

Abstract

This paper describes a rule-based technique for analyzing the morphology of Arabic words. The proposed 'Morphological Analyzer' processes the input word in order to determine its lexical form. The lexical form of the majority of Arabic words consists of a root and a morphological pattern. The analyzer applies a set of predefined rules in order to analyze the morphology of Arabic words as they appear in real text. It is able to recognize diacriticized, undiacriticized or partially diacriticized Arabic words generated from N-letter roots. In order to determine the possible meanings of a word, the Morphological Analyzer also provides some useful attributes of the word such as its type, gender, tense and number. The proposed Morphological Analyzer is a general-purpose technique that can be integrated into larger scale systems such as automatic translation applications, text summarization applications, text correction applications, web search engines, automatic vowelization of Arabic text applications and other natural language processing applications.