

ABSTRACT

The Qur'an is a holy book that is a role model and reference for all Muslims in the world. The Qur'an consists of 114 letters, 6236 verses, and 77845 words. Because the Qur'an has a lot of information in it, one effort to facilitate the reader in learning and filtering information from the Qur'an is to make a brief paraphrase by using the Rule-Based Chunking system.

Rule-based chunking can shorten a sentence by grouping words that are considered important based on grammar rules. The process of filtering information in this study is carried out through several stages, namely breaking down words in each sentence in each verse, labeling Part of Speech (POS) for each word, search for words that are in accordance with the rules based on grammar rules so that they can be grouped and shortened by rulebased chunking. The system has input in the form of verses of the Qur'an Indonesian translation that have been sought Longest Common Substring and Longest Common Subsequence with output in the form of paraphrases.

From 2341 repetitive texts obtained from previous research, we produced 1261 chunks. After evaluating the system, accuracy we obtained is 91%, with 63.5% precision and 74.4% recall.

Keywords: Al-Quran, Rule-Based Chunking, paraphrases