Abstract

This paper describes experimental results in an attempt to identify the Transformation techniques which can be adopted to improve features for the automation of classification of Swahili documents. This means improving classification rate by enhancing separability and accuracy. The experiment involved Relative Frequency (RF), Power transformation (PT) and Relative Frequency with Power transformation (RFPT). The Term weighting with TFIDF and the absolute features (AF) were also studied. The features' dimension reduction was done by using the statistical techniques of Principal Component Analysis. In learning algorithm, the Support vector machine for classification and the k-NN were used, and in evaluating the effect of features' performance with the classifiers the micro averaged f-measure were adopted. The extensive experimental results demonstrated that the RFPT features worked better with the Support Vector Machine classifiers unlike k-NN in improving the classification rate by enhancing document separability and accuracy in Automation of Swahili document classification.

Index Terms

Computer Science  Artificial Intelligence

Keywords
Machine learning algorithm, Support vector machine, Swahili, Swahili document classification