

## Chapter 40

# The Use of Arabic WordNet in Arabic Information Retrieval

**Ahmed Abbache**

*University of Oran 1, Algeria*

**Fatma Zohra Belkredim**

*University Hassiba Ben bouali of Chlef, Algeria*

**Fatiha Barigou**

*University of Oran 1, Algeria*

**Ghalem Belalem**

*University of Oran 1, Algeria*

### ABSTRACT

*Research and experimentation using Arabic WordNet in the field of information retrieval are relatively new. It is limited compared to the research that has been done using Princeton WordNet. This work attempts to study the impact of Arabic WordNet on the performance of Arabic information retrieval. We extend Lucene with Arabic WordNet to expand user's queries. The major contribution of this study is to propose an interactive query expansion (IQE) methodology using the word's part-of-speech, according to the part it plays in a query. First, the user selects the appropriate part of speech for each term in the original query, and then he reselects the appropriate synonyms. Experimental results show that our IQE strategy produces a good Mean Average Precision (MAP), it is able to improve MAP by 12.6%, but no variant of automatic query expansion (AQE) strategies did. Nevertheless, the experiments allow us to conclude that with an appropriate use of Arabic WordNet as a source of linguistic information for AQE can improve effectiveness for Arabic information retrieval.*

### INTRODUCTION

The amount of electronic information is increasing, so Information Retrieval (IR) should provide the user with easy access to the information in which he is interested. The user should formulate his information need into a query, usually a set of keywords that summarizes the description of his information need. Given the user query, the

key goal of an IR system is to locate information that is relevant to a user's query (Baeza-Yates & Ribeiro-Neto, 1999).

When searching information, users use different terms to describe a similar concept or need. Vocabulary differences have created difficulties for Information Retrieval Systems (IRS) for decades; this is called the vocabulary problem. One well-known method to solve this problem is the

DOI: 10.4018/978-1-4666-9562-7.ch040

automatic query expansion (Voorhees, 1994); it aims to generate a new query called the expanded query that contains not only the terms of the user's query but also the ones relevant to the query.

Cui et al. (2002) classified query expansion techniques into two main groups: global analysis and local analysis.

Through global methods, query expansion works independently from the initial query and the results returned from it. It is depending on find word relationships from all the documents in the corpus, however, they are relevant to the query or not. Or use external knowledge sources to select terms for expansion. For example, query expansion with a thesaurus or WordNet.

Unlike Global methods, Local methods use documents that are retrieved using the initial query. And expansion terms are selected from them, like relevance Feedback.

Although most of the research has focused on the use of English WordNet to improve the effectiveness in information retrieval, and other WordNet like EuroWordNet for several European languages. There have been a few of work and effort to improve effectiveness in Arabic information retrieval systems (Abderrahim, 2013; Brahmi, 2012). Research and experimentation using Arabic WordNet (AWN) in the field of information retrieval are relatively new, and limited compared to the research that has been done using English WordNet, which has been used in the field of information retrieval for a long while.

Our contribution is guided by the lack that we observed around the use of Arabic WordNet in the field of information retrieval. In this paper, we will focus on automatic/interactive query expansion for Arabic Information Retrieval; the expansion method proposed here is a global expansion technique based on Arabic WordNet. In this study, we propose to use the word's part-of-speech tagging, according to the part it plays in a query. This tagging will allow the system to provide the user with an appropriate list of synonyms. So first, the user must select the adequate part of speech for each

term in the original query, and then according to this choice he reselects the appropriate synonyms from a list of suggestions proposed by the system.

This paper is organized as follows: Section two illustrates an overview of the similar work, highlighting the methodology and the results founded. Section three presents the proposed technique for automatic and interactive query expansion. Section four describes the experiments themselves, and section five summarizes the conclusions.

## **RELATED WORK**

The existing work on Arabic WordNet can be classified mainly into two classes: AWN development and/or enrichment and AWN exploitation.

In this paper, we considered the researches which exploit AWN in the field of information retrieval. Other researches that aimed to the development and/or enrichment of the current AWN are not included. Despite the very little efforts that exploit AWN in: Text Categorization, Semantic Web Applications, Question/Answering, Lexical Semantic Annotation and Information Retrieval. This section presents and analyses the few numbers of studies that have used AWN in information retrieval.

Since summarization is needed the most in the process of searching for information on the web, Imam et al. (2013) introduced a query based single document Summarizer for Arabic text; by using AWN (to provide them with the synonyms of lexical words) and a knowledge base (represented by topic related concepts/keywords and the lexical relations among them). The user's query is expanded once by using an AWN thesaurus and then by adding the domain specific knowledge base to the expansion. To test performance, Essex Arabic Summaries Corpus was used. Experiments have shown that using their extracted knowledge base enhance performance than using only AWN.

Arabic contents exist in many forms on the web – HTML, word documents, XML, PDF, etc.

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/the-use-of-arabic-wordnet-in-arabic-information-retrieval/142650](http://www.igi-global.com/chapter/the-use-of-arabic-wordnet-in-arabic-information-retrieval/142650)

## Related Content

---

### Business Intelligence as a Service: A Vendor's Approach

Marco Spruit and Tim de Boer (2014). *International Journal of Business Intelligence Research* (pp. 26-43).

[www.irma-international.org/article/business-intelligence-as-a-service/126896](http://www.irma-international.org/article/business-intelligence-as-a-service/126896)

### Big Data Technologies and Analytics: A Review of Emerging Solutions

Hoda Ahmed Abdelhafez (2014). *International Journal of Business Analytics* (pp. 1-17).

[www.irma-international.org/article/big-data-technologies-and-analytics/115517](http://www.irma-international.org/article/big-data-technologies-and-analytics/115517)

### Real-Time Data Quality Monitoring System for Data Cleansing

Cihan Varol and Henry Neumann (2012). *International Journal of Business Intelligence Research* (pp. 83-93).

[www.irma-international.org/article/real-time-data-quality-monitoring/62024](http://www.irma-international.org/article/real-time-data-quality-monitoring/62024)

### Predicting Wastewater BOD Levels with Neural Network Time Series Models

David West and Scott Dellana (2004). *Neural Networks in Business Forecasting* (pp. 102-120).

[www.irma-international.org/chapter/predicting-wastewater-bod-levels-neural-network/27246](http://www.irma-international.org/chapter/predicting-wastewater-bod-levels-neural-network/27246)

### Knowledge Analytics: A Constituent of Educational Analytics

Shouhong Wang and Hai Wang (2020). *International Journal of Business Analytics* (pp. 14-23).

[www.irma-international.org/article/knowledge-analytics/264260](http://www.irma-international.org/article/knowledge-analytics/264260)