

In this work I developed an application capable of extracting data from scanned documents. For optical character recognition, I used external OCR engine Tesseract, but it can be easily changed. I use document templates, which have informations about data areas and its data types. I tried to automatize most of the steps which are required to extract data or create new data template. User can improve or change results of these steps. For export from application I implemented components, which export data to XML, HTML or plain text. Another components can be easily added, to adapt application for various uses.