

In the present thesis we solve the problem of symbol extraction and recognition from printed documents digitized by the scanner or camera. We introduce a noise resistant algorithm of document lighting normalization. We continue with the extraction of individual characters from the document and their recognition with a system of feedforward multilayer neural networks. We also focus on processing of the resulting set of recognized characters, which is necessary for further use of the extracted text. The last step is correction of the output based on surrounding letters of each character. We have successfully implemented an automatic system containing all the above components.