

USAGE AND PERCEPTIONS OF WEB SEARCH FEATURES

J. Christopher Sandvig, Western Washington University, chris.sandvig@wwu.edu

Deepinder S. Bajwa, Western Washington University, deepinder.bajwa@wwu.edu

Steven C. Ross, Western Washington University, steve.ross@wwu.edu

ABSTRACT

This study assesses perceptions of web search result enhancements on two types of search tasks, factual and investigative, as well as the amount of time required for users to complete each type of search. Enhancements are determined to be more effective with factual searches, where the result of the search often answers a question. Investigate searches take more time to complete and enhancements are viewed as less effective.

Keywords: Web Search, Search Engine Features, Google, Bing, Search Time, User Perceptions

INTRODUCTION

In recent years many commercial web search engines have augmented their search results with additional features such as business results, maps, image results, table-of-contents, query suggestions, shopping results, and page previews. The objective of these additional features, which we refer to as search enhancements, is to provide web searchers with richer and more relevant information than is provided by simple text results. The widespread adoption of such enhancements by search engine providers suggests that these features are perceived to be useful by many searchers.

A potential disadvantage of search enhancements is that they add complexity to the search results, possibly increasing the cognitive load on searchers [3, 8, 10]. Beaulieu concludes that the benefits of enhancements need to be balanced against the additional cognitive load that they impose upon searchers:

The effectiveness of search enhancements is influenced by many factors, including how they are presented, the richness of the data being searched, the number of terms in the query, and the design of the user interface [1].

This study examines how utilization of web search enhancements influences searcher perceptions of enhancement effectiveness and search time. Since previous studies have found that searcher perceptions can be affected by query complexity, two types of queries are utilized. The first are simple, factual queries. The second are more complex investigative queries that require the searcher to make comparisons and choices. The paper also explores the relationship between time spent conducting a search task and searcher perceptions of enhancement effectiveness.

LITERATURE REVIEW

A large body of research on search enhancements has revealed useful insights about their effectiveness and factors that influence user perceptions.

Early research into query expansion enhancements found that searchers did not utilize the enhancements [2]. However a later study found that search enhancements improved user search performance (as measured by the number of relevant documents found) by approximately 20% [6]. Similarly, White and Marchionini [9] studied the use of one enhancement, query expansion, and found that its usage led to higher quality queries and greater user engagement. Surprisingly, they also found that users were more satisfied with overall effectiveness and usability when the expansion enhancement was not available.

A more recent 2011 study utilizing commercial search engines found that perceptions of search engine effectiveness increased when enhancements were utilized. The study also found that perceptions decreased when more than one

enhancement was used on complex search tasks [7]. The second finding suggests that there is an interaction between search task complexity and user perceptions of search enhancements.

A potential disadvantage of search enhancements is that they can reduce the quality of search results. Dou, Song, and Wen [4] found that the effectiveness of inferring user intent is dependent upon the specificity of the query and the accuracy of the data used to make the inference. For an unambiguous query, the quality of inferred intent is higher when the inference is based upon large quantities of query data collected from many users. This allows the search engine to make high quality inferences even for queries that the searchers has not previously issued. However when the query is ambiguous and the searcher has an extensive search history, the quality of inferred intent is higher when the inference is based upon personal search history.

Most search research has examined either user perceptions or explicit measures of search behavior, such as number of queries conducted, number of results clicked, rank of results clicked, usage of back button, and time spent evaluating each result. Fox, Karnawat, Mydland, Dumais, and White [5] used an instrumented browser to examine the relationship between the two types of measures. The modified browser recorded the details of searcher behavior. At the conclusion of each search session searchers received a popup dialogue box that surveyed their perceptions of the search. The study found a high correlation between explicit measure and user perceptions. By utilizing a combination of implicit measures the authors were able to correctly predict user's perceptions of satisfaction 74% of the time. They also found that the longer users search through a particular result list, the less likely they are to be satisfied the search session.

RESEARCH MODEL

Figure 1 shows the research model. We propose that effectiveness of enhancements will be higher for factual tasks than for investigative tasks. We also propose that factual searches will be conducted faster than investigative searches. Finally, search time will influence effectiveness of enhancements, i.e., effectiveness of enhancements will be higher for faster searches. In general, factual searches will be conducted faster than investigative searches.

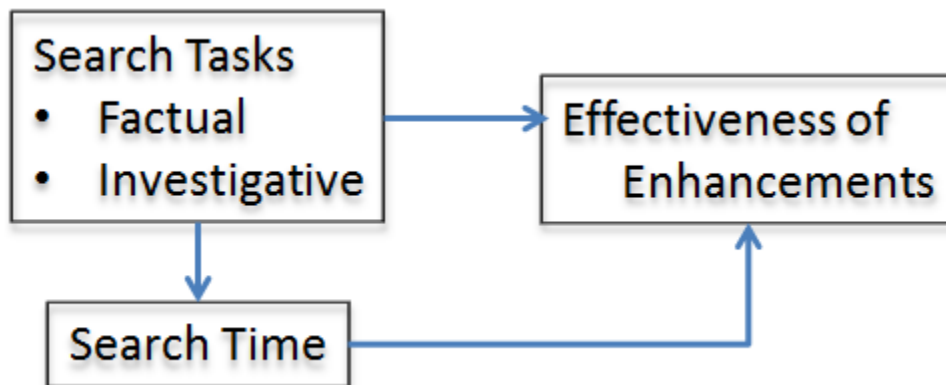


Figure 1: Research Model

METHODOLOGY

We designed an empirical study to validate the research model. Thirty five undergraduate business students enrolled in an introductory Management of Information Systems (MIS) class at a Pacific Northwest University participated in the empirical study. Each participant was required to conduct fourteen information-seeking tasks. The fourteen information-seeking tasks were formulated as questions. The participants completed their tasks using either the Bing or Google search engine. To avoid any sequence bias, the search tasks and the search engine were randomly ordered and assigned. We designed a custom web-based application to collect participant responses. The application also

Issues in Information Systems

Volume 13, Issue 2, pp. 164-169, 2012

captured the time elapsed to complete each information seeking task. The clock stopped as soon as the participant pressed a key to enter their response to the search task. Once the participant entered his/her response to the search task, he/she was required to identify the enhancements used (if any) and answer five questions that related to the effectiveness of the enhancements deployed. The process repeated itself until all the fourteen searches were completed. Study participants were then required to complete a post-study questionnaire that asked them about the positive and negative attributes of the Google and Bing search engines.

Measures

Given that the effectiveness of search perceptions can be influenced by search task complexity, we assigned each participant to conduct two types of search tasks: factual and investigative. Each participant conducted seven factual search tasks and seven investigative search tasks. The fourteen questions representing factual and investigative search tasks are listed in Table 1. The questions were pre-tested by the researchers to ensure that they were unambiguous and neither too easy nor too difficult for undergraduates to successfully answer.

Table 1. Search Task and Questions

Task Type	Question
Factual	What is tomorrow's local weather forecast?
	Who is the current Commerce Secretary?
	What is today's exchange rate between the Euro and the U.S. dollar? (\$ per euro)
	Which team won the World Series in 1952?
	Why did Eliot Spitzer resign?
	What is the phone number for Apex property management?
	Which movie won the Oscar for "best picture" in 2008?
Investigative	Find the lowest on-line price (including shipping) for a Kodak Easyshare C180 camera.
	Which university has the highest undergraduate enrollment: UCLA or Purdue?
	Find a microbrewery within walking distance of MIT (Massachusetts Institute of Technology).
	Name a major city along the most direct driving route between Seattle and Las Vegas.
	Which country has the second largest military budget?
	Which law school has higher median GPA for admitted applicants: UW or Seattle University?
	What is the lowest price you can find for a Nike rugby shoe?

To identify the enhancements deployed in the search tasks, we provided the study participants with a detailed checklist of enhancements available on the Google and Bing search engines. The participants were instructed to select the enhancements that they had utilized during each search task. The enhancements list included: standard text results, query suggestions, entity cards, image results, local business results, maps, news results, quick preview, related searches, search history, search options, site links, spelling suggestions, table of contents, and video results. Perceptions of effectiveness of enhancements were measured using five items. These included confidence (I am confident that I found the right answer), ease of use (enhancements were easy to use), relevancy (enhancements provided me with relevant information), quickness (enhancements helped me quickly find the answer), and satisfaction (overall I was satisfied with the enhancements). A five-point Likert type scale (strongly disagree, disagree, neutral, agree, strongly agree) was used for each measurement item.

Data Collection

Before collecting data, the broad purpose of the study and detailed instructions on how to complete the study were explained to the student participants. An extra credit for participating in the study was given to each student. The students conducted the searches in a computer lab located on the university campus. The students were allowed to

conduct the searches at any time during a five-day period. Experimenter influence was minimized by providing the participants with a familiar environment. Five designated computers with hard copies of instructions were used for the study. The instructions (also discussed in class) were included in the web application so that the participants could read them prior to logging into the web application. To ensure that the participants accurately identify the search enhancements that they utilized in their searches, we included hard copies of screen shots of the Google and Bing search results pages overlaid with red arrows and text that identified the search enhancements. Each student was also provided with a URL to a password protected web application. Once the students had logged into the study the web application directed them through an online version of the printed instructions, including screenshots identifying the various search enhancements. The application then randomly assigned search task and a randomly generated hyperlink to either the Google or Bing search provider. Clicking on the hyperlink opened a new tab and after conducting the search task the participants entered their answer into a textbox and checked all of the search enhancements that they utilized to answer the question. They then answered the search enhancement effectiveness questions. The application repeated itself till all the fourteen search tasks were completed. Extensive validation was deployed to assure that the participants completed each step before continuing to the next stage.

Data Analyses

A total of 488 searches were performed. Two hundred and thirty six of these were conducted using Bing and 252 using the Google search engine. Two hundred and forty two searches were conducted to respond to factual questions and 245 for investigative type of questions. Enhancements were used in 246 of the 488 searches. Only those searches that used enhancements were used in this research.

Prior to analyzing the data, stem-and-leaf plots were used to identify outliers for measured search time. Students could potentially become distracted during the search task, leading to inaccurate measured search task times. Twenty of the searches were identified as extremes and were discarded. Thus the remaining part of data analysis focuses on the remaining 226 searches where enhancements were used to respond to the factual and investigative questions.

Bartlett's test of sphericity was conducted for the presence of correlations amongst the five item measuring effectiveness of enhancements. The null hypothesis of variable independence was rejected ($p = 0.000$). To further assess the degree of intercorrelations between the variable items, the measure of sampling adequacy was found to be 0.91, suggesting that all the five items were appropriate for factor analysis. Principal component factor analysis was conducted on all the five items. The results are shown in Table 2. As expected, all the five items loaded on a single factor explaining 83.8% of variance. All the loadings were above 0.85. The descriptive statistics are shown in Table 3. Cronbach's alpha of 0.95 suggests high reliability of the measurement items. Due to large variance in search time, a natural logarithm of time was computed.

Table 2. Factor Analysis

Variable	Loadings
Effectiveness of Enhancements [Eigenvalue = 4.19]	
- Confidence	0.86
- Ease of use	0.91
- Relevancy	0.94
- Quickness	0.91
- Satisfaction	0.95

Table 3. Descriptive Statistics

Items and Variables	N	Mean	SD
Log_time	226	11.0922	.74822
Effectiveness of Enhancements [Alpha = 0.95]	226	3.3496	.80162
- Confidence	226	3.42	.831
- Ease of use	226	3.39	.827
- Relevance	226	3.31	.881
- Quickness	226	3.27	.973
- Satisfaction	226	3.34	.866

To test for search engine bias, we conducted independent sample t-tests between searchers that used Bing and those that used the Google search engine. No significant differences were observed in the effectiveness of enhancement item means between the two groups that utilized different search engines.

Next, we conducted t-tests to compare differences in means of effectiveness of enhancements and search time between factual and investigative search groups. The group statistics are shown in Table 4. Significant differences in means of search time and effectiveness of enhancements were detected between the factual and investigative search groups ($p \leq 0.005$). The results suggest that the mean of effectiveness of enhancements was significantly higher when factual searches were conducted than when investigative searches were conducted. Furthermore, factual searches were conducted significantly faster than investigative searches.

Table 4. Group Statistics

	Task Type	N	Mean	Std. Deviation	Std. Error Mean
Log_time***	F	113	10.7020	.69560	.06544
	I	113	11.4823	.57764	.05434
Effectiveness of enhancements***	F	113	3.5416	.71797	.06754
	I	113	3.1575	.83728	.07876

*** $p = 0.000$

To test for the relationship between search time and effectiveness of enhancements, we conducted correlation and regression analysis. A significant negative correlation was observed between search time and effectiveness of enhancements. The Pearson correlation coefficient was -0.322 and since we have proposed a direction relationship, the correlation was found to be significant at $p = 0.000$ (1-tailed). A simple linear regression was also conducted with search time as the independent variable and effectiveness of the enhancements as the dependent variable. The model (Standardized Beta = -0.322 and $T = -5.083$) was found to be significant at $p=0.000$. The adjusted R-square was 9.9%.

CONCLUSIONS

The empirical study of 226 search tasks found that web searchers spent significantly more time conducting investigative searches than factual searches. Searchers also perceived enhancements to be significantly more effective on factual searches than on investigative searches. There were no significant differences between the Google and Bing Search engines with respect to perceptions of effectiveness.

There are several possible explanations for why users perceive enhancements to be more effective for simple factual searches. Some of the enhancements may be targeted at simple factual searches and in many cases the enhancements help identify the answer. For instance, one of the factual queries utilized in the study was "What is tomorrow's weather forecast?" This query is made so frequently that both of the search engines utilized, Google and Bing, have special weather enhancements that provide the local weather forecast in a simple graphical format. The users do not even need to provide their location in the query.

On the other hand, enhancements may not be so useful for investigative queries, where the searcher must collect data from multiple sources for additional analysis. These enhancements may even distract the user by cluttering the screen or providing irrelevant (but eye-catching) content.

Further research into the interactions between types of enhancements and types of queries should be conducted to further identify best practices. For instance, when is a map useful, and when is it not? Does a link to recent news add to a better result, or one in which the searcher feels more certain?

REFERENCES

1. Beaulieu, M. "Experiments on Interfaces to Support Query Expansion," *Journal of Documentation*, (53:1), 1997, 8-19
2. Beaulieu, M., Payne, A., Do T., and Jones, S. The Impracticality of Relevance Feedback in Internet Search Engines - ENQUIRE Okapi Project Report 17. British Library Research and Innovation Center, Wetherby, West Yorkshire, UK 1996.
3. Bruza, P., McArthur, R., and Dennis, S. "Interactive Internet Search: Keyword, Directory and Query Reformulation Mechanisms Compared," Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in information Retrieval. SIGIR '00. ACM, New York, NY, 2000, 280-287.
4. Dou, Z., Song, R., and Wen, J-R, "A Large-scale Evaluation and Analysis of Personalized Search Strategies," WWW 2007, International World Wide Web Conference Committee, Banff, Alberta, Canada May 18-12, 2007.
5. Fox, S., Karnawat, K, Mydland, M. Dumais, S. and White, T. "Evaluation Implicit Measures to Improve Web Search," *ACM Transactions on Information Systems*, (23:2), 2005, 147-168.
6. Jansen, B. and McNeese M. "Evaluating the Effectiveness of and Patterns of Interactions with Automated Searching Assistance," *Journal of the American Society for Information Science and Technology*, (56:14), 2005, 1480-1503.
7. Sandvig, J. and D. Bajwa (2011) "User Perceptions of Search Enhancements in Web Search," *Journal of Computer Information Systems*, Winter 2011.
8. Schwartz, B. (2005) *The Paradox of Choice: Why More is Less*. Harper Perennial, New York.
9. White, R. and Marchionini, G. "A Study of Real-Time Query Expansion Effectiveness," Proceedings of the 29th Annual international ACM SIGIR Conference on Research and Development in information Retrieval, Seattle WA USA, 2006.
10. Wilson, Max L. and Schraefel, M.C. (2008) "Improving Exploratory Search Interfaces: Adding Value or Information Overload?" Second Workshop on Human-Computer Interaction and Information Retrieval, Redmond, WA, USA, 81-84.