

Abstract

One way to understand the Quran is to do interpretations correctly and not deviate, is by paying attention to the word editor that used in composing verses of the Quran. In this way, another verse can be found that has a similar set of words. One way to search for the same verse based on these words is to use the Longest Common Subsequence (LCS) approach which can find the longest shared word sequence of a text pair. The results of the same verse search are then collected to become a corpus which is expected to help humanity in interpreting the Quran. This research produces a system that can search for the same verse using the LCS approach, then results will be collected into a corpus based on the LCS results. The average results obtained from several tests that have been done are Arabic data getting the precision value is 46.84%, the recall value is 96.13%, and the f1-score value is 62.96%. While Indonesian Data getting the precision value is 40.57%, the recall value is 97.56%, and the f1-score value is 57.04%.

Keywords: Corpus, Longest Common Subsequence, The Quran, Word Sequence.