Chapter 54
Arabic Query Expansion Using WordNet and Association Rules

Ahmed Abbache
University of Oran 1, Ahmed Ben Bella, Algeria

Farid Meziane
University of Salford, UK

Ghalem Belalem
University of Oran 1, Ahmed Ben Bella, Algeria

Fatma Zohra Belkredim
University Hassiba Ben Bouali of Chlef, Algeria

ABSTRACT

Query expansion is the process of adding additional relevant terms to the original queries to improve the performance of information retrieval systems. However, previous studies showed that automatic query expansion using WordNet do not lead to an improvement in the performance. One of the main challenges of query expansion is the selection of appropriate terms. In this paper, the authors review this problem using Arabic WordNet and Association Rules within the context of Arabic Language. The results obtained confirmed that with an appropriate selection method, the authors are able to exploit Arabic WordNet to improve the retrieval performance. Their empirical results on a sub-corpus from the Xinhua collection showed that their automatic selection method has achieved a significant performance improvement in terms of MAP and recall and a better precision with the first top retrieved documents.

INTRODUCTION AND MOTIVATION

Information Retrieval (IR) is concerned with organizing, storing, retrieving and displaying information. IR systems aim to provide the user with an easy access to the information he/she is interested in. Usually the user is required to formulate his information need through a query; the IR system then provides the user with the relevant information in return (Baeza-Yates & Ribeiro-Neto, 1999). While interacting with the users, IR systems face many challenges, one of these being the vocabulary problem also referred to

as vocabulary mismatch (Carpineto & Romano, 2012). To address this issue, researcher in the IR field proposed many solutions and the latest being the Automatic Query Expansion (AQE). This technique aims at reformulating the original query by adding new terms into it to achieve a better accuracy for the IR system. Various AQE techniques have been proposed and Cui et al. (2002) split them into two major classes: global analysis and local analysis.

The global analysis approaches are independent from the initial query or its result. Generally, they use external knowledge sources to select terms for expansion such as thesaurus or WordNet. Local analysis approaches formulate a new query on the basis of some retrieved documents of a previous search, for example relevance feedback (Bilel et al, 2011).

Adding new terms to the initial Query can take place prior to either the initial search or the relevance-feedback search (Cuna et al., 1992). The selection of these terms is a key phase in the IR process. There are several sources for terms selection, WordNet has been recognized as an important source of selection for query expansion. It is one of the largest and most widely used in the tasks of natural language processing (NLP), counting Word Sense Disambiguation (WSD) and Question Answering Systems (QAS) (Tingting et al., 1992).

Arabic is a vocalized language. It requires the adding of signs to the consonants to precisely define the pronunciation of a word. Hence, the non-vocalized Arabic word may have several possible meanings. Unfortunately, texts in Arabic languages, mainly Modern Standard Arabic (MSA) are not vocalized. For example, the non-vocalized word: (عالم) may mean by way of its vocalization: scientist (عَالِم) or world (عَالَم). This phenomenon makes the selection of appropriate synonyms for expansion more difficult in Arabic, a problem that is not faced by other languages.

Association rules have been used in several areas, including clustering and IR (Picariello & Rinaldi, 2007; Veeramalai & Kannan, 2011). In AQE, they have been used to provide semantic links between terms. In a previous study (Abbache et al., 2014), we have shown that AQE does not improve retrieval; but on the other hand, Interactive Query Expansion (IQE) improves retrieval. We concluded that if we can find a way to select appropriate terms from the Arabic WordNet instead of taking all the returned terms, we may achieve better results. In this study, we investigate the possibility of using association rules between terms based on the assumption that words in documents that associate with a word in the query are more likely to be related to that query word.

The remaining of the paper is organized as follows. Section two summarizes some related and similar work highlighting specifically the methodology used and the results obtained; Section three attempts to highlight the source of terms selection (Arabic WordNet). Section four presents the proposed technique for automatic query expansion. Section five describes the experiments and section six summarizes the conclusions.

RELATED WORK

Various methods and techniques have been proposed or used for AQE and these have been extensively studied in IR. Query expansion is the process of adding additional terms to the original query to improve the IR systems' performance. AQE is not new; in fact, it has been mentioned in earlier 1960 but has not reached maturity until very recently (Carpineto & Romano, 2012).

Related Content

Cloud4NFICA-Nearness Factor-Based Incremental Clustering Algorithm Using Microsoft Azure for the Analysis of Intelligent Meter Data
Archana Yashodip Chaudhariand Preeti Mulay (2020). *International Journal of Information Retrieval Research (pp. 21-39).*
www.irma-international.org/article/cloud4nfica-nearness-factor-based-incremental-clustering-algorithm-using-microsoft-azure-for-the-analysis-of-intelligent-meter-data/249699

Theory of Cognitive Constructivism
Kijpokin Kasemsap (2015). *Information Seeking Behavior and Technology Adoption: Theories and Trends (pp. 1-25).*
www.irma-international.org/chapter/theory-of-cognitive-constructivism/127119

Theory of Planned Behavior (TPB) Ajzen (1988)
Azzah Al Maskari (2015). *Information Seeking Behavior and Technology Adoption: Theories and Trends (pp. 237-260).*
www.irma-international.org/chapter/theory-of-planned-behavior-tpb-ajzen-1988/127135

Innovative Product Design using Metaontology with Semantic TRIZ
Koswatte R. C. Koswatte, Incheon Paik, Wonhee Parkand Banage T. G. S. Kumara (2015). *International Journal of Information Retrieval Research (pp. 43-65).*
www.irma-international.org/article/innovative-product-design-using-metaontology-with-semantic-triz/130007

The Importance of Authoritative URI Design Schemes for Open Government Data
Alexei Bulazel, Dominic DiFranzo, John S. Ericksonand James A. Hendler (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications (pp. 2181-2199).*
www.irma-international.org/chapter/the-importance-of-authoritative-uri-design-schemes-for-open-government-data/198642