A survey of statistical approaches for query expansion

Muhammad Ahsan Raza[1] · Rahmah Mokhtar[1] · Noraziah Ahmad[1]

[1]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Kuantan, Malaysia

Abstract

A major issue in effective information retrieval is the problem of vocabulary mismatches. The method called query expansion addresses this issue by reformulating each search query with additional terms that better define the information needs of the user. Many researchers have contributed to improving the accuracy of information retrieval systems, through different approaches to query expansion. In this article, we primarily discuss statistical query expansion approaches that include document analysis, search and browse log analyses, and web knowledge analyses. In addition to proposing a comprehensive classification for these approaches, we also briefly analyse the pros and cons of each technique. Finally, we evaluate these techniques using five functional features and experimental settings such as TREC collection and results of performance metrics. An in-depth survey of different statistical query expansion approaches suggests that the selection of the best approach depends on the type of search query, the nature and availability of data resources, and performance efficiency requirements.

Keywords Information retrieval · Statistical approaches · Query expansion · Document analysis · Query log analysis