Abstract

This research proposes a method to help correct typo errors, such as wrong letters, missing letters, and additional letters. To date, the Lafzi search system is an example of an effective application for searching Arabic queries based on sound similarity. The Lafzi application helps correct typos due to the sound of Arabic letters that are almost the same as the pronunciation but not correct errors due to typos. Typos could prevent the search system not to display the desired results. This research proposes a solution by employing auto-complete to equipped missing trigram and the edit distance metric to calculate the differentiation value between the corrected query with the initial query. The way the system works is by separating and sorting trigram tokens from queries (user inputs) based on the verse. Each verse that has a missing trigram token will be equipped and re-transformed into a corrected query. Each corrected query will be compared to the edit distance value against the initial query (input from the user), then a corrected query will be taken which has the smallest edit distance value and will be made as a suggested query. The evaluation shows that the proposed method produces the highest recall value at 93.40% and the highest MAP value at 86%. This outperforms the previous Lafzi system approach which achieves recall at 85.23% and MAP at 79.83%.