RESEARCH ON DIGITAL FINANCIAL DOCUMENTS PROCESSING USING OPTICAL CHARACTER RECOGNITION TECHNIQUES

Dmytro Slobodeniuk, Andrii Kopp
National Technical University «Kharkiv Polytechnic Institute», Kharkiv

In this study, we address the practical task of automation and optimization of document processing flow, namely processing of digital financial documents by means of optical character recognition process to improve accuracy of information extraction and increase the efficiency when working with huge amount of documents.

The volume of data and documents rapidly grows and it is important to have efficient tools to process and analyze this information. Also with the development of automatic computing engines and the widespread distribution of multimedia content, most financial information is presented in the form of digital images. The vast majority of such images represent printed financial information [1].

The documents processing flow can be briefly described as follows:

1. Operator takes a set of financial documents.
2. Sorts documents by date of receiving.
3. Selects one document and saves the values from all fields to a database.

Manual processing of paper documents can become very tedious and exhausting for operators and is a costly and inaccurate process that requires significant time and resources, as operators have to manually select documents one by one and extract information to add it to database for further analysis.

Therefore, the research and development of OCR (Optical Character Recognition) systems is becoming an important task to automate the process of obtaining textual information and improve work efficiency in various industries [2].

This study proposes to conduct a detailed analysis of the process of working with OCR software components to improve the efficiency of document processing and improve the overall workflow by developing an appropriate software solution using OCR techniques. The main OCR approaches, including pattern-matching and machine learning methods [2], were considered and discussed.

Therefore, the following goals were achieved in the work:

1) investigated existing methods for solving the problem of processing digital documents by means of optical text recognition;
2) analyzed the existing software solutions for processing digital documents by means of optical text recognition.

In the future research, we plan to formulate requirements for a software solution for digital financial documents processing using OCR techniques, select software development tools, design and develop the software solution, perform experiments using the developed software, and analyze the obtained results.

References:

1. Awel, M. A., Abidi, A. I. (2019). Review on optical character recognition. *International Research Journal of Engineering and Technology (IRJET)*, 6(6), 3666- 3669.
2. Optical Character Recognition (OCR): Definition & How To Guide. (2021). *Www.v7labs.com*. https://www.v7labs.com/blog/ocr-guide.