Part-Of-Speech Tag Studi Kasus Kelas Kata dalam Al-Quran

Muhammad Amal Dzaky[1], Moch. Arif Bijaksana[2], Arif Fatchul Huda[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung

[1]mhdamaldzaky@students.telkomuniversity.ac.id,
[2]arifbijaksana@telkomuniversity.ac.id, [3]afh@uinsgd.ac.id

Abstract

Indonesia is the fourth most populous country and is the largest country for the world's Muslim population. As it should be, the Qur'an becomes a guideline for the lives of Muslims. Al-Qur'an was written 1400 years ago and now Muslims contain the contents of the Qur'an, even starting from a small age. The Qur'an has 114 Surahs, 6000 more verses, and 128219 words. Because the holy Qur'an uses Arabic, which is how Arabic is morphologically complicated. Because the Qur'an has many words, it would be better if the learning process of the Qur'an is helped by utilizing Natural language Processing (NLP). In this study, the Part-Of-Speech Tagging will be conducted on the holy Qur'an using the Term Frequency Inverse Document Frequency (TF-IDF). The results of the classification have the accuracy of Naive Bayes 24.7725% for the Arabic-language corpus, 54.9836% for the normalized Arabic writing, and 64.4283% for the normalized Buckwalter absorption. For the Random Forest, 38,077% for the Arabic-language corpus, 69.6771% for the normalized Arabic writing, and 81.568% Buckwalter anotation.

Keywords: *Part-Of-Speech*, *Random Forest*, *Naive Bayes*, *Term Frequency Inverse Document Frequency*, *Na- tural language Procesing*, Al-Qur'an.