

Optical Character Recognition (OCR) is a process of converting text from images to a machine-readable text. We run three OCR systems (Tesseract, Ocrad and GOCR) on an original multilingual OCR dataset and perform statistical and linguistic analysis of the results in order to compare the tested systems and investigate typical OCR errors.