forensic evidence in cases"
At 2010-12-02 15:00:59, 273 triggered the result_reader function call.
At 2010-12-02 15:00:59, 273 opened file LAT20030919.0114.
At 2010-12-02 15:00:59, 273 went to url http://ruby.ils.unc.edu:9191/open?query\_text=innocent+prisoner+freed+dna&filename=LAT20030919.0114&fileid=488579&headline=++La.+Man+Serving+Time+on+Rape+Conviction+Released+From+Prison++.
At 2010-12-02 15:01:00, 273 saved file LAT20030919.0114.
At 2010-12-02 15:01:00, 273 triggered the saveFile function call.
At 2010-12-02 15:01:05, 273 triggered the results function call.
At 2010-12-02 15:01:05, 273 went to url http://ruby.ils.unc.edu:9191/results?recommended=1&query=innocent%20prisoner%20freed%20dna&ref=61863.

```

Figure 4. Log File from the Search System

Following the search tasks, participants were presented with a search engagement questionnaire (O'Brien & Toms, 2010). This measure consisted of 19 questions

examining multiple aspects of the participants' search experience. Each item asked participants to respond on a five-point agreement scale, from one (strongly disagree) to five (strongly agree). The measure contains four subscales that evaluate focused attention (items 2, 6, 10, 14, 17), perceived usability (items 1, 3, 7, 11, 15, 18, 19), durability (items 4, 8, 12, 16), and feeling of involvement (items 5, 9, 13).

1. I could not do some of the things I needed to do with this search system.	11. I felt discouraged while using this search system.
2. I was so involved in my search experiences that I lost track of time.	12. My search experiences did not work out as I had planned.
3. I found this search system confusing to use.	13. The search experience was fun.
4. Searching on this system was worthwhile.	14. I was absorbed in the searching tasks.
5. I was really drawn into my searching tasks.	15. Using this search system was mentally taxing.
6. When I was searching, I lost track of the world around me.	16. My search experiences were rewarding.
7. I felt annoyed while using this search system.	17. During the searching tasks I let myself go.
8. I consider my search experiences successful.	18. The search experience was demanding.
9. I felt involved in the searching tasks.	19. I felt in control of my searching experience.
10. The time I spent searching just slipped away.	

Figure 5. Search Engagement Questionnaire.

Once participants completed the search engagement questionnaire, they were asked whether they noticed the query suggestions as well as whether they used the query suggestions. Participants who noticed but did not use the query suggestions were asked why they chose to ignore the suggestions; those who both noticed and used the query suggestions were asked to complete a questionnaire evaluating their quality and usefulness. Each of the nine questions asked participants to respond on a five-point scale, from 1 (strongly disagree) to 5 (strongly agree).

1. The query suggestions were useful.	6. The query suggestions helped me modify my own queries.
2. The quality of the query suggestions was good.	7. The query suggestions helped me better understand the topic.
3. The query suggestions led me to relevant documents.	8. The query suggestions helped me think of new approaches to searching for the topic.
4. The query suggestions helped me think of new queries.	9. Overall, the query suggestions made searching easier.
5. The query suggestions helped find relevant documents.	

Figure 6. Query Suggestion Questionnaire

All participants were then given an opportunity to provide feedback regarding their search experience and the search system. Finally, the participants were presented with a debriefing screen thanking them for their participation.

Results

To examine the effects of labeling query suggestions as either system-generated or as coming from fellow users of the IR system, the queries submitted by participants in each group were examined. One participant did not use any of the query suggestions and was therefore excluded from these analyses. (The participant actually reported using query suggestions and hence completed the questionnaire for evaluating the suggestions; however, the system logs for this person indicate no use of query suggestions). Means were calculated for the number of query suggestions used across all three search tasks, the total number of queries submitted across all tasks, and the proportion of total queries that were the result of clicking on query suggestions.

Variable	Group	N	Mean	SD
Number of query suggestions used	System	4	8.75	4.27
	Other Users	7	11.86	9.77
Total number of queries	System	4	34.00	18.57
	Other Users	7	36.86	4.34
Proportion of query suggestions to total number of queries	System	4	0.285	0.092
	Other Users	7	0.324	0.248

Table 2. Query Means by Heading

As Table 2 shows, the differences between participants who received the “system” heading and those who received the “other users” heading are minor. In particular, the total number of queries for both groups of users is nearly indistinguishable. The high standard deviations for the number of query suggestions and total number of queries indicate that the slight differences for these variables are not likely to be statistically meaningful. It is noteworthy, however, that both the absolute number of query suggestions used as well as the proportion of queries that came from query suggestions are higher for participants who believed the suggestions came from other users.

The effects of varying the query suggestion heading were also examined with regard to the participants' experience of the system and their evaluation of the query suggestions. In order to examine these results, a mean value was calculated for each of the four subscales of the Search Engagement questionnaire as well as for the Query Suggestion questionnaire based on their respective component items. Means were then calculated for each of the two query heading conditions.

Variable	Group	N	Mean	SD
Attention	System	4	2.35	0.60
	Other Users	7	2.74	0.99
Usability	System	4	2.89	1.15
	Other Users	7	3.33	0.56
Endurability	System	4	2.81	1.21
	Other Users	7	3.25	0.43
Involvement	System	4	3.17	1.04
	Other Users	7	3.14	0.60
Query Suggestion	System	4	3.61	0.66
	Other Users	7	3.86	0.41

Table 3. Scale Means by Heading

As Table 3 shows, there appear to be at least some minor differences between the two groups for these measures. While the differences for Involvement and Query Suggestion are relatively small, the variation between groups for Attention and Endurability are more noticeable. Nonetheless, neither Attention ($F(1, 9) = 0.507, p = .49$), nor Usability ($F(1, 9) = 0.738, p = .41$), nor Endurability ($F(1, 9) = 0.791, p = .40$) meet the level of statistical significance. Despite the lack of significant differences between means, there is a general trend apparent in the data. With the exception of Involvement, which was nearly equal between groups, participants in the "other users" condition had higher mean scores for each of the other four scales. A consistent trend such as this suggests that further study, particularly with a larger sample, might prove fruitful.

Next the data was examined for differences based on whether the query suggestions appeared on the left or right side of the search system interface. Mean values for the number of query suggestions, total number of queries, and proportion of queries resulting from query suggestions were calculated for each group; the values are shown in Table 4.

Variable	Group	N	Mean	SD
Number of query suggestions used	Left	7	11.43	10.21
	Right	4	9.50	2.38
Total number of queries	Left	7	34.29	12.31
	Right	4	38.50	8.43
Proportion of query suggestions to total number of queries	Left	7	0.339	0.243
	Right	4	0.259	0.095

Table 4. Query Means by Position

There is very little difference in the number of query suggestions used; the differences are greater for the total number of queries and the proportion of queries coming from query suggestions. A higher value for the proportion of queries coming from query suggestions for participants who received query suggestions on the left of the screen could directly indicate an increased reliance on query suggestions. Combined with a lower value for total number of queries, the findings may suggest that the increased use of query suggestions improved information retrieval efficiency, reducing the need for additional querying. However, given the within-group variation illustrated by the high standard deviations for these variables, neither the total number of queries ($F(1, 9) = 0.362$, $p = .56$) nor the proportion of suggested queries ($F(1, 9) = 0.384$, $p = .55$) approached statistical significance.

The scale means were also examined for differences between those who were presented with query suggestions on the left and those for whom the suggestions appeared on the right.

Variable	Group	N	Mean	SD
Attention	Left	7	2.66	0.78
	Right	4	2.50	1.10
Usability	Left	7	3.47	0.52
	Right	4	2.64	0.99
Endurability	Left	7	3.43	0.51
	Right	4	2.50	0.87
Involvement	Left	7	3.48	0.57
	Right	4	2.58	0.69
Query Suggestion	Left	7	3.78	0.53
	Right	4	3.75	0.51

Table 5. Scale Means by Position

Differences between groups for Attention and Query Suggestion are relatively minor. Potentially meaningful differences appear to exist for Usability, Endurability, and Involvement. Using a one-way analysis of variance (ANOVA), the usability of the search system was not found to be significantly different between groups ($F(1, 9) = 3.417$, $p = .10$) However, the endurability of using the search system was found to be significantly higher for participants presented with query suggestions on the left of the page ($F(1, 9) = 5.145$, $p = .050$). The difference for Involvement was also significant ($F(1, 9) = 5.395$, $p = .045$), with scores higher for participants with query suggestions on the left of the page. The higher scores indicate that displaying the suggestions to the left of the page resulted in a more positive and involving user experience. This finding is consistent with the concept proposed by Pandey and colleagues (2010) that users may have become “trained” to less positively evaluate content placed on the right side of a search interface due to the traditional presence of sponsored links in that location. The finding is further bolstered by the overall pattern of the data. While not all of the scales

showed significant differences between groups, the mean score of participants who viewed the query suggestions on the left of the page was higher for each of the five scales.

Examining the interaction between query heading and query position is tenuous at best, given the low number of participants and uneven distribution of participants across conditions. An examination of the mean number of query suggestions utilized by each group shows that participants who were presented with the other users heading on the left of the page used the greatest number of query suggestions, while those shown the system heading on the right of the page used the fewest query suggestions. However, the differences between groups are relatively small and unable to support any definitive conclusions.

	Left	Right
System	9.0 (5.2)	8.0 (*)
Other Users	13.3 (13.4)	10.0 (2.6)

* Only one person assigned to the system heading/right position group completed the study, so the standard deviation for this group cannot be calculated.

Table 6. Mean (SD) number of query suggestions by heading and position

Prior to the debriefing at the conclusion of the study, participants were given the opportunity to provide feedback in an open-ended format. Nine of the twelve participants provided comments in response to this question. While the responses given were somewhat unique to each participant, there were a handful of themes that appeared in multiple responses. Three users identified different limitations of the search system used that affected their search experience (inability to use advanced search operators, inability to restrict date range of results, and restriction to a single browser tab). Two participants expressed a lack of clarity in the study instructions regarding the ability to save relevant documents found via the search system. A pair of participants also mentioned that there

were relatively few document results for each query. Two responses also addressed more general design issues, such as the overall color scheme and the desire for search terms to be highlighted in the documents returned.

Conclusion

The goal of this study was to examine differences in the use of query suggestions based on their position within the interface and their source. In addition to the number and proportion of query suggestions used, ratings of the query suggestions and engagement with the search system were evaluated for differences based on the source and position of query suggestions. The initial plan for this study involved recruiting a much larger sample of participants in order to allow for more valid comparisons. While the small sample size of this study limits the ability to draw definitive conclusions, the findings do suggest some potential areas for further investigation.

No significant differences were found between participants who received query suggestions identified as coming from other users and those who received suggestions labeled as system-generated. A definite pattern was apparent within the results, however. For three of the four scales measuring engagement as well as the query suggestion rating scale, participants in the “other users” group had higher mean scores. Additional confirmation of this trend could suggest that users more positively view interactions with an information search system that possesses a human or social aspect within its interface.

The differences found between users who saw the suggestions on the left and those who saw them on the right indicate that positioning may be a meaningful factor in how users perceive their experience with an information retrieval system. Specifically, the more positive evaluations of the search experience when query suggestions were placed on the left supports the idea that search engine users have become inured to the presence of sponsored links on the right side of the search interface. Further research could serve to confirm the “banner blindness” phenomenon, and it is possible that

additional differences could be identified given a larger pool of participants. With additional support, these findings might help interface designers create layouts that help users take better advantage of helpful resources such as query suggestions.

Notes

1. I would like to express my gratitude to Dr. Karl Gyllstrom for his work in adapting the search system to address my research questions, without which this study would have been impossible.

References

- Aharoni, E., & Fridlund, A.J. (2007). Social reactions toward people vs. computers: How mere labels shape interactions. *Computers in Human Behavior*, 23, 2175–2189.
- Anick, P. (2003). Using terminological feedback for Web search refinement - a log-based study. In *Proceedings of the 26th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, 88–95.
- Chatterjee, P., Hoffman, D.L., & Novak, T.P. (2003). Modeling the clickstream: Implications for Web-based advertising efforts. *Marketing Science* 22(4), 520–541.
- Cosley, D., Lam, S.K., Albert, I., Konstan, J.A., & Riedl, J. (2003). Is seeing believing? How recommender interfaces affect users' opinions. *CHI 2003* (Ft. Lauderdale, FL, United States, April 5–10, 2003). ACM, New York, NY, 585–592.
- Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwar, B., Herlocker, J., & Riedl, J. (1999). Combining collaborative filtering with personal agents for better recommendations. In *Proceedings of the AAAI '99 Conference*, 439–446.
- Kelly, D., Cushing, A., Dostert, M., Niu, X., & Gyllstrom, K. (2010). Effects of popularity and quality on the usage of query suggestions during information search. *CHI 2010* (Atlanta, Georgia, United States, April 10–15, 2010). ACM, New York, NY, 45–54.

- Kelly, D., Gyllstrom, K., & Bailey, E.W. (2009). A comparison of query and term suggestion features for interactive searching. *SIGIR '09* (Boston, MA, United States, July 19–23, 2009). ACM, New York, NY, 371–378.
- Krishnan, V., Narayanashetty, P.K., Nathan, M., Davies, R.T., & Konstan, J.A. (2008). Who predicts better? – Results from an online study comparing humans and an online recommender system. *RecSys '08* (Lausanne, Switzerland, October 23–28, 2008). ACM, New York, NY, 211–218.
- Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human–computer interaction. *International Journal of Human-Computer Studies*, 49, 79–94.
- Nass, C., Steuer, J., & Tauber, E.R. (1994). Computers are social actors. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating interdependence. B. Adelson, S. Dumais, and J. Olson, Eds. *CHI '94* (Boston, MA, United States, April 24–28, 1994). ACM, New York, NY, 72–78.
- Pandey, S., Punera, K., Fontoura, M., & Josifovski, V. (2010). Estimating advertisability of tail queries for sponsored search. *SIGIR '10* (Geneva, Switzerland, July 19–23, 2010). ACM, New York, NY, 563–570.
- O'Brien, H.L., & Toms, E.G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society of Information Science and Technology*, 61(1), 50–69.

- Portnoy, F., & Marchionini, G. (2010). Modeling the Effect of Habituation on Banner Blindness as a Function of Repetition and Search Type: Gap analysis for future work. *CHI 2010* (Atlanta, Georgia, USA, April 10–15, 2010). ACM, New York, NY, 4297–4302.
- Rose, D.E. (2006). Reconciling information-seeking behavior with search user interfaces for the Web. *Journal of the American Society for Information Science and Technology*, 57(6), 797–799.
- Shechtman, N., & Horowitz, L.M. (2003). Media inequality in conversation: How people behave differently when interacting with computers and people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA, April 05 - 10, 2003). CHI '03. ACM, New York, NY, 281–288.
- Sundar, S.S., Knobloch-Westerwick, S., & Hastall, M.R. (2006). News cues: Information scent and cognitive heuristics. *Journal of the American Society for Information Science and Technology*, 58(3), 366–378.
- Sundar, S.S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51(1), 52-72.
- Terveen, L. G. & Hill, W. (2001). Beyond recommender systems: Helping people help each other. In Carroll, J. (ed.) *HCI in the New Millennium*. Addison Wesley.
- User Centric. (2011). *Eye tracking Bing vs. Google: A second look*. Retrieved from <http://www.usercentric.com/news/2011/01/26/eye-tracking-bing-vs-google-second-look>

- Voorhees, E. M. (2006). Overview of the TREC 2005 Robust Retrieval Track. *Proceedings of 14th Annual Text Retrieval Conference, (TREC-14)*. ACM, New York, NY, 41–48.
- White, R.W., Bilenko, M., & Cucerzan, S. (2007). Studying the use of popular destinations to enhance Web search interaction. *SIGIR '07* (Amsterdam, The Netherlands, July 23–27, 2007). ACM, New York, NY, 159–166.
- White, R.W., & Marchionini, G. (2007). Examining the effectiveness of real-time query expansion. *Information Processing & Management*, 43(3), 685–704.