

The Search for the Searcher: Features Influencing the Knowledge Gain by the User

Yash Mundra

MAI in Computer Engineering

Supervisors: Prof. Seamus Lawless and Dr. Annalina Caputo

April 25, 2019

Abstract

The digital presence and accessibility made the internet connection viable to all irrespective of ages. The purpose of exploring internet is the need for information. People use search engine for getting knowledge in a particular area or new topic. It may even be for upgrading their knowledge on a specific topic. They used to search by posing some questions in the search engine and these queries are informational queries. Though the search machine presents many results, there is no appropriate quantification about the gaining of knowledge by the searchers in a particular search session. The search engine can be tailor made once this factor is known. This paper intends to design such prediction model and aims to analyse the impact of features on the users' knowledge gain using search log analysis. The feature selection for selecting only the imperative features are performed using three models of regression. Two different sets of features are considered for analysis to have a comparative analysis and to ensure the effect of features on the user's knowledge gain by an efficient prediction model. The set 1 will carry more features that are considered important and have impact on the model and the other set will carry the features selected by an appropriate regression model. These set of features are fed to various models of classification and the performance of models are evaluated using various performance indices. The prediction models are premeditated to have three different labels (low, medium, and high) with regard to the knowledge gained by the user. The model is validated using openly available crowdsourcing dataset. Majority of the designed models show promising results.