A Dataset for Evaluating Query Suggestion Algorithms in Information Retrieval

Ioan Bădărînză
Department of Computer Science
Babeş-Bolyai University
Cluj-Napoca, Romania
ionutb@cs.ubbcluj.ro

Adrian Sterca
Department of Computer Science
Babeş-Bolyai University
Cluj-Napoca, Romania
forest@cs.ubbcluj.ro

Darius Bufnea
Department of Computer Science
Babeş-Bolyai University
Cluj-Napoca, Romania
bufny@cs.ubbcluj.ro

Abstract—This paper presents a dataset that can be used for evaluating query suggestion algorithms in textual information retrieval. The dataset is public and offered free of charge to the information retrieval research community. The data was gathered in an experiment that lasted more than 2 months and to which participated a number of 119 users, mainly faculty students. The dataset contains web browsing history and query history (submitted to the Google search engine) from all these users. The data is indexed in a database and downloadable in a database dump format. The dataset is very useful for evaluating general query suggestion algorithms by themselves (in a standalone manner) or against Google's MPC query suggestion algorithm. At the same time, the dataset supports building and testing personalized query suggestion algorithms that consider the user context/profile when computing query suggestions.

Index Terms—dataset, query suggestion, information retrieval, search engine