

Indirect Assessment of Web Navigation Success

Jacek Gwizdka

University of Toronto
Toronto, ON, Canada
chi2005@gwizdka.com
+1 416 946 5813

Ian Spence

University of Toronto
Toronto, ON, Canada
spence@psych.utoronto.ca
+1 416 978 7623

Abstract

Despite much research on hypertext and web navigation, relatively little is known about the relationship between web navigation strategies and success. We present two exploratory studies designed to explore the relationships between several web navigation metrics that are based on similarity to an optimal path to predict task success. The data suggest that the relationships between these measures depend on the particular web navigation task.

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

General Terms: Experimentation; Measurement.

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

INTRODUCTION

Navigating large, complex, web-based collections of pages places a demanding cognitive burden on users; many users experience difficulty in forming mental models that will allow them to complete navigation tasks successfully. Despite the long history of research on hypertext and web navigation, relatively little is known about the relationship between web navigation strategies and success on information seeking tasks. A constructive first step in gaining an improved understanding of the characteristics of success and failure in user navigation is to explore the properties of measures that capture the diverse factors that underlie the navigation behaviors of users. User web sessions may then be characterized by relating these measures to navigational task outcomes. We present the results of two exploratory studies designed to examine relationships between web navigation patterns and task success.

RELATED WORK

Web navigation paths may be represented by graphs, where the nodes represent web pages, and the edges represent the links traversed. Graph properties such as *compactness* and *stratum* [1] correlate with user task outcomes. Learning task

success, according to McEneaney [4], was associated with shallow hierarchical navigation paths (high compactness), while task failure was related to linear paths (high stratum). Pei-Chun Shih et al. [7] used web-based courseware to study navigation behavior and found that the navigation graphs of people with greater prior experience with web-based instructional tools was more linear and disperse (high stratum & low compactness). Do similar relationships between graph-based measures and navigation task success hold for all types of search tasks?

A series of web pages visited by a user can be represented as a sequence of symbols. The similarity between web pages can be based on their URLs, position in the page hierarchy, text content, or layout, among other properties. Similarity between entire sequences can then be calculated using longest common sequence or gene sequence matching algorithms. For example, Wang & Zaïane [9] used a sequence alignment algorithm to cluster user web navigation sessions. We use a similar algorithm to assess similarity between the user navigation path and the optimal navigation path.

RESEARCH OBJECTIVES

Our goal is to establish which quantitative measures of navigational behavior are the best predictors of success in information seeking tasks. Our approach is to build predictive models using empirical data collected from web navigation sessions using large, natural web sites.

METHODOLOGY

Navigation Task

We examine goal-directed information seeking, since these are among the most common information tasks on the Web, accounting for 25% of web search purposes, according to Morrison et al. [5]. Examples of goal-directed tasks include searching for a specific fact, a specific document, or information about a specific product.

User Studies

Participants in two studies performed question-driven information seeking tasks using large Canadian government websites. The studies were conducted in a university laboratory using a PC running the Microsoft Windows 2000 operating system. Each participant was given a starting

page (a Government of Canada home page or a Health Canada home page. In both cases the page was in English). Participants were asked to find a single web page containing information that was specified in each of a number of questions. Participants were asked to navigate in a single browser window, without using a search engine. URLs of all pages visited were logged to a file, with timestamps, and the coordinates of the link, or the button, or the graphic, that was clicked.

Samples of the information goals were:

- “Find information on the history of the West Nile virus”
- “Find a short description of Ottawa (including its population and area covered)”

Time-limit Web Navigation Study (TL study)

Forty eight adults (29 females & 19 males) took part in individual sessions. Each participant was asked to complete eight search tasks. The tasks were the same for all participants and were asked in the same order. In contrast to the TA study, the time allowed for each task was limited to three minutes. If the requested information was not found in that time, the participant stopped and moved on to the next task.

Talk-aloud Web Navigation Study (TA study)

Fourteen adults (6 females & 8 males) took part in individual sessions. Each participant was asked to complete ten search tasks. The tasks were the same for all participants and were asked in the same order. Participants were required to talk aloud during the search. There was no time limit. The sessions were recorded using the Camtasia software for capturing computer screens (along with ambient sound).

MEASURES

A number of measures characterizing user navigation were computed. Some measures were derived directly from the raw navigation data, while others were obtained from more complex calculations.

Simple navigation metrics

Simple metrics, such as, the number of web pages visited in a session (N), the number of unique web pages visited (U), time spent on each web page (Time_per_page), and the total time on each question (Total_time) were obtained directly from the web session logs. In addition, a simple derived metric was calculated from the ratio of visited and optimal node counts: $\text{Revisits} = 1 - U/N$ [2].

Properties of the web navigation graph

Stratum and *compactness* [1] were calculated: these provide quantitative measures of the shape of the web navigation graph. Compactness refers to the connectedness of a graph and it 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. Stratum varies

between zero and one with values closer to zero indicating a less linear navigation path, and values closer to one indicating a more nearly linear path.

Similarity to the optimal path

The third set of metrics measured similarity between the user path and the optimal path. The optimal path is the shortest route to the web page containing the sought for information. Thus we made the assumption that such a path exists for each information finding task and that this path is unique.

Two similarity measures based on a well-known procedure by Needleman and Wunsch [6] were calculated. This method uses a global sequence alignment algorithm with a non-zero gap cost and an arbitrary distance function. The algorithm was applied based on the following assumptions:

- The user navigation path is a sequence of web pages, referred to as nodes;
- Each node is uniquely identified by a URL, and treated as 3-tuple: $\langle \text{host}, \langle \text{path}, \langle \text{query} \rangle \rangle$;
- The distance function between sequences is calculated based on similarity between URLs treated as 3-tuples.

Two metrics were derived from the N-W algorithm: *LCSMax*, *LCSlenMax*. *LCSMax* is a similarity score obtained from the N-W algorithm for a given user path and the optimal path. The score was normalized to the maximum possible score (an ideal match) for a path length equal to the user path length. *LCSlenMax* is the ratio of the length of the longest common subsequence (LCS) to the user path length (N). The LCS is calculated between a given user path and the optimal path. If there were no gaps in the common subsequence allowed, *LCSlenMax* would correspond to the ratio of the number of all visited pages which are on the optimal navigation path to the user path length. Since the N-W algorithm has a non-zero gap cost, the common subsequence may contain pages which appear only on the optimal path or on the user path, thus interpretation of *LCSlenMax* is less simple.

Navigation Task Outcomes

Task success was the main dependent measure in these studies; it was defined as finding a web page containing information specified in a question. Task success was measured differently in each study.

In the TL study, task success was inferred from the time spent on each question. All sessions where participants spent more than three minutes and ten seconds on one question were considered to be a failure to find an answer to a question (the time limit to answer each question was three minutes and we allowed a ten second period of grace). In the TA study, task success was evaluated subjectively after the session. The evaluation was based on the participant's own assessment of success. This subjective judgment was further verified by the experimenter, who

checked the content of the final webpage visited by the participant. In both cases, task success was coded as a binary outcome (true or false).

RESULTS

The first step in the data analysis was intended to provide a better understanding of the commonalities and differences among the navigational metrics. Based on these results, we built regression models to predict navigation task success.

Second Order Navigation Factors

As some of the first-order navigation metrics were related (by definition), second-order navigation metrics were calculated.

Rotated Factor Pattern			
Variable	Factor1	Factor2	Factor3
Total time	0.85	0.41	0.15
U (unique pages)	0.80	0.04	-0.49
N (total pages)	0.73	0.39	-0.47
LCSMax	-0.83	-0.28	0.01
LCSlenMax	-0.84	-0.35	0.12
Compact	0.19	0.96	-0.02
Revisits	0.43	0.79	-0.14
Stratum	-0.29	-0.90	0.12
Time_per_page	-0.06	-0.09	0.96

Table 1. Principal component analysis results – variable loadings on three factors (after varimax rotation).

Principal component analysis with varimax rotation was applied to the data from the TL study (Table 1). Three factors were extracted and interpreted as:

1. *Effort and navigational inefficiency*: characterized by a high number of pages visited, more time spent on the task, low task success, low similarity to the optimal path. This factor accounted for 45% of the variance.
2. *“Laborious” navigation*: characterized by high compactness, high proportion of revisited web pages, and low stratum – low linearity of the user path. This factor bears some similarity to the same-named factor identified by Juvina & van Oostendorp [3] (we have borrowed their name for this factor). The factor accounted for 36% of the variance.
3. *Navigation speed*: only one component loaded highly (.96) on this factor – average time per click.

Prediction of Task Success

Logistic regression was used to predict task success. LCSMax was found to be the best predictor of task success in both the TL study and the TA study (Table 2). This result confirmed our *a priori* intuitions: the higher the similarity

to an optimal path, the better the chances for success on an information finding task.

Study	R ²	% agree	Wald χ^2	p
TL	.30	81.3	62.2	<.001
	Best predictor: LCSMax		32.6	<.0001
TA	.28	86.8	12.8	.005
	Best predictor: LCSMax		10.3	.0013

Table 2. Regression models and best predictors of task success for study TL and TA.

For the TL study¹, we also examined whether task success was best predicted by LCSMax in each of the individual information-seeking tasks. Significant regression models were obtained for five out of eight questions. Three of those models confirmed LCSMax to be the best predictor of task success. In two models, however, Compactness or Stratum turned out to be the best predictors. In the first case, lower values of compactness (sparsely linked web usage graphs corresponding to fewer returns to previously visited pages) were related to higher task success. In the second case, higher values of Stratum (a more linear user navigation path) were related to higher task success. Although LCSMax was, in general, the best single predictor of task success, on particular questions measures of graph compactness or linearity proved to be better predictors.

The two main second-order factors established by principal component analysis (“Effort & inefficiency” and “Laboriousness”) were significant predictors of task success (Table 3). For the TL study, lower “Effort & inefficiency” and lower “Laboriousness” are associated with a higher probability of success.

R ²	% agree	Wald χ^2	p
.42	89.4	64.6	<.0001
Best predictor: Effort & inefficiency		61.4	<.0001
Best predictor: Laboriousness		50.1	<.0001

Table 3. Second order factors as predictors of task success.

DISCUSSION

Our results suggest that lower values of compactness and higher values of stratum tend to be associated with a higher probability of task success. This finding is opposite to the one reported by McEneaney [4]. However, there is no necessary contradiction. McEneaney used a quite different navigational task (learning from a hypertext handbook vs. goal-directed information finding). It seems that the navigation strategies which prove to be successful can be quite different in the two situations. The variation in the power of the same predictors, depending on the particular

¹ In the TA study, there was not enough data to perform a similar analysis.

question, further supports the conclusion that the success of a search strategy is highly dependent on the information-seeking task. Hence, while stratum and compactness measures may be good predictors in some situations, they are not necessarily as effective in predicting task success with all types of navigational tasks.

We showed that similarity to the optimal path is a good predictor of success in our tasks. However, these results should be considered in the context of our goal-directed navigational tasks. For tasks other than goal-directed information seeking the results may be different. Depending on the task, the notion of an optimal path may not make sense, or an optimal path may not exist, and therefore the use of similarity metrics would be problematic. Success on some tasks may depend on the user finding one of a set of “good” linear paths while success on other tasks may be associated with exploration paths of quite different “shape”.

CONCLUSIONS AND FUTURE WORK

Appropriate metrics can provide useful characterizations of user web navigation behavior and can help to diagnose a variety of problems (such as getting lost, and not finding the desired information) that users encounter when navigating hypertext documents. The ability to predict task success with reasonable precision would be extremely useful in practice. An effective diagnostic capability could be used to help build adaptive web solutions.

Our basic goal in this research is to find ways of improving the navigation experience by suggesting new design principles for information architectures to assist users to find the information that they are seeking without feeling lost. Appropriate metrics can help identify the local web structures that are conducive to successful navigation and will also be useful in identifying those structures that tend to inhibit the progress of the navigation.

In future research we plan to explore a wider variety of information seeking tasks, for example, characterizing metrics on tasks such as broad browsing. We also plan to determine which influences encourage deviation from the optimal or intended path, by local examination of the website structure at points in the search path where users frequently become lost.

ACKNOWLEDGMENTS

This research was supported by CITO.

REFERENCES

1. Botafogo, R. A., Rivlin, E., and Shneiderman, B., "Structural analysis of hypertexts: Identifying hierarchies and useful metrics," *ACM Transactions on Information Systems*, vol. 10, no. 2, pp. 142-180, 1992.
2. Herder, E. and van Dijk, B., "Site Structure and User Navigation: Models, Measures and Methods," in S.Y.Chen and G.D.Magoulas (eds.) *Adaptable and Adaptive Hypermedia Systems* Idea Group Publishing, 2004, pp. 19-34.
3. Juvina, I. and van Oostendorp, H., "Individual Differences and Behavioral Aspects Involved in Modeling Web Navigation," *LECTURE NOTES IN COMPUTER SCIENCE*, vol. 3196 pp. 77-95, 2004.
4. McEneaney, J. E., "Graphic and numerical methods to assess navigation in hypertext," *Intl. Journal of Human Computer Studies*, vol. 55, no. 5, pp. 761-786, 2001.
5. Morrison, J., Pirolli, P., and Card, S. K. A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. 161-162. 2001. *Proc. of CHI' 2001*. Extended abstracts.
6. Needleman, S. B. and Wunsch, C. D., "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of molecular biology*, vol. 48, no. 3, pp. 443-453, 1970.
7. Shih, P.-C., Sanchez, F., and Mate, R. Quantifying user-navigation patterns: a methodology proposal. 2004. Poster presented at the 28th International Congress of Psychology, Beijing.
8. Smith, P. A., "Towards a practical measure of hypertext usability," *Interacting with computers*, vol. 8, no. 4, pp. 365-381, 1996.
9. Wang, W. and Zaïane, O. R. Clustering Web Sessions by Sequence Alignment. 394-398. 2002. *Proc. DEXA Workshops*.