

Stacja 1 — popraw zepsuty skan (OCR)

Plik źródłowy: przyklad-korpus/poznan-1956/wycinki-prasowe/wycinek-01-ocr.txt

To komunikat prasowy z 1956 r. „przepuszczony przez zły OCR”: 1 zamiast l, 0 zamiast o, pogubione polskie znaki, ucięte wyrazy.

Prompt (wklej, a pod nim wklej tekst z pliku)

Popraw błędy OCR w poniższym tekście. Odtwórz poprawną polszczyznę z polskimi znakami (ą, ę, ł, ó, ż, ź...), napraw pomyłone znaki (1→l, 0→o), połącz ucięte wyrazy. Zachowaj oryginalny sens i styl. Nie dodawaj nic od siebie, nie komentuj.

Czego się spodziewać: czysty, czytelny tekst komunikatu w kilka sekund — gotowy do dalszej pracy.

Wariant trudniejszy

Po poprawieniu tekstu dopisz krótką analizę: jakim językiem jest napisany ten komunikat (neutralny / propagandowy)? Wskaż 3 sformułowania, które o tym świadczą.

Po co to w pracy naukowej: stare PDF-y, skany z archiwum i zdjęcia dokumentów prawie zawsze mają błędy OCR. Model czyści je szybciej niż ręczna korekta — a Ty zostajesz przy weryfikacji.