ABSTRACT

DATA ENTRY VALIDATION DESIGN USING OCR AND NLP INTEGRATION WITH CONDITIONAL RANDOM FIELD (CRF) METHOD

Goes Rizky Apriadi Nurramdahan [1)], Safitri Jaya [2)]

[1)] *Student of Informatics Departement*, Universitas Pembangunan Jaya

[2)] *Lecturer of Informatics Departement*, Universitas Pembangunan Jaya

Errors in the manual data entry process, especially from printed or handwritten documents, are still an obstacle in maintaining the quality and reliability of digital data. To overcome these problems, this study aims to design an automatic data entry validation system by combining Optical Character Recognition (OCR) and Natural Language Processing (NLP) technologies. The approach used in this system is the Conditional Random Field (CRF) method which functions to extract and recognize entities from the results of OCR text conversion. Testing was carried out on a collection of semi-structured document data to measure the level of system accuracy in recognizing and validating information. Based on the evaluation results, the system managed to achieve an accuracy of up to 92% in the extraction and validation process, and was able to minimize errors that commonly occur in manual data entry. Thus, the application of a combination of OCR, NLP, and CRF methods has proven effective in increasing the accuracy and efficiency of the automatic data validation process.

Keywords: Optical Character Recognition (OCR), Natural Language Processing (NLP), Conditional Random Field (CRF), Data Validation, Automated Data Entry.

Libraries :

Publication Years : 2020 - 2025