Abstract

Today many experts in the field of information technology have been designing and developing algorithms to solve stemming problems, especially in Arabic. But, from many stemming analysis in Arabic, there is no standardization of good stemming algorithm in analyzing the accuracy of the text in the Quran. The construction of stemming in the Quran is an important work because it supports the classification of *sharaf* in the Quran to understand the meaning of every word in the Quran. One stemmer or stemming algorithm to find the basic form of a word in Arabic is the Khoja Stemmer algorithm. The workings of Khoja Stemmer is to try to search root in a word of Arabic by removing the longest prefix and longest suffix of a word, then trying to determine the root of the remaining word using the root dictionary. In this research, the built of Khoja Stemmer is able to calculate the average stemming in the Quran of 95.295%. However, the roots produced by Khoja Stemmer are still found some errors when manually checked. Thus, it takes a dictionary of the Quran to analyze every result of stemming done by Khoja Stemmer in order to stemming the Quran.

Keywords: quran, stemming, khoja stemmer, sharaf, arabic