

Optické rozpoznávání znaků (OCR) je proces převodu textu z obrázku do strojově čitelného textu. Spustili jsme tři OCR systémy (Tesseract, Ocrad a GOCR) na vytvořeném vícejazyčném datasetu a provedli statistickou a lingvistickou analýzu výsledků za účelem porovnání testovaných systémů a identifikování typických OCR chyb.