Processing and Categorization of Czech Written Documents Using Neural Networks

The Kohonen Self-organizing Feature Map (SOM) has been developed for clustering input vectors and for projection of continuous high-dimensional signal to discrete low-dimensional space. The application area, where the map can be also used, is the processing of text documents. Within the project WEBSOM, some methods based on SOM have been developed. These methods are suitable either for text documents information retrieval or for organization of large document collections. All methods have been tested on collections of English and Finnish written documents. This article deals with the application of WEBSOM methods to Czech written documents collections. The basic principles of WEBSOM methods, transformation of text information into the real components feature vector and results of documents classification are described. The Carpenter-Grossberg ART-2 neural network, usually used for adaptive vector clustering, was also tested as a document categorization tool. The results achieved by using this network are also presented.