ABSTRACT

AUTHOR RECOGNITION FOR TURKISH DOCUMENTS

Özcan KOLYİĞİT

M.Sc. Thesis, Department of Mathematics
Supervisor: Assist. Prof. Dr. Rıfat AŞLIYAN
2013, 57 pages

Today, the studies of author recognition have been made for providing the solutions of the problems which occur by developing and growing of information technology. Some of these problems are to specify the authors who the papers are exactly written by. In this study, some systems about author recognition for Turkish documents have been developed. For generating the systems, we have used the columns which belong to six authors in some newspapers. A corpus which includes totally 420 documents is constructed for training and testing of the systems. Each author has seventy documents. Twenty documents of every author are used for training operation. But, the other documents are utilized for testing stage. The features of word, stem, syllable, character and their *n*-grams are decided for each documents of these six author. Author recognition systems have been developed with the methods as K-Nearest Neighbor, Support Vector Machine, Multi-Layer Perceptron and Learning Vector Quantization. The feature vectors' lengths of the systems developed by K-Nearest Neighbor have been chosen as 120, 180 and 240. Because the most successful results are obtained as the length of the feature vectors is 120, we have used this length for the other methods. After the developed systems are trained the methods, the systems have been tested and evaluated according to accuracy and F-measure values.

Key words: Author Recognition, K-Nearest Neighbor, Multi-Layer Perceptron, Support Vector Machine, Learning Vector Quantization, *n*-gram.