Search Result Exploration: A Preliminary Study of Blind and Sighted Users' Decision Making and Performance

Melody Y. Ivory
The Information School
University of Washington, USA
myivory@u.washington.edu

Shiqing Yu
Encysys Consulting
Seattle, WA, USA
syu@encysys.com

Kathryn Gronemyer
The Information School
University of Washington, USA
ksg2202@u.washington.edu

Abstract
We conducted a preliminary study to examine sighted and blind users' decision-making behavior and performance during the search process. We manipulated the search result's relevance to a task, the search result presentation, and the effort required to process the corresponding web page. We found that users leveraged page features to gauge the amount of effort that is required to explore search pages and made exploration decisions accordingly. Users' desire to know additional page details varied based on their visual ability and the results' relevance. We quantified the cost/benefit tradeoff of additional page features and suggest ways to better support diverse Web searchers.

Categories & Subject Descriptors: H3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval - *search process, selection process*; H5.m. [**Information Interfaces and Presentation (e.g., HCI)**]: Miscellaneous.

General Terms: Design, Human Factors.

Keywords: WWW, search, user interface, user study, usability, accessibility, visual ability.

INTRODUCTION

Web searching is the primary means for locating information online. However, it can still be a complex and inefficient task [6]. Researchers have worked to improve search engine indexing and retrieval mechanisms, search result ranking, and query formulation. There have also been efforts to improve search result presentation. Our objective is to identify page features that could be presented in result displays and circumstances in which these features might help searchers to decide on exploring certain search results. We are also interested in how unexplored factors—users' visual or cognitive abilities—affect their use of search result displays. For instance, sighted users can scan and scroll around results or web pages, but blind users use screen readers to read results or pages linearly. Hence, there is a major gap between blind and sighted users' efficiency in completing online tasks.

We describe the Web search process and related work on improving it. We describe factors that influence users' exploration decisions and present our study and findings.

BACKGROUND AND RELATED WORK

Web searching is a complex process comprised of: (1) formulating a query for an information need, (2) inspecting search results to identify relevant results, and (3) exploring potentially relevant pages to locate desired information [3]. Users often revise queries and repeat these steps. According to information foraging theory [5], while searching for information, users attempt to optimize benefits (finding desired information) and minimize costs (effort to find desired information). We are interested in strategies that users employ to accomplish this objective. We examine: (1) the information consulted and time expended to make exploration decisions (step 2), and (2) the time expended or cost of exploring web pages (step 3). We also examine whether users consult additional page details like content or advertising volumes, if they are presented, and the cost/benefit tradeoff of doing so.

Researchers have examined various ways to help users to make exploration decisions (see [1] for a survey). Search result displays are augmented with contextual information (e.g., content categories or links to nearby pages) or thumbnail previews. Page summaries are generated dynamically based on search terms or keywords are highlighted within them. In general, the space of page features that could be useful to users, if presented or incorporated into result rankings, needs more exploration.

INFLUENCES ON USERS' EXPLORATION DECISIONS

Our literature survey revealed thirteen high-level factors that play a role in a user's decision to explore a search result. Factors address the search task (ST), search result (SR), corresponding web page (WP), and user (U).

- ST1: Type of search task (factoid, comparison, etc.)
- SR1: Features presented in the search result (e.g., URL)
- SR2: Quality of the presented features (descriptiveness, accuracy, etc.)
- SR3: Order in which the features are presented
- SR4: Formatting or visual display of the result
- SR5: Perceived relevance of the result
- SR6: Bias associated with the result (e.g., advertiser)
- SR7: Order in which the result is presented (ranking)
- WP1: Design of the corresponding web page (usability, accessibility, complexity, advertisement volume, etc.)
- WP2: Medium of the web page (HTML, image, etc.)
- U1: User's level of interest in the search task
- U2: User's ability (vision, search expertise, etc.)
- U3: User's technology (e.g., connection speed)

For this first study, we examined four factors that relate to SR1, SR5, WP1, and U2. We identified useful page features

CHI 2004, April 24–29, 2004, Vienna, Austria.
ACM 1-58113-703-6/04/0004.

(SR1) by informally asking subscribers to the Information School's Chat List to: (1) imagine that they were using a search engine to search for information and need to decide to explore a search result and (2) rank ten page features in terms of how important they thought they would be for making the decision (1-very important to 10-very unimportant). Features addressed the page's design (WP1) and consisted of the number of ads (textual, graphical, and pop-up), links, images, and words, download speed, page quality, bytes for the HTML file, and total bytes for the page. We analyzed twelve responses by computing median feature rankings. Page quality and the number of words were ranked as most important (median rankings of 1.5 and 3.5, respectively). Our study included these two features and the number of graphical ads (median ranking of 6.0); we included the latter feature, because respondents' rankings for the three types of ads suggested the number of ads on a page as an important feature.

EXPERIMENT

Study Objective

Our objective was to examine users' decision-making behavior on individual search results; however, we did not study how users' exploration decisions may be influenced by the entire set of results (e.g., by scanning the result list). We wanted to answer the following questions for both sighted and blind users.

- What information is used and desired to inform exploration decisions?
- How long does it take to make exploration decisions?
- How long does it take to explore web pages? What is the cost/benefit tradeoff of exploration decisions?
- What preferences do users have for incorporating additional page features into search result displays?

Study Factors and Levels

Search Result Relevance (SR5). A true positive (TP) is when the search result seems relevant to the task and the corresponding page contains the desired information. In a false positive (FP), the corresponding page does not contain the desired information, but the result seems relevant. We examine two levels for this factor—TP and FP.

Search Result Features (SR1). The control condition has the page's title, URL, and a summary. Three additional conditions contain the control information plus the number of graphical ads (ads), number of words (words), or estimated page quality (quality; i.e., usability, accessibility, etc.). For this preliminary study, the quality feature is based on our subjective judgment and an automated analysis of pages [4]. We examine four levels for this factor—control, ads, words, and quality.

Page Processing Effort (WP1). For simplicity, we estimated the effort required to process a page, based on the additional search result features. We used our automated web site evaluation work to identify quantitative ranges for each feature [4]. Low-effort pages contain 73–371 words, have no more than one graphical ad, and have good designs (i.e., they are highly usable and accessible). Medium-effort pages contain 219–517 words, have 2–3 graphical ads, and have mediocre designs. High-effort pages contain 710–1009 words, have more than four graphical ads, and have poor designs. A higher word count did not necessarily signify longer pages or additional scrolling, because most pages used multi-column layouts. We did not conduct user studies to test our assignment of pages to these categories. Although these categories simplify our study design, they are somewhat biased, for example, against information-dense sites in which well-designed pages contain a large amount of text but require relatively little effort to process. We use three levels for this factor—low, medium, and high.

Users' Visual Ability (U2). We examine two levels of visual ability: no visual impairment (sighted) and blindness which is counteracted by the use of a screen reader (blind).

Method

Search Tasks. Study factors resulted in a 2x2x3x4 full factorial design. Hence, we developed a partial factorial design based on the result relevance, result features, and page processing effort factors. We used factor and level analysis techniques to determine 18 experiment conditions (the minimum number required for study reliability) and distributed them into two experiments of nine conditions. We used random start order with rotation to determine nine task completion sequences for each experiment; a participant's task completion sequence was randomly chosen from the stored sequences. We counterbalanced participant assignment to the experiments; experiment completion was between subjects.

We created nine factoid search tasks, which include finding the "Freebird" ringtone for a cellphone and the weight of a full-grown Golden Retriever. For each task, we conducted a search with the Google search engine and identified both a TP and a FP result; participants completed only one version of each task. For each search result, we recorded the control information (title, URL, and summary), downloaded the corresponding web page, and used our automated evaluation tool to compute the additional features (ads, words, and quality). For some study pages, we had to modify their designs and the amounts of text to conform to specific effort levels.

Study Procedure. Participants completed study sessions within a testing lab. They used PCs with Windows XP and campus network connections. They used Internet Explorer for task completion; blind participants used the JAWS screen reader software [2]. Page caching was disabled. We used screen and audio capture software to record sessions.

During study sessions, participants provided demographic information (age, gender, search experience, etc.), discussed what they liked and disliked about search results, and described processes that they use to decide whether or not to explore a result. They completed two training tasks, followed by nine study tasks; we asked them to think aloud during task completion. Each task entailed evaluating both the search result and the corresponding web page by following five

steps. We developed a CGI script to go through the five steps for each task; each step was presented on a new web page.

1. User reads search task and clicks continue.
2. User reads search result and indicates whether or not she would explore the result (yes or no); the study script records the time that the user spends on this step. Figure 1 shows how results were presented to participants: the title was colored blue and underlined, but not active; the summary and URL were not emphasized; and the additional page feature was bolded and formatted with a larger font size than the size used for control information.
3. User completes first questionnaire to indicate the result features that influenced his exploration decision, rate the helpfulness of the result (7-point scale from very unhelpful to very helpful), and provide freeform comments.
4. User explores corresponding page and attempts to locate information to complete the task. We used the captured screen activity to estimate the time spent exploring pages.
5. User completes second questionnaire to indicate whether or not the page matched her expectations and contained the desired information. User also rates the page's quality (7-point scale from very low to very high), re-rates the search result's helpfulness, and indicates the page features that he would have liked to have seen before exploring the page.

Participants

Sixteen users (ten sighted and six blind) participated. Over two-thirds were age 18–25 and 56% were males. Most had some college education, eight years of computer experience, five years of Internet experience, and conducted Web searches multiple times daily. Equal numbers of participants from each user group completed the two experiments.

RESULTS

Influence of Search Result Features

Participants indicated mostly that they would explore results (92% on TP and 63% on FP tasks; F=8.045, p=0.005); visual ability had no significant effect. They used an average of two features to determine whether or not to explore a result. Feature use was as follows: summary (94%), title (64%), URL (42%), words (23%), quality (18%), and ads (16%). Results suggest that participants have consistent feature use patterns. For instance, one participant never used the title or URL, but used the summary in most cases. These tendencies could be supported by enabling users to control search result features.

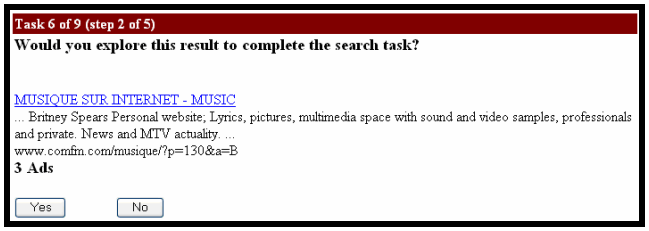

Figure 1. Search result that was presented to participants.

Table 1. Percentage of decisions to explore search results.

Rel.	Features				Sig.
	Control	Words	Ads	Quality	
TP	85.7%	95.7%	93.8%	85.5%	F=0.410, p=0.747
FP	87.5%	43.8%	68.8%	56.5%	F=2.572, p=0.061
Sig.	F=0.000, p=1.000	**F=19.217, p=0.000**	F=3.429, p=0.074	**F=4.526, p=0.040**	

Table 1 shows that participants were more likely to explore a result on FP tasks in control conditions than when additional features were presented. Hence, without additional page features, users may spend considerable time to explore search pages that do not contain the needed information. Participants were slightly more willing to explore search results on TP tasks when the number of words or ads was presented. Their written comments suggest that they used these features to estimate the amount of effort required to process a result and then made decisions accordingly. Half of their comments mentioned that the features had a major effect on their decision making. Example responses are below.

- "poor quality is the major thing in making my decision. If I haven't seen the 'poor quality,' I would use this page."
- "110 words, small page, won't take long to load so what the heck."
- "I like the fact that there is 0 ads."

After viewing search results and corresponding pages, participants reported for 63% of the search tasks that they would have wanted to know one or more of the following features upfront: download time (23%), number of links (21%), page quality (18%), number of ads (18%), images (17%), and words (18%), and file size (4%). They requested additional features mostly on FP tasks (72% vs. 54%; F=4.246, p=0.024), and sighted participants were most likely to request additional features (71% vs. 48%; F=8.141, p=0.005). Estimated page processing effort did not have a significant effect, but information needs varied across tasks. For instance, blind participants wanted to know the file size (10% vs. 1%; F=5.949, p=0.016), whereas sighted participants wanted to know the download time (29% vs. 12%; F=5.366, p=0.022) and number of words (26% vs. 5%; F=4.568, p=0.035). Given that screen readers can only process the contents of the HTML file, blind participants could use the file size to estimate the effort required to explore a page, whereas sighted users could use the download time to make this estimate. Results suggest the need for an efficient way to request or incorporate additional page details dynamically.

It could be argued that by displaying additional features, participants may adopt decision-making strategies that may prevent them from finding the needed information. A naturalistic study is needed to understand whether users prefer to avoid encounters with certain types of pages (e.g., pages with a lot of text, poor quality, or numerous ads) more so than finding needed information and potential implications.

Table 2. Mean search result and web page evaluation times (seconds). Standard deviations are in parentheses.

Visual Ability	Search Result	Web Page
Sighted	17.5 (11.0)	35.8 (23.4)
Blind	40.1 (17.3)	106.4 (49.8)
Sig.	**F=90.66, p=0.000**	**F=112.39, p=0.000**

Search Result and Web Page Evaluation Times

Table 2 shows that blind participants took over twice as long as sighted participants to explore a search result; result relevance and features had no significant effect. The table also shows that blind participants took about three times as long as sighted participants to look for information on web pages. Result relevance and estimated page processing effort also had a significant effect on evaluation time. Participants' web page evaluation time was mostly proportional to the estimated effort level; specifically, they spent less time on low effort pages and more time on medium and high effort pages. They spent the most time on FP pages, although the difference was not significant. For sighted participants, the cost/benefit tradeoff for the additional page features was: spending an extra second to explore an additional feature, but not spending 36 seconds to explore a page. The cost/benefit tradeoff for blind participants was an extra 7 seconds to explore an additional feature versus 106 seconds to explore the page. In both cases, the tradeoffs suggest that users' search efficiency can be affected considerably when they explore pages that require a lot of processing effort, especially when they do not satisfy their information needs.

Preferences for Incorporating Search Result Features

To characterize participants' "ideal search result displays," we asked them to indicate their desire to: (1) have additional page features presented, (2) control how page features are presented, (3) have results sorted based on criteria like usability, and (4) control how results are sorted. Blind and sighted participants preferred mostly to specify criteria for controlling search result sorting or ranking (4). The benefits of this approach include: users would not have to read extra information, they could process search result lists faster, and they could possibly find the desired information faster.

Sighted participants expressed interest in having additional information displayed about search pages (1), whereas blind participants expressed interest in controlling search result displays (2). No participants had a strong preference for having results sorted by usability, accessibility, etc. (3). Their written comments suggest that they need to know how a pages' usability or accessibility is assessed to determine the credibility of such assessments.

DISCUSSION AND CONCLUSIONS

Results suggest that initially participants used the page's summary, title, and URL to predict search result relevance. They then considered additional features (words, ads, and quality) to decide whether or not to explore the page, regardless of their relevance predictions. Their decisions seemed related to the amount of effort that they thought would be required for them to explore pages. For instance, participants were most likely to explore search pages on the FP tasks when additional page features were unavailable. Blind participants took twice as long as sighted participants to explore search results and three times as long to explore web pages. In most cases, participants expressed a desire for additional page features, which varied based on their visual ability and the result's relevance. The cost/benefit tradeoff of displaying additional features suggest that users' search efficiency can be improved considerably when they choose not to explore pages that may not satisfy their information needs. Participants mostly preferred to use additional page features to control result order or ranking.

Although we consider the study to be preliminary, it suggests that there are other ways in which we can potentially improve the user's search experience. For instance, users could benefit from having an easy, flexible, and efficient way to modify results to match their typical exploration strategies. More studies need to be conducted to better understand the effect of incorporating additional page features into result ranking or displays such that they do not interfere with users finding the information that they need. We will refine and repeat our study and conduct future experiments to examine the effects of other factors like the type of search task. We also plan to carry out a naturalistic study to examine whether users prefer to avoid certain types of pages more than they desire to find information to satisfy their need. We will also explore the development and use of a proxy-based system to augment search results with page features and allow diverse users to control result displays and sorting. We intend to leverage the SCONE architecture (www.scone.de) in the proxy.

ACKNOWLEDGMENTS

We thank Dave Hendry and WenChun Wang for help with the study's design, Judy Ramey for use of the LUTE lab, Sheryl Burgstahler for help recruiting participants and use of the JAWS software, and participants and reviewers.

REFERENCES

1. O. Drori, Display of Search Results in Google-Based Yahoo! vs. LCC&K Interfaces: A Comparison Study, Proceedings of the Informing Science Conference, 2003.
2. Freedom Scientific, Jaws for Windows Overview, 2003.
3. M. A. Hearst, User Interfaces and Visualization, Modern Information Retrieval, R. Baeza-Yates and B. Ribeiro-Neto, Eds.: New York: ACM Press, 1999, 257-323.
4. M. Y. Ivory, Characteristics of Web Site Designs: Reality vs. Recommendations, Proceedings of the 10th International Conference on Human-Computer Interaction, 2003.
5. P. Pirolli and S. Card, Information Foraging in Information Access Environments, Proceedings of the ACM CHI Conference, 1995.
6. C. Sherman, Why Search Engines Fail, in SearchDay, August 29, 2002.