ABSTRACT

Along with the growth of sites on the Internet that is very fast, needs to be done the preparation and organization of web documents (webpages) for easier search, manage, and obtain information in accordance with their needs. The process of classification is the solution that is by using machine learning.

The method of machine learning that used before is *Support Vector Machine* (SVM), but that was one weakness is the classification classifier produced by the SVM can't be known whether an assumpsion or definite answer. In this final task will be done a new approach in classfying web documents using the SVM in order to the classification that produced become reliable, namely by applying the *Version Space* (VS). Version space is an approach to obtain a reliable classification. The main idea is to build a version space that contains a set of function hypothesis. Rule of the version space, called the unanimous vote rule, will be used to ensure that if a new data classified then it correctly classified.

In this final task will be to do an analysis of the results of training and testing. Testing will also apply feature selection. Results from the study showed that the optimal performance obtained when applying the feature selection is 82.35% in the first domain data by using only 60-90 attributes of total 175 attributes, and 80% in the second domain by using only 30 attributes of total 214 attributes.

Keywords: classification, preprocesing, reliable, SVM, VS, webpage.