

The work presents the field of document classification. It describes existing techniques with emphasis on the Naive Bayes' classifier. Several existing feature selection methods suitable for the Naive Bayes' classifier are discussed.

This theoretical background is the basis for the implementation of a classification library based on the Naive Bayes' method. Besides the classification program, the library provides a range of document preprocessing tools. They allow to work with different types of documents and, more importantly, they significantly reduce redundant document dimensions.

Eventually, we tested the library on two different datasets and compared implemented feature selection methods. The functionality of the whole library is practically verified by including it into the open-source email client Mailpuccino.