

Abstract

In the world of internet, the amount of information and its variety are always continuously increased. Therefore, it is resulted to become unclassified and unorganized that could make number of important information becoming left out and the user would have difficulty to access his or her desired information. The most obvious example is the case information in the form of article on Indonesian based language. Through process of text mining with the method of categorization, the selected number of articles could be well organized through classification based on the topic and its content so that user can access his or her desired information easily.

RIPPER is the method of categorization that is based on contexts; these contexts are the relationship between the existence and non-existence of one word with another word. This final project would implement *RIPPER* in the selected number of articles in Indonesian based language from some of news source that come from web and offline based. In order to analyze the performance of *RIPPER*, the author correlated it with *Naïve Bayes* that has method of categorization based on the non-contexts in the clean and noise data.

Based on the overall test result, it showed that *RIPPER* has weaker performance than *Naïve Bayes* and slower in the development model on the clean and noise data. This was suggested that the use of classification method based on contexts was less effective to classify document than the use of classification method based on non-contexts. Moreover, since the project was looking at the contexts from its content, then time in developing classification method of contexts-sensitive was longer than insensitive.

Keywords: text mining, categorization method, context-sensitive, insensitive, *RIPPER*, *Naïve Bayes*