Predicting Outcomes of Web Navigation

Jacek Gwizdka Ian Spence
Department of Psychology
University of Toronto
Toronto, Canada
+1 416 946 5813 +1 416 978 7623
www2005@gwizdka.com spence@psych.utoronto.ca

ABSTRACT
Two exploratory studies examined the relationships among web navigation metrics, measures of lostness, and success on web navigation tasks. The web metrics were based on counts of visits to web pages, properties of the web usage graph, and similarity to an optimal path. Metrics based on similarity to an optimal path were good predictors of lostness and task success.

Categories and Subject Descriptors
H.1.2 [**Models & Principles**]: User/Machine Systems — Human factors; H.5.4 Hypertext/Hypermedia – Navigation; User issues.

Keywords
Web navigation; Lostness; Path similarity; Compactness; Stratum.

1. INTRODUCTION
Quantitative assessment of navigational behavior is fundamental to understand the phenomenon of lostness in web navigation. While anecdotes and informal observations may be suggestive, quantitative measures of user behavior will provide a better characterization of user navigation and this will, in turn, suggest better ways of designing the structure of web sites.

We present two exploratory studies that examined relationships between web navigation patterns, lostness, and task success.

2. RELATED WORK
The properties of web-usage graphs have been correlated with user task outcomes. McEneaney [3] demonstrated that learning task success was correlated with shallow hierarchical navigation (high compactness), while task failure was related to a linear style of navigation (high stratum). It is natural to enquire whether these results hold for all types of search tasks.

Pitkow & Pirolli [7] used the longest repeated sequence algorithm to predict user surfing behavior and Wang & Zaïane [1] employed a sequence alignment algorithm to cluster user web navigation sessions. We used a similar algorithm to assess similarity between the user navigation path and the optimal navigation path.

Otter and Johnson [6] suggested that to capture lostness in hypertext, a battery of measures was needed. Herder [2] supported this view. He found that user disorientation in hypermedia could be characterized by combining an index of revisitation behavior with median page view times. Can other measures also predict lostness and can they do so with a wide variety of search tasks?

WWW 2005, May 10-14, 2005, Chiba, Japan.
ACM 1-59593-051-5/05/0005.

3. METHODOLOGY

3.1 User Studies
Participants in two studies performed question-driven information finding on large Canadian government websites. Goal-directed finding of information is one of the most common information seeking tasks on the Web (accounting for 25% [4]). Participants were asked to find a single web page containing information that was specified in each of several questions. Sample question: "Find addresses of passport offices in Ontario". Participants were asked to navigate to the page in the same web browser window, without using a search engine.

In the first study (TA—Talk Aloud), fourteen adults (6 females & 8 males) were tested separately. Each participant answered ten questions, which were the same for all participants. They were asked to talk aloud while they were navigating. There was no time limit on finding an answer to each of ten questions. All sessions were recorded using screen capture software.

In the second study (TL—Time Limit), forty eight adults (29 females & 19 males) were tested separately. Each participant answered eight questions, which were the same for all participants. In contrast to the TA study, the time allowed for each question was limited to three minutes. If the requested information was not found in that time, the participant moved on to the next question.

3.2 Measures

3.2.1 Simple navigation metrics
Simple metrics, such as the number of web pages visited in a session, the number of unique web pages visited, the time spent on each web page, and the total time on each question were obtained directly from the web session logs.

3.2.2 Properties of the user web navigation graph
The nodes of the graph (see example Figure 1) are the individual pages traversed. Two measures of the shape of the graph were calculated: *stratum* and *compactness* [1]. *Compactness* refers to the connectedness of the graph and varies between zero and one; it is close to zero for sparsely linked graphs and close to one for highly connected graphs. *Stratum* measures how close the navigation path is to a linear ordering. Like compactness, stratum varies between zero and one. A stratum close to zero indicates a less linear navigation path; a stratum close to one indicates a more nearly linear navigation path.

3.2.3 Similarity to the optimal path
We defined the optimal path as the shortest path to the page containing the desired information. We assumed that such a path existed for each question and that the path was unique. Two

similarity measures (*LCSMax*, *LCSlenMax)* were calculated based on a procedure by Needleman and Wunsch [5], which uses a global sequence alignment algorithm with a non-zero gap cost and an arbitrary distance function. LCSMax is a normalized similarity score between the user path and the optimal path. LCSlenMax is the ratio of the length of the longest common subsequence (in the user and optimal paths) to the user path length.

3.3 Navigation Task Outcomes

Task success was defined as finding a page with the information specified in the question. In the TA study, task success was evaluated subjectively by each participant and verified by the experimenter after the session. In the TL study, task success was inferred if the task was completed within the time limit. In both cases, task success was assigned a binary value (true or false). *Lostness* was operationalized as the participant's feeling of disorientation on the web navigation task. This could be assessed in the TA study only, where participants were asked to talk aloud during the sessions which were screen-cam recorded. A trained rater watched the audio-video record and assessed the apparent lostness of the participant on a 4-point scale.

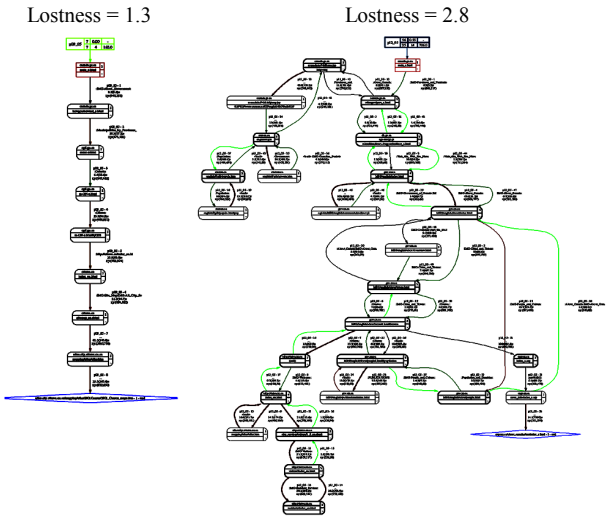

Figure 1. Different shapes of web navigation graphs on the same task

4. RESULTS

Using linear regression, LCSlenMax and total time were found to the best predictors of subjective lostness, accounting for over 90% of the total variance in the fitted regression model.

Using logistic regression in both studies, LCSMax was found to be the best predictor of *task success*. This confirmed our *a priori* intuitions: the higher the similarity to the optimal path, the better the chances for success on an information finding task.

In the TL study, significant models were fit for five of the eight questions. Three models confirmed LCSMax as the best predictor of task success. Compactness or stratum were the best predictors of task success in the other two other models. Lower values of compactness (sparsely linked web usage graphs corresponding to fewer returns to previously visited pages) and higher values of stratum (more linear user navigation path) were related to task success. Although LCSMax was, in general, the best single predictor of task success, linearity of the search path proved to be a better predictor on some questions.

5. DISCUSSION

Similarity to the optimal path is a good predictor of both lostness and task success, for the navigational tasks that we employed. Tasks other than goal-directed search may yield different results. Depending on the task, the notion of an optimal path might not make sense, or an optimal path might not exist, and therefore similarity metrics would be ill-defined. Success may depend on the user finding one of a set of "good" paths and such situations may be better characterized by considering the "shape" (e.g., compactness and stratum) of the exploration path.

Higher values of stratum (typically along with lower values of compactness) tend to be associated with a higher probability of task success. This finding is opposite to McEneaney [3]. However, there is no necessary contradiction; McEneaney used a different navigational task (learning using a hypertext handbook) which likely requires a rather different navigation strategy. The viability of a search strategy is probably contingent on the information-seeking task.

6. CONCLUSIONS & FUTURE WORK

Appropriate metrics can provide useful characterizations of user web navigation behavior and can diagnose a variety of problems (such as getting lost, and not finding the desired information). The ability to predict task outcomes with precision would be extremely useful in practice. More needs to be done to define and to select the best metrics. Effective diagnostics could be of great assistance in building adaptive web solutions.

We plan to investigate metrics for a wider variety of information seeking tasks, for example, on tasks such as broad browsing. We plan to determine which factors induce deviation from the optimal path, by examination of the local website structure at points in the search path where users frequently become lost.

7. REFERENCES

[1] Botafogo, R.A., Rivlin, E., & Shneiderman, B., Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. ACM Transactions on Information Systems, 1992. 10(2): 142-180.

[2] Herder, E., & Van Dijk, B. Site Structure and User Navigation: Models, Measures and Methods. In Adaptable and Adaptive Hypermedia Systems, S.Y. Chen and G.D. Magoulas (eds), Idea Group Publishing, 19-34, 2004.

[3] McEneaney, J.E. Graphic and numerical methods to assess navigation in hypertext. In Intl. Journal of Human-Computer Studies, 55, pages 761-786, 2001.

[4] Morrison, J.B., Pirolli, P., and Card, S. K., A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions, In the Proceedings of CHI'2001. Extended abstracts. 163 – 164. 2001.

[5] Needleman S.B., Wunsch C.D. A general method applicable to the search for similarities in the amino acid sequences of two proteins. J. Mol. Biol. 48:443-453, 1970.

[6] Otter, M. and Johnson, H. Lost in hyperspace: metrics and mental models. Interacting with Computers, 13(1), 1-40, 2000.

[7] Pitkow, J. E., Pirolli, P. Mining longest repeated subsequences to predict World Wide Web surfing. Second USENIX Symposium on Internet Technologies and Systems. 1999.

[8] Wang, W. and Zaïane, O. R. Clustering Web Sessions by Sequence Alignment. 394-398. 2002. Proceedings of DEXA Workshops 2002.