

## Abstract

To understand the Quran better, correct interpretation requires paying more attention to verses in the Quran. From 114 surahs of the Quran, according to al-Khatib al-Iskafi (dead 420 H./1026 M.), only 28 or about 25% do not contain similar or repeated verses. Currently, to get longest recurring verses in the Quran requires manual effort, by searching the Quran verse by verse from beginning to end which takes a long time. This is a problem of longest common substring (LCS) in the Quran. We propose a system using Ukkonen's algorithm to build a database of repetitive text sets. Ukkonen's algorithm is chosen because this method of solving the longest common substring (LCS) problem has better efficiency than previous methods. We use two language translations, Indonesian and Arabic. We use 789 verses or 5 surahs in the Quran. We limit longest repeated text sets containing at least three words. For Indonesian we produced 2341 repetitive text sets, for Arabic we produced 2567 repetitive text sets, which are then stored in the database. After evaluating this system, for Indonesian we obtained precision of 45.9%. Whereas for Arabic, we obtained precision of 55.5%. This research produced a new database that did not exist previously in scientific papers. The resulting database can be used for various studies on the Quran and can also be used to compare with other studies about the Quran.

The code and resulting database is available at: <https://github.com/deviokta27/Al-Quran.git>.

**Keywords:** Database, Longest Common Substring (LCS), Quran, Ukkonen's Algorithm.