Retrieval Efficiency of Normalized Query Expansion

Sofia Stamou and Dimitris Christodoulakis

Computer Engineering and Informatics Department, Patras University
Research Academic Computer Technology Institute, 26221, Greece
{stamou, dxri}@cti.gr

Abstract. In this paper we experimentally study the impact of normalized query expansion on Web Information Retrieval. In this respect, we have implemented a query expansion module, which firstly normalizes the user submitted queries and subsequently attempts to enrich them with semantically related terms that are obtained from WordNet. Experimental results demonstrate that for certain query types our module has a potential in giving improved search results in terms of relevance, compared to the results retrieved for the same queries by other retrieval methods.

1 Introduction

To support information seekers in overcoming terminological problems when searching for information on the Web, several approaches have been addressed in the literature, the most prominent of which imply the expansion of the issued queries with semantically related terms. In this paper we seek to get an improved insight on how normalized query expansion can effectively cope with vocabulary mismatches, in an attempt to improve retrieval relevance when querying the Web. To challenge that, we built a prototype query expansion module that interacts with normalization techniques, applied to the subjected queries. The query expansion module we introduce explores the semantic information encoded in WordNet and determines which terms are the most suitable to be used for enriching a given query.

We evaluated our system's performance in successfully enriching queries by having humans judge the relevance of the results retrieved in response to a set of queries after these have been expanded by our system, and compare them to the relevance of the results retrieved for the same queries after employing other searching techniques. Our findings indicate that our expansion approach improves retrieval performance for certain query types, compared to the performance of other retrieval techniques. Retrieval improvements are pronounced for long queries (i.e. multiword and phrase queries) due to our system's effectiveness in disambiguating them.

In the remaining of this paper, we describe how our system proceeds in generating expanded queries (Section 2) and we report on our empirical findings (Section 3) that proof the impact of our approach in helping information seekers find alternative wordings for expressing their information needs. We conclude our work (Section 4) with a discussion on our system's contribution in retrieval performance. Before pre-