Abstract

Statistically the Koran has 77845 words. So we need a system that can facilitate someone to look for the vocabulary of the Arabic-scripted Arabic with input in the Latin form. In addition, the large number of existing population causes large-scale caselpelling. So that a word search system is needed in the Koran that is tolerant of variations in writing and writing errors. Phonetic coding method is based on the rules of reading the Koran (Tajweed) to match words in Arabic and query in Latin form. Trigram index is used to match queries with datasets. For the suggestion system, several manuals are used to complete completing and edit distances for differences between queries and the results of rules. The result is that the phonetic search system produces MAP 79,16% and Recall produces 79,16%. In the suggestion system the MAP value is 57,87% and the recall is 58,14%

Keywords: Al-Quran, Trigram, Edit Distance, Tolerant Retrieval, Spelling Correction