

## ***Abstract***

*In the Muslims life, the understanding of the Holy Qur'an contents as a life guide is very important. Holy Qur'an as the main Muslims references are written in Arabic language. To process the text, one of the initial preprocessing is tokenization. So all the text processing requires tokenization done beforehand. In Indonesian and English languages, tokenization can simply be done, where tokens are bounded by a space. But in Arabic, as well as in several other languages, such as Japanese and Chinese, spaces can not be used as a token boundaries. In this final project, tokenization will be carried to Arabic in a Holy Qur'an case. By using a Maximum Matching (also called Greedy) as a baseline, the matching process was done against the dictionary.*

**Keywords:** *Tokenization , Arabic , Qur'an*