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.

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.

\[
\begin{align*}
\text{Lostness} & = \sqrt{(U/N-1)^2 + (O/U-1)^2} \\
\text{Revisits} & = 1 - U/N \\
\end{align*}
\]

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.

\[
\begin{align*}
\text{Stratum} & = \text{a measure of linearity} \\
\text{Compactness} & = \text{a measure of connectedness} \\
\end{align*}
\]

(Botafogo et al. 1992)

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: \(n_1, n_2, \ldots, n_x\)

Each node is identified by a URL and treated as a 3-tuple: \(<\text{host}>, <\text{path}>, <\text{query}>\)

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).

(LCS similarity was highly related to LD similarity and related to Lostness. (Needelman & Wunsch 1970)

Metrics in Practice
User Task, Q8: “Find page describing how to deal with stress for women”

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\(^2\), 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.