NEWS ARTICLES CLASSIFICATION IN INDONESIAN LANGUAGE BASED ON NAIVE BAYES CLASSIFIER USING CONFIX-STRIPPING STEMMER

Name : Amalia Anjani Arifiyanti
NRP : 5210 100 106
Departement : Information System FTIF-ITS
Supervisor : Prof. Ir. Arif Djunaidy, M.Sc, Ph.D
Renny Pradina Kusumawardani, S.T., M.T

Abstract

Dynamic as well as rapid changing news stream needs news article classifier to ease the process of news document classification. Nowadays, there are already many methods for classifying news. Most of them, however, are limited only in the use of corpus for certain foreign languages. Indonesian language corpus has its own challenging section due to its different suffixes compared with other languages.

In this final project, an application for classifying news article in Indonesian language based on Naive Bayes classifier using confix-stripping stemmer is developed. Before classifying process is performed, a preprocessing step needs to be done to eliminate stopwords and split the article into a number of tokens. The confix-stripping stemmer algorithm is then applied to obtain the term corresponding to each token being produced. The news article classifier developed in this final project was modeled using all terms obtained from several news articles representing all news categories to be accommodated.

A news article classifier application that has been successfully developed in this final project was tested using a set of news articles consisting of 12 categories. The test results showed that the application, which is developed under the windows operating system, is capable of producing sufficiently good classifier with accuracy, precision, recall, and f-measure of 87.675%, 88.21%, 87.65% and 87.69%, respectively.

Keywords: confix-stripping stemmer, Indonesian language stemming, naïve bayes classifier, news articles, text classification.
Halaman ini sengaja dikosongkan.