Quantifying Navigation PATHS

Characterizing User Navigation in Hypermedia

Jacek Gwizdka, Ian Spence, Harumi Takeshita, & Ken Seergobin

Department of Psychology, University of Toronto
(jacek@psych.utoronto.ca ; spence@psych.utoronto.ca)

Introduction

Motivation

Despite many research efforts in the area of web navigation, relatively little is known about the relationship between web navigation strategies and success on information seeking tasks. The first step in gaining a better understanding of user navigation is to investigate measures that capture different aspects of user behaviour.

The goal of the research presented in this poster is to review, evaluate, and select appropriate metrics to characterize user navigation strategies in typical information seeking tasks on large websites.

Definitions of Metrics

Working Assumption

There exists a unique, optimal navigation path that leads to the target webpage.

Simple Metrics

This set of metrics is based on the ratio of visited and optimal node counts to characterize user navigation behaviour.

Lostness = $\sqrt{(U/N-1)^2 + (O/U-1)^2}$ (Smith 1996)

Revisits = $1 - U/N$

where: N – number of nodes visited; U – number of unique nodes visited;
O – number of nodes on the optimal path.

Graph-Based Metrics

The second set of metrics uses two formal properties of the graph that describes the user's navigation path.

Stratum = a measure of linearity (Botafogo et al. 1992)
Compactness = a measure of connectedness

Navigation Path Similarity Metrics

The third set of metrics measures similarity between user path and the optimal path.

The navigation path is treated as a sequence of nodes: $<n_0, n_1, ..., n_i>$
Each node is identified by a URL and treated as 3-tuple: <host>,<path>,<query>

Similarity measures were calculated using two well-known algorithms:
1) Levenstein distance (**LD**) = smallest number of string edits required to match two strings
2) Longest common sequence (**LCS**) that uses the Needleman-Wunsch global sequence alignment algorithm with a non-zero gap cost and an arbitrary distance function. The distance function was calculated based on similarity between URLs treated as 3-tuples (see above). (Needelman & Wunsch 1970)

Summary

Key Findings

- A number of metrics (Lostness, Stratum, Similarity with optimal path) were shown to be strongly associated with search time.
- More complex and computationally expensive metrics (e.g. LCS similarity) do not necessarily provide different (or better) information than very simple metrics (e.g. lostness).
- We have demonstrated the usefulness of constructing graphical visualizations of user navigation paths and the relationship of these visualizations to a variety of numerical metrics that characterize the navigation path.

Summary

Appropriate metrics can provide useful characterizations of user web navigation behaviour and can help to diagnose a variety of problems that users encounter when navigating hypertext documents. This diagnostic capability could be used, for example, to build adaptive web solutions. Extensions of this research include characterizing metrics on other tasks (i.e. on broad browsing). The ultimate goal is to inform web design and to improve the information architecture and information design of large, complex websites and hypertext documents.

Metrics in Practice

The Web Navigation Study

Context: Question-driven information seeking on government websites.
Participants: 14 adults (8 males, 6 females)
User task: find one web page containing information specified in each question.
Task constraints: navigate to the result, do not use search.
Sample question: "Find passport offices in Ontario".
In addition to the navigation metrics, the following measures were collected: time, time per click, task success (objective & subjective).

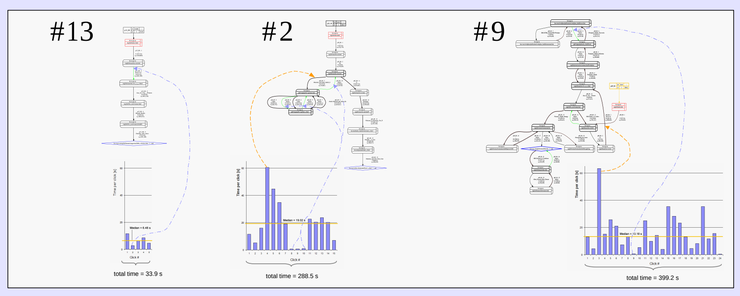
Characterizing One Task

User Task, Q8: "Find page describing how to deal with stress for women"

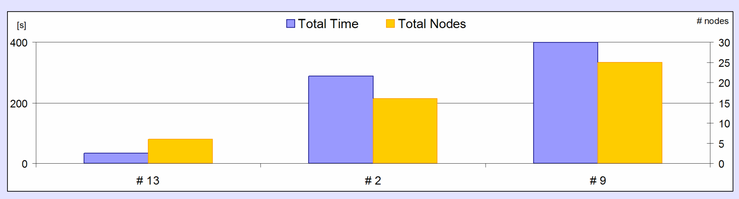
Subject Number

User Navigation Path
from linear to "bushy"

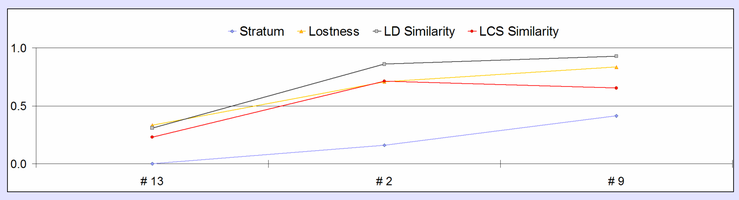
Time per Click
slowest & fastest clicks are linked to appropriate spots on user paths



Total Time & Total Nodes



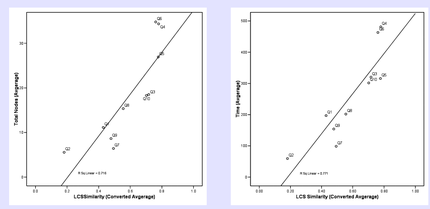
Stratum*, Lostness, & Similarity*
* to facilitate comparisons, Stratum and Similarity were rescaled to [0, 1]



Relationships Between Metrics

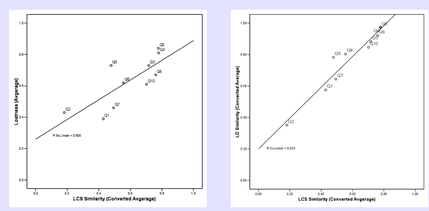
LCS Similarity and Task Performance

User paths similar to the optimal path tended to be associated with faster performance and smaller number of clicks.



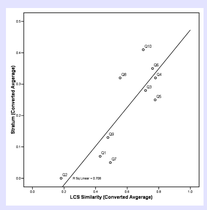
LCS and LD Similarity, and Lostness

LCS similarity was highly related to LD Similarity and related to Lostness.



LCS Similarity and Stratum

High LCS Similarity tended to correspond to more linear user navigation graphs.



Subjective Task Success

When participants felt that they had been successful in the information-seeking task, their navigation paths were close to linear and similar to the optimal path.