Abstract

This thesis offers a new and effective contribution to the analysis and classification of Arabic texts (poems). Despite the great difficulties in this analysis and classification of Arabic texts, there are many ramifications and limitations based on the complex rules of the Arabic language. Additionally, such studies require the researcher to master the rules, (for example, "Al Aurde" which was invented by Al-Khalil bin Ahmed Al Farahidi) upon which the poem structure is based. This theory is based on several concepts like "Tafila" and "Buhur". This thesis helps in the analysis and classification of the large and increasing amount of Arabic texts (poems), especially taking into consideration the widespread adoption of the Internet which requires finding new ways to analyze and classify texts to shorten the time and cost spent by users to find and understand new information. This study used natural language processing (NLP) and the theory of "Al Aurde" to create a novel automatic way of finding the rhythm of Arabic poetry. This method opens the way for several studies, in addition to finding the rhythm of Arabic texts. For example, it can be also used to compare different types of Arabic poems to extract their features. This study also proposed a new method of finding and classifying the rhyme of Arabic poetry by employing the theory of "Al Aurde" and natural language processing. Rhyme is considered as a complementary part of the rhythm (the ringtone of Arabic poems), both give the Arabic poem a special musical tone and influence on the listeners.

Studying the Arabic language and finding the characteristics of Arabic poems are difficult for English speakers. In this thesis, we implemented a system for analyzing classical Arabic poetry for teaching purposes. Several NLP technologies are used to create this system which also employs the theory of "Al Aurde". Using this system, students can be taught on the automatic way of finding most of the features of classical Arabic poems, like the case of the letters and the percentages for each short vowels in the poem. This system is then used to compare different types of classical Arabic poems and to find the properties of each type. In this study we tried to cover the difficulties of the Arabic language and were able to implement the method for the classification of modern Arabic poetry using Machine Learning (ML). In this study, we employed 3 ML algorithms (NB, SVM, and LSVC) for the classification task. We built and used our own dataset because of the lack of a free dataset for such studies for the Arabic language. Several steps of data pre-processing were performed to increase the accuracy of the classification. From the results" comparison, the most appropriate algorithm for the task. It was also found that machine learning is a good tool for the classification of the modern Arabic poetry.