Preliminary Evidence for Top-down and Bottom-up Processes in Web Search Navigation

Shu-Chieh Wu
San Jose State University
NASA Ames Research Center
Moffett Field, CA 94035 USA
scwu@mail.arc.nasa.gov

Craig S. Miller
School of Computer Science,
Telecommunications and
Information Systems
DePaul University
Chicago, IL 60657 USA
cmiller@cs.depaul.edu

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.
ACM 978-1-59593-642-4/07/0004.

Abstract
In current theories of web navigation, link evaluation has been treated primarily as a bottom-up process involving assessing the semantic distance between a search goal and a given link in the information source. In this paper we investigate whether link evaluation could be subject to top-down influence from knowledge of the information source. We measured fixation durations that occurred during link evaluation and found shorter durations in the search for easy goals. This preliminary finding suggests that for goals with category names readily retrievable from knowledge of the information source, search is likely aided by top-down influences.

Keywords
Web navigation, eye movements, scanning, search, menu organization

ACM Classification Keywords
H.5.2 Information Interfaces and Representation: User Interfaces. H.3.3 Information Search and Retrieval: Search processes, Selection processes

Introduction

Despite advances in query-based search engines that have dramatically improved the efficiency and accuracy of keyword search, search through navigation remains an indispensable method for locating unfamiliar information goals. Almost all popular search engines still retain directories of information which offer users the ability to locate a search goal without knowing its proper terminology or to browse information within the same category. How users search among provided categories and choose one that they believe will lead them to their goal is a question of both theoretical interests and practical implications on the design of efficient information architecture.

In this paper, we examine the process underlying a critical element of web search navigation – link evaluation. Almost all aspects of search behavior are determined in one way or another by the result of link evaluation, whether it is to proceed with a link or to backtrack from a dead-end page. Theories of web navigation have proposed various mechanisms for link evaluation and gained moderate success in predicting among a group of links which one will more likely be selected in actual user behavior. There has been less attention, however, on the mechanism underlying the evaluative process, per se, of a single link. This paper represents our preliminary attempt in empirically investigating this issue.

Related Work

Despite its central role in web search navigation, the mechanism underlying link evaluation is not always explicitly defined in current theories of web navigation. For example, in MESA (Method for Evaluating Site Architectures), developed by Miller and Remington, links on a webpage are assigned relevance scores [4]. MESA selects a link when it encounters a link with a relevance score above an internal threshold. Relevance scores are obtained using separate procedures and provided to the model. While the model can usefully predict actual search times by human users, primarily a model for navigation, MESA does not offer a processing account for individual link evaluation.

CoLiDeS (Comprehension-based Linked model of Deliberate Search), developed by Kitajima, Blackmon, and Polson, attempts to provide a more encompassing account of web navigation [2]. The model takes into account the physical layout of texts and objects on a webpage as well as the goal directed action planning process of the user. The relevance of a link is determined by its similarity to the goal, which is defined by their distance in a semantic space and determined in the model by Latent Semantic Analysis (LSA).

Similarly, SNIF-ACT (Scent-based Navigation and Information Foraging in the ACT architecture), developed by Fu and Pirolli, determines link relevance by the similarity of a link to the search goal [1]. Termed information scent, the link relevance measure is derived from the result of a spreading activation mechanism operating within a user’s linguistic knowledge.

Both CoLiDeS and SNIF-ACT model the mechanism underlying individual link evaluation as measuring the distance between the link and search goal in a semantic space. There is little doubt that link evaluation involves some form of comparison and for information goals the comparison naturally concerns semantics. The question is whether the comparison is really between the particular link being evaluated and the search goal, an assumption shared by both models.

There is a paradoxical aspect to web search that, although users seek information that they do not possess, they will not likely find the information if they do not possess any knowledge of the information they seek. The reality is that users often possess certain amount of knowledge not only of their search goal but also of the search paths available to them. For example, a shopper who wishes to purchase a television set from an online store would not expect to necessarily find the word television on the front page of the store. Rather, she or he would expect to find links like electronics or video equipments. In other words, to facilitate search users are likely to take into account what opportunities are available based on their prior experience and apply that knowledge to rephrase their search goals in ways conforming to what they believe to be conventional contents of the information database.

What is being suggested here is that link evaluation likely involves not only a bottom-up process in which the meaning of a countered link is compared to that of the search goal but also a top-down process in which a user rephrases the search goal in languages closer to those available from the information source. This idea is akin to the elaboration of goals in CoLiDeS but with an emphasis on elaboration targeted to match choices available from the information source. By reformatting the search goal according to the information source, the comparison process of link evaluation then turns into recognition, arguably the simplest form of comparison.

In this preliminary work, we investigate whether there is evidence for top-down processing in individual link evaluation. We hypothesize that the distinction between top-down and bottom-up processes is most likely revealed in the comparison between search for easy and hard goals. When the search goal is familiar and easy, users are more likely to apply their prior knowledge of the information source to reformat the search goal and transform the process of link evaluation into recognition. Conversely, when the search goal is unfamiliar and hard, users are more likely to depend on bottom-up processes and compare the search goal against each encountered link. The prediction is that individual link evaluation should take less time with easy than hard goals.

Research Approach

To identify easy and hard search goals, we used the Gini index, a measure of inequality with values between 0 and 1. The Gini index associated with a search goal was calculated by having prospective users classify the goal into the category that they believed best described it. A Gini index of 0 indicates that all users classify a given goal into the same category. A Gini index of 1 indicates that users lack consensus on which category a given goal should belong.

We believe that Gini index can be a reasonable measure of task difficulty. A lack of consensus across users suggests the presence of competing categories, even within a user. Competing categories could inhibit the user's ability to recall the most likely category and make the search more difficult than when the search goal can be readily associated with a known category.

To assess individual link evaluation time, we monitored the eye movements of participants doing a search task. It is widely known that duration of fixation reflects underlying processes [5]. Assuming easy search goals are associated with top-down processes, we predict that they will yield shorter fixation durations on individual link evaluations.

Empirical Study

Apparatus

The study was carried out on a Pentium 4 PC running Internet Explorer. Eye movements were monitored using a head-mounted high-speed eye tracker (Applied Sciences Laboratory, Model 501) with eye-head integration function, sampling at 120Hz. Gaze positions were then synchronized with recorded scenes using GazeTracker software (Eye Response Technology), which records video at 640x480 pixel resolution and samples at 40 frames per second.

Task and Design

The website used in the study was generated based on an “expert database” maintained by the Media Relation Department at DePaul University, which contains descriptions of 970 university faculty members and their respective areas of expertise as a resource for journalists in need for a subject-matter expert.

The expert database was implemented in a browser-based web application. The database has 9 top level categories, displayed in 3 columns and 3 rows. Each category is a link that links to subcategories of the respective category or expert descriptions. The construction of the website is described in Miller et al.[3]

The interface of the expert database featured a back button which returns to the previously displayed page. Participants were asked to only use navigation functions provided by the web application and not those by the browser. The web application recorded and time-stamped every selection performed by the participant.

In the study, participants searched for 16 experts of two task types:

- Exact description task: Participants were asked to locate a specific expert in the database.
- Scenario task: Participants were given a scenario and asked to identify an appropriate expert for the scenario in the database.

Each participant received the same 6 scenario task trials. Two of the 10 exact description task trials were pre-selected and given to all participants. The other 8 were randomly selected for each participant. The order of the trials was randomized for each participant.

Computation of the Gini Index

Fifteen volunteers classified all 970 expert descriptions into 9 categories. Details of the procedure are described in Miller et al. [3] The category distributions were used to calculate the Gini index for each expert description.

Procedure

Instructions were presented online prior to the session. The presentation of each trial target was accompanied by a “continue” button which – when pressed – displayed the top level menu of the database and started the timer. The target description was continuously viewable from the top region of the display during a trial. Each trial was terminated upon finding the target expert or after four minutes have elapsed. Then the next trial was presented.

Results

Two participants (CM and EO) recruited from local colleges participated. Both of them are experienced computer users familiar with web browsers. They had no prior experience with the expert database and were naïve to the purpose of the study.

Because there were no definite correct answers for the scenario task trials, the present analyses focused on the results from the exact description task trials (10 in total). Further, for consistency purposes, analyses were limited to visits (initial and revisits) to the top level menu (identical to all task trials) within the first four minutes.[1]

Gini Index and Task Difficulty

As a first step, we sought validation of using the Gini index as a difficulty measure. The top level menu on each trial was visited on average 2.5 times, with a range between 1 to 6. The number of top level visits may also reflect the difficult of the search goal, and indeed there were significant correlations between the Gini index and the number of top level visits for both participants (r = .55 and r = .61, df = 8, p < .05).

1 There existed an unknown glitch in the system so that the task did not terminate after four minutes even though the target expert had not been located. The analysis was restricted to the first four minutes of each task to reflect the intended task feature.

Fixation Duration

On the top level menu, we defined 10 non-overlapping areas of interest (AOI) which included one AOI for each of the 9 categories and one for the top area where the current trial scenario was displayed. Next, we identified fixations within each AOI. Fixations were defined as 3 or more sampled gaze points falling within an area of 60 pixels and with a total duration of at least 250 ms. When calculating fixation durations, successive fixations within the same AOI were combined, along with intervening saccade intervals. For the analysis, we included only fixations that occurred during the very first visit to the top level menu on each trial because

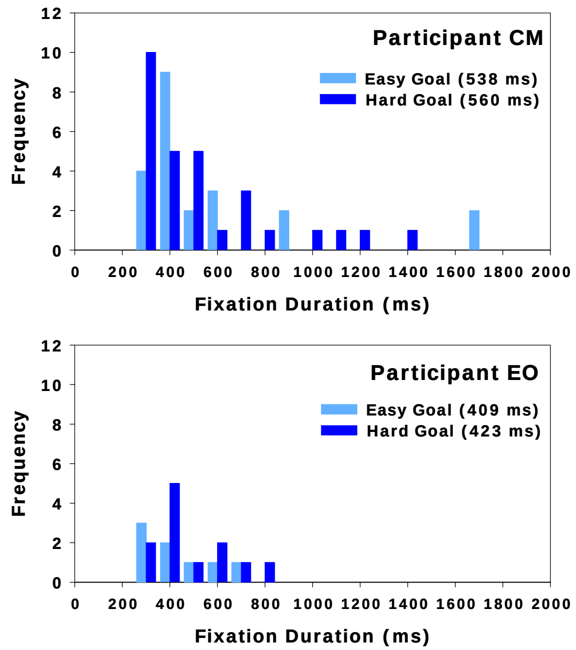

Figure 1. Fixation duration frequency distribution for easy and hard search goals

subsequent visits are subject to greater noises from other sources, such as frustration for having to backtrack from a dead end. We also excluded last fixations on the first top menu visits because their durations often include the time to make link selection.

Using these criteria, we collected 42 valid fixations from CM's results and 20 from EO's. We then divided these fixations based on the difficulty of the search goals into two categories: easy goal and hard goal. Both participants generated more fixations on trials with hard goals (30 by CM, 12 by EO). Figure 1 shows the frequency distributions of fixation durations, with mean fixation durations of the respective category listed in the legend. On average, fixation durations in the search for easy goals were 18 ms shorter than in the search for hard goals.

Discussion

The two participants we tested showed fixation duration patterns consistent with our prediction. Specifically, fixation durations were shorter in the search for easy goals, suggesting the possible use of top-down knowledge. Obviously these results are very preliminary and need to be replicated in a larger sample of participants.

There are alternative explanations for reduced fixation durations on easy goals. Some task trial descriptions used in the study contained category names, which could prime category recognition. Spreading activation could occur faster between familiar search goals and categories due to their close semantic distance; it could lead to faster link evaluation with no relation to top-down or bottom-up processing. Future work should address how we may distinguish between faster fixation durations as a consequence of priming and as consequence of efficient link recognition.

Understanding users' ability to employ more efficient top-down processes has important implications for the successful design of web sites and other menu-based systems. As users become more familiar with the information source, they may be better served with more selection items per page in order to reduce the number of user actions needed to reach their information goal.

Acknowledgements

This work was funded by grant NNA06CA99A from NASA's Human Research Program. We thank Eric Landrum, Shane Sestito, and Elizabeth Williams for assistance with data collection and analysis.

References

[1] Fu, W.-T., and Pirolli, P. A cognitive model of user navigation on the World Wide Web. *Human-Computer Interaction,* in press.

[2] Kitajima, M., Blackmon, M.H., and Polson, P.G. A comprehension-based model of web navigation and its application to web usability analysis. *Proc. CHI 2000,* ACM Press (2000), 357-373.

[3] Miller, C.S., Fuchs, S., Anantharaman, N.S., and Kulkarni, P. Evaluating category membership for information architecture. DePaul CTI Technical Report (2007).

[4] Miller, C.S., and Remington, R.W. Modeling information navigation: Implications for information architecture. *Human-Computer Interaction, 19* (2004), 225-271.

Salthouse, T. A., and Ellis, C. L. Determinants of eye-fixation duration. *Am J of Psych, 93* (1980), 207-234.