

## ABSTRACT

Understanding the meaning of a text, entirely or partially, that is written in a document is an activity that is essential in the process of understanding the document as a whole. Errors in understanding the meaning of a text that is a part of the document, especially the important words, will create inaccuracies in making the overall perception of a document. The scripture book of Al-Quran is an important document in the lives of Muslims around the world. It consists of 114 chapters (surah in Arabic), which are divided into verses (ayah in Arabic). Practical Indonesian Muslims, because they do not speak Arabic, rely heavily on the document translation of the Quran in the Indonesian language as a means of understanding the meaning of the contents of the document in the Quran itself.

In the context of the Al-Quran in Arabic, it is known that one word in Arabic can be translated into two or more words in Indonesian. In this case, we have to parse the words in the document and combine several words into a single word originally in Arabic. The process of decomposition and the determination of the words in these segments is called phrase segmentation.

Our research proposes to perform a process phrase-based segmentation for document translation Quran Indonesian, in this case the surah Yasin, the approach method monolingual alignment and parsing, which made the process of tokenization, word similarity with the reference Quran phrase database, post tagging and parsing resulting level suitability considerable when compared with the gold standard

Keywords: phrase-based segmentation, monolingual alignment