"RESUME ANALYZER"

AUTOMATED CV ANALYZING SYSTEM USING NLP AND OCR

Chamodh Randil Karunarathna

A dissertation submitted in partial fulfilment of the requirement for

Bachelor of Engineering (Honours) degree in Software Engineering

Department of Computing

Informatics Institute of Technology, Sri Lanka
In collaboration with
University of Westminster, UK

2021

Abstract

The potential to use automation in the recruitment process is more appropriate for large companies, which can revolutionize the recruitment process. Although the company has a high level of skilled human resources team, it takes a lot of manpower, time and effort to select the best candidate for an open position in any large company with the number of resumes received when advertised.

The recruitment process is now semi-automated. It's fair to believe that the first time you were selected for a job vacancy in a company was a computer program, not people. The recruitment process is a large process consisting of several stages. AI will be used to facilitate inefficient stages of the recruitment process. It is simply the elimination of the human process of reading CVs.

Candidates are now tempted to include text in CVs using beautiful images, charts and fancy looks. CVs can be sent in PDF type document to companies in various encodings such as UTF-8, ASCI, Unicode. Existing solutions to automate the recruitment process are not a method of extracting text from images and reading PDF documents with various encodings and extracting data from those documents. These are the basic problems that this research hopes to solve. If semi-automated software that can filter an applicant's CV cannot read CVs and extract critical information, applicants will be excluded from the job competition. Every candidate has many factors to compare among other candidates. Within them, Skills play an important role. So that Existing CV analysis systems do not provide a CVs comparison process.

Combining OCR technology tools with natural language Processing presented a solution that could extract text from images and read PDF documents with various encodings such as UTF-8, ASCI, Unicode. The entire system was developed with a pre-evaluation by Domain experts. The system stands out for its ability to read all encoded PDF formats and read text contained in images while scaling with existing solutions. The system was evaluated and produced an accuracy of 92% and a recall value of 82%, which was found to be very satisfactory in benchmarking similar solutions.

Key words: Human Resource Management, Recruiting Process, Natural Language Processing, Optical Character Recognition, Similarity Matching