Abstract

In the Text Preprocessing, term weighting is a step that is very important. This step is applied in order to give a value/weight on the term that is contained in a document. The weight given to a term depends on the method that is used for the weighting. In the text mining, there are some term weighting method such as TF, TF·IDF, WIDF, TF·IG, and TF·RF.

In this Final Task, some term weighting methods like TF·IDF, and TF·RF, are compared each other by seeing the output of the text categorization performance. Some parameters that will be used as a measurement for comparing the text categorization performance are recall, precision, and f-measure. To test the output of the weighting result, it will be used a classification tool like Weka, with NaiveBayes and Naïve Bayes Updateable as its classifier.

Based on the result, it is concluded that the TF·RF weighting method has better performance compared with TF·IDF weighting method. In general, the TF·RF outperformed TF·IDF in some testing done. Excess TF·RF which takes into account the frequency of occurrence of a term in a category and menormalisasikannya to the whole document, make this method better than the TF·IDF.

Keyword: *text preprocessing, term, term weighting, TF·IDF, TF·RF.*