

Text clustering and classification are important machine learning tasks. In this work, a combination of their approaches is presented. The main purpose was to automatically prepare a set of clusters (or generally concepts), which would subsequently serve as a training data for learning of a classifier. This work comprises of theoretical background, implementation details and experimental results of clustering and classification of text documents. A train set of documents is first hierarchically clustered by the bisecting k-means algorithm. The result is offered to an expert for modifications and possible improvements of the hierarchy. Following this, the resulting structure is used for learning of a naive Bayes classifier and a test set of documents is classified by it. A program was developed to perform these tasks and its results are evaluated and compared in processing document collections written in both English and Czech.