AN ACCURACY-ENHANCED STEMMING ALGORITHM FOR ARABIC INFORMATION RETRIEVAL

*Sadik Bessou, Mohamed Touahria**

Abstract: This paper provides a method for indexing and retrieving Arabic texts, based on natural language processing. Our approach exploits the notion of template in word stemming and replaces the words by their stems. This technique has proven to be effective since it has returned significant relevant retrieval results by decreasing silence during the retrieval phase. Series of experiments have been conducted to test the performance of the proposed algorithm ESAIR (Enhanced Stemmer for Arabic Information Retrieval). The results obtained indicate that the algorithm extracts the exact root with an accuracy rate up to 96% and hence, improving information retrieval.

Key words: *Arabic morphological analysis, stemming, information retrieval, machine translation*

Received: January 23, 2013 **DOI:** 10.14311/NNW.2014.24.007
Revised and accepted: February 13, 2014

1. Introduction

The accessible electronic documents in websites constitute a field of documentary research that is extensively growing [1]. According to the web father, these documents are intended to be decoded by humans rather than being data that can be automatically analysed [2]. The challenge is to automatically extract the information contained in these documents which are written in natural language, since "the power of the natural language creates an obstacle to its use for data processing" [3]. Nowadays, various languages are successfully processed. However, indexing Arabic language documents remains a big challenge towards its integration in the information technology, given its power and its wealth.

Automatic indexing of Arabic documents raises major problems [4, 5]:

- the problem of ambiguity caused by the absence of vowels, which requires complex morphological rules [6]; and
- the problem of inflected forms recognition, because Arabic derivational morphology is productive [7].

*Sadik Bessou, Mohamed Touahria, Department of Computer Science, University of Sétif 1, Sétif, 19000, Algeria, E-mail: bessou.s@gmail.com, touahria_momo@yahoo.fr

