

INDONESIAN LANGUAGE DOCUMENT CLASSIFICATION USING SEMANTIC SMOOTHING

Aprilia Ramadhina, Julio Adisantoso

Department of Computer Science
Faculty of Mathematics and Natural Sciences
Bogor Agricultural University
julioipb@gmail.com

August 13, 2011

Abstract

The first supervised learning method for document classification is Naive Bayes classifier. A common problem that often occurs on simple methods like Naive Bayes is data sparsity, this problem especially occurs when the size of document training and dokument testing is too small. Smoothing technique is method for handle the sparsity problem, one of method from smoothing technique is Semantic Smoothing. Document Agricultural Research Journal of holticulture domain is used for this research, this document contain of three class. The average for accuracy of document classification on Semantic Smoothing is 92.88%. Results of the classification with Semantic Smoothing has been able to classify document of Agriculture Research Journal at holticulture domain.