

This thesis follows up text categorization. In the first part are described several chosen algorithms for a categorization of documents - the Bayesian model, a categorization with a neural networks and a vector model. Practice part is focused on a algorithm vector model. The vector model is based on idea of two vectors. One vector represents a pattern and second a query. In our case first vector corresponds with a category and the second one with the document. Coordinates of the vector are weights of single words in the text or in the branch depends on, which vector we think about. For comparing are possible to use several procedures like Dice coefficient similarity, Jaccard coefficient or cosine similarity. In my thesis is used cosine similarity. Computing weights is based on frequency of the term in the document and on frequency of documents, which contain the term. Relevant terms are selected on Luhn simple ideas of significance words.