Abstract

Gemma O'Rourke
MAI Computer Engineering 2018
The Search For The Searcher: Understanding User Behaviour through Search Log Analysis
Supervisors: Prof. Seamus Lawless and Dr. Annalina Caputo

When users search online, it is important to understand what they are looking for and how they go about looking for it. This project analyses search engine logs to gain insight into why users behave the way they do. In particular, how does user behaviour change within a search session, as the user works towards a particular information need? And based on that behaviour, can search logs provide sufficient information to indicate how much a user already knows about that topic?

The aim of this research is to investigate how useful search logs can be for understanding and assessing users. The research question is as follows: 'to what extent can the analysis of search logs provide insight into how an individual user develops their knowledge of a topic within a search session?'

To provide answers to this question, users are randomly sampled from the AOL search logs dataset, based on the number of search sessions they have. Within a session, a query complexity score is calculated based on term specificity and reading difficulty. The topic of the query is found, and the depth and overlap of the topic with other topics in the session is evaluated.

Query complexity is found have a tendency to increase within a search session and has a positive correlation with topic depth. Topic depth and overlap also tend to increase within a search session. This indicates that the user is increasing the complexity of their queries towards a specific information need.

These results indicate that search logs can provide useful insights into user behaviour and their knowledge development within a session. It is possible that future work could be able to categorise users into different knowledge levels based on their search behaviour. There is also scope for developing more sophisticated methods of calculating query complexity and analysing topics.