

Complete OCR Solution for Image Analysis of World War 2 Documents

Friday 16 September 2022 14:20 (20 minutes)

The field of Optical Character Recognition (OCR) consists of techniques that are mainly focused on document image analysis. Aside from generating significant speedups of everyday procedures, OCR has a considerable role in the preservation of historical sources of information. Most of the World War 2 (WW2) documents are of great importance, especially with applications in virtual archives, museums, and research. The situation asks for an efficient, yet not aggressive, transcribing method using OCR tools. This paper describes an approach in the context of the given problem. The focus is oriented towards extracting the information from documents affected by their age, but with simpler structures, mainly split into paragraphs, such as letters and military reports. The approach is based on combining the results of multiple OCR engines, with the final objective of achieving better performance compared to the individual performance of each engine.

Authors: Mr GRĂDINARU, Mihai (University Politehnica of Bucharest); Mr NEGRU, Andrei (University Politehnica of Bucharest); Prof. BOIANGIU, Costin-Anton (University Politehnica of Bucharest); Mr TARBĂ, Nicolae (University Politehnica of Bucharest); Mr VONCILĂ, Mihai (University Politehnica of Bucharest); Prof. DEACONESCU, Răzvan-Adrian (University Politehnica of Bucharest)

Presenter: Mr TARBĂ, Nicolae (University Politehnica of Bucharest)

Session Classification: Session 3A - Grid, Cloud & High Performance Computing

Track Classification: Grid, Cloud & High Performance Computing in Science