International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 6 ǁ 2021 ǁ PP. 01-06

Resume Parser Using Natural Language Processing Techniques

Shubham Bhor[1], Vivek Gupta[2], Vishak Nair[3], Harish Shinde[4],
Prof. Manasi S.Kulkarni[5]

[1]*Department of Computer Engineering, Pillai College Of Engineering, New Panvel, Maharashtra*
[2]*Department of Computer Engineering, Pillai College Of Engineering, New Panvel, Maharashtra*
[3]*Department of Computer Engineering, Pillai College Of Engineering, New Panvel, Maharashtra*
[4]*Department of Computer Engineering, Pillai College Of Engineering, New Panvel, Maharashtra*
[5]*Professor (Department of Computer Engineering), Pillai College Of Engineering, New Panvel, Maharashtra*

Abstract
Job Requirement is considered one of the major activities for humans which is a very strenuous job to find a fruitful talent. Our proposed model is basically to extract the details and statistics from the resume and ranking the resume based on the preference of the company associated and its requirements using the Natural Language Processing (NLP) techniques. Parsing and ranking the resume makes the hiring process easy and efficient. A resume contains various minute data within it and any respectable parser needs to extract out these data such as education,experience,project,address etc. So,basically we are going to build a job portal where the employees and applicants would upload their resume for any particular job and using the NLP technique, the necessary information will be parsed and a structured resume with information will be generated and also the resumes of employee will be ranked according to the requirement of the company skill set and employees skills in the provided resume.

Keywords: *Resumes,NLP,Parser,Extract Information,Skillset,Ranking.*

Date of Submission: 10-06-2021 Date of acceptance: 24-06-2021

I. INTRODUCTION

Corporate companies and recruitment agencies process numerous resumes daily. This is no task for humans. An automated intelligent system is required which can take out all the vital information from the unstructured resumes and transform all of them to a common structured format which can then be ranked for a specific job position.Parsed information include name, email address, social profiles, personal websites, years of work experience, work experiences, years of education, education experiences, publications, certifications, volunteer experiences, keywords and finally the cluster of the resume (ex: computer science, human resource, etc.). The parsed information is then stored in a database (NoSQL in this case) for later use. Unlike other unstructured data (ex: email body, web page contents, etc.), resumes are a bit structured. Information is stored in discrete sets. Each set contains data about the person's contact, work experience or education details. In spite of this, resumes are difficult to parse. This is because they vary in types of information, their order, writing style, etc. Moreover, they can be written in various formats. Some of the common ones include '.txt', '.pdf', '.doc', '.docx', '.odt', '.rtf' etc. To parse the data from different kinds of resumes effectively and efficiently, the model must not rely on the order or type of data.

II. LITERATURE SURVEY

A. NLP Based Extraction of Relevant Resume using Machine Learning:

This technique stated parsing of the resumes with least limit and the parser works the utilization of two or three rules which train the call and address.Scout bundles use the CV parser system for the determination of resumes. As resumes are in amazing arrangements and it has different sorts of real factors like set up and unstructured estimations, meta experiences, etc. The proposed CV parser approach gives the component extraction method from the moved CV's.

B. E-Recruitment System Through Resume Parsing, Psychometric Test And Social Media Analysis:

It follows an approach of 4 stages, the first stage was to get the data (resume) and convert them into structured format and then perform the analysis using deep learning techniques. Second step includes the psychometric test where the text mining is used to generate scores for each candidate. In the third step they perform web scraping on various social media sites to get the additional information about the candidates and recommend suitable jobs to them. In the fourth step, the system will recommend the skills and requirements in which the students are lacking and also help them to get recruited in the desired company.

C. Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing:

The techniques used in this category are neural networks and CRF to segment and extract various information from resumes. CNN model is used for segmentation and compared with a Bi LSTM model. A CRF based model is chosen for information extraction and compared with a Bi-LSTM-CNN model. They segmented and extracted several pieces of information from personal, educational and occupational blocks. The results are promising and the output JSON file contains 23 data fields.

D. A CV Parser Model using Entity Extraction Process and Big Data Tools:

Here the problem definition was based on designing an automated resume parser system, which will parse the uploaded resume according to the job profile. And it will transform the unstructured resumes into structured format. It will also main-tains a ranking system on the resumes. Ranking will depend on the basis of information extracted i.e according to technical skills, education etc. Here the CV parser is used.CV parsing is such a technique for collecting CV's. CV parser supports multiple languages, Semantic mapping for skills, job boards, recruiter, ease of customization. Parsing with hire ability provides us accurate results. Its integration makes users API key for integration efforts. The parser operates using some rules which instructs the name and address. Recruiter companies use CV parser technique for selection of resumes. As resumes are in different formats and it has different types of data like structured and unstructured data, metadata etc. The proposed CV parser technique provides the entity extraction method from the uploaded CV's.

E. An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes:

In this work, a qualitative assessment of resumes on the basis of different quality parameters using a simple text analytic based approach for a resume collection was described. The resume collection was processed for two qualitative coverage, comprehensibility and the aspects; and extracted ratings are modified into a quality rating which is comprehensive. All the parameters were collectively uniformed into a combined 1 to 5 rating scale for associating a quality metric for resumes. The qualitative evaluation results obtained through the algorithmic approach were congruent to and were hence validated through the wisdom of crowds.

No.	Author & Year of Publications	Paper Title	Observations and remarks
1.	Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh 2020,IJITEE	NLP based Extraction of Relevant Resume using Machine Learning	● Parsing with lease limit ● Similarity index for skill sets
2.	Dr.Parkavi A,Pooja Pandey,Poornima J,Vaibhavi G S Kaveri B W2019,IJARBEST	E-Recruitment System Through Resume Parsing,Psychometric Test And Social Media Analysis	● Text mining is used to generate scores ● Web scraping
3.	Ayishathahira C H,Sreejith C,Raseek C 2018, International CET Conference on Control, Communication, and Computing (IC4)	Combination of Neural Network and Conditional Random Fields for Efficient Resume Parsing	● Classify resume into three segments ● Extract 23 different data fields
4.	Papiya Das,Manjusha Pandey and Siddharth Swarup Rautaray 2018, IJITCS	A CV Parser Model using Entity Extraction Process and Big Data Tools	● Convert unstructured resumes to structured ● Supports multiple language
5.	Satyaki Sanyal,Neelanjan Ghosh, Souvik Hazra, Soumyashree Adhikary	Scrape keywords Conversion of different formats of resumes to text	● Scrape keywords ● Conversion of different formats of resumes to text
6.	Vinaya R. Kudtarkar, Manjula Ramannav ar, Dr.Nandini S. Sidnal 2015 IJIRAE	An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes	● A simple text analytic ● Quality metric for resume ● Comprehensive quality rating

Table 1: Literature Table

III. PROPOSED WORK

In this proposed methodology we are using Natural Language Processing technique for parsing the resume according to the particular companies.. A common job portal for employers as well as employees is provided to apply and create the job. The resumes received would be parsed and ranked according to company requirements.Additionally our other goal is to extract the data from Social Media like Linkedin for applying jobs which will make the recruitment process easier getting quality applications from various regions by avoiding unfair and discriminatory practice.

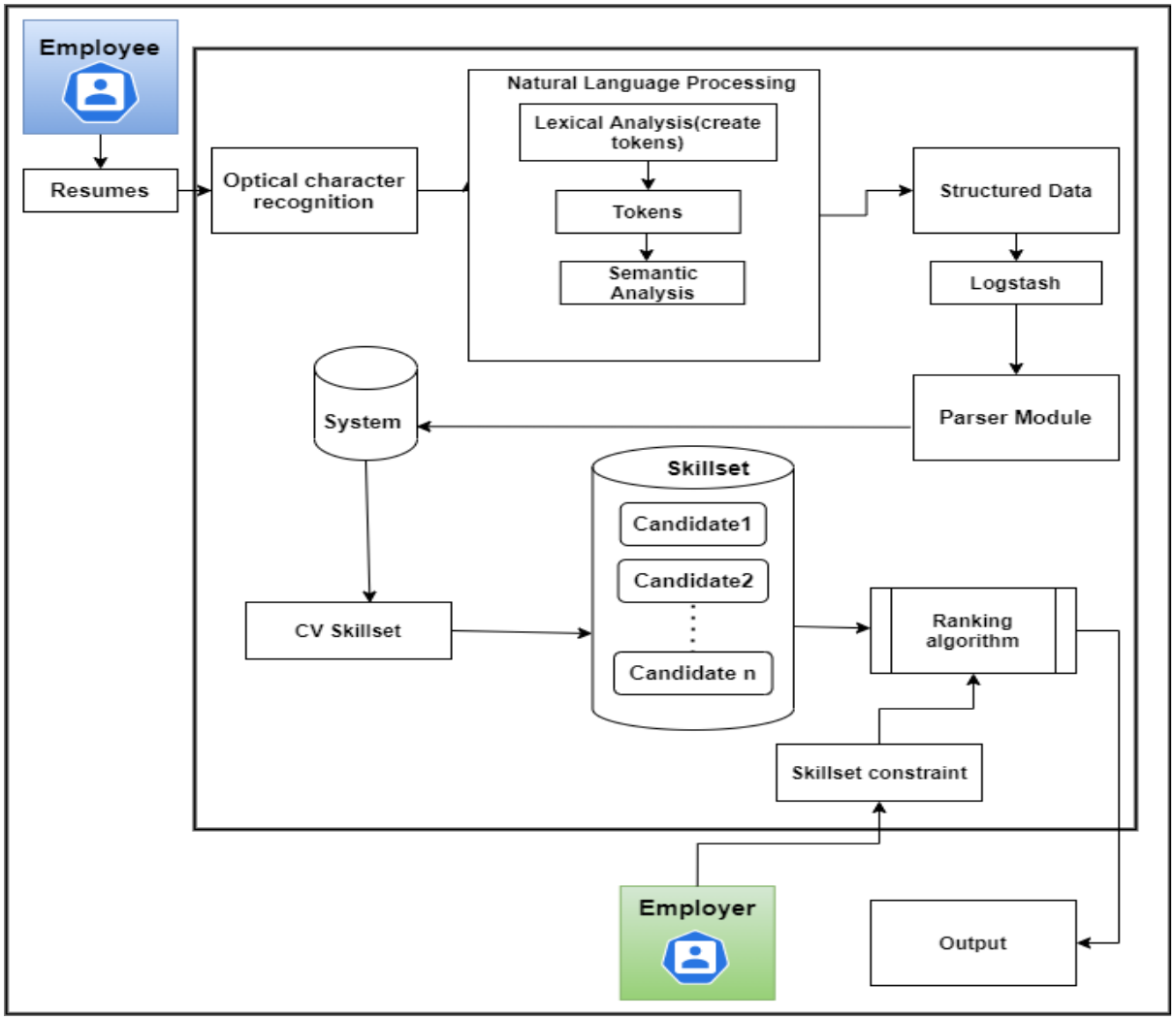

Fig.1: System Architecture

To use the Resume Parser system both the employee and company needs to provide the detailed resume by employee and job requirements skill set constraints by company. The working of the system architecture is as follows:

1. A web portal is provided to the HR, to define the constraints and the required skill sets of the company, on which the applicants are to be judged.
2. The Candidate also needs to use the web portal to create his account and upload the resume.The Resumes uploaded by the candidate are fetched and fed to the OCR.
3. As the resumes uploaded can be of any format such as '.txt', '.pdf', '.doc', '.docx', '.odt',etc. we will use Optical Character Recognition to convert the resume to a single text format.
4. The Converted resume is then Fed to the Natural Language Processing module it takes the plain text as input and converts it into meaningful data.Using NLP, we are going to parse the resume, NLP requires the following for parsing:

1) Lexical Analysis:It is the first phase of NLP parsing,in which the plain text input is segmented into words and paragraphs and then the tokens are created.

2) Syntactic Analysis:In Syntactic analysis the analysis of the grammar and the arrangement of words in a meaningful manner is checked,sentences like "College goes to girl" is rejected.
3) Named Entity Recognition (NER):One of the problems with using the same NLP module for all the companies is the jargon and words that mean something for that company's domain and may mean something else in general. This hindrance is overcome in our system with the help of "Named Entity Recognition" or NER.A named entity is an object that exists in the real world. With NER, we can fine tune our NLP module to understand the real word objects from a domain.
5. The company or the HR of the company will provide the skill set requirement for the job posted by them. Further the skills mentioned by the applicants in the resume which was tokenized is compared with the required skill set.
6. Using queries inbuilt in the whole process, the applicant resumes will be scored and then they will be provided in the form of a bar graph and pie chart with whole statistics .
7. At the end, the percentage of the bar graph will be used to sort the applicants. A final list of shortlisted applicants for the further placement process will be generated..

Fig. 2: Overview

IV. RESULTS AND DISCUSSIONS

The result involves parsing the resume from /pdf/doc/rtf format into plain document and tokenized data entities. Resumes were ranked by comparing extracted entities and required keywords and the result was finally provided in the form of Pie Chart and Bar Graph.

Fig. 3: Entities Extracted From Resumes

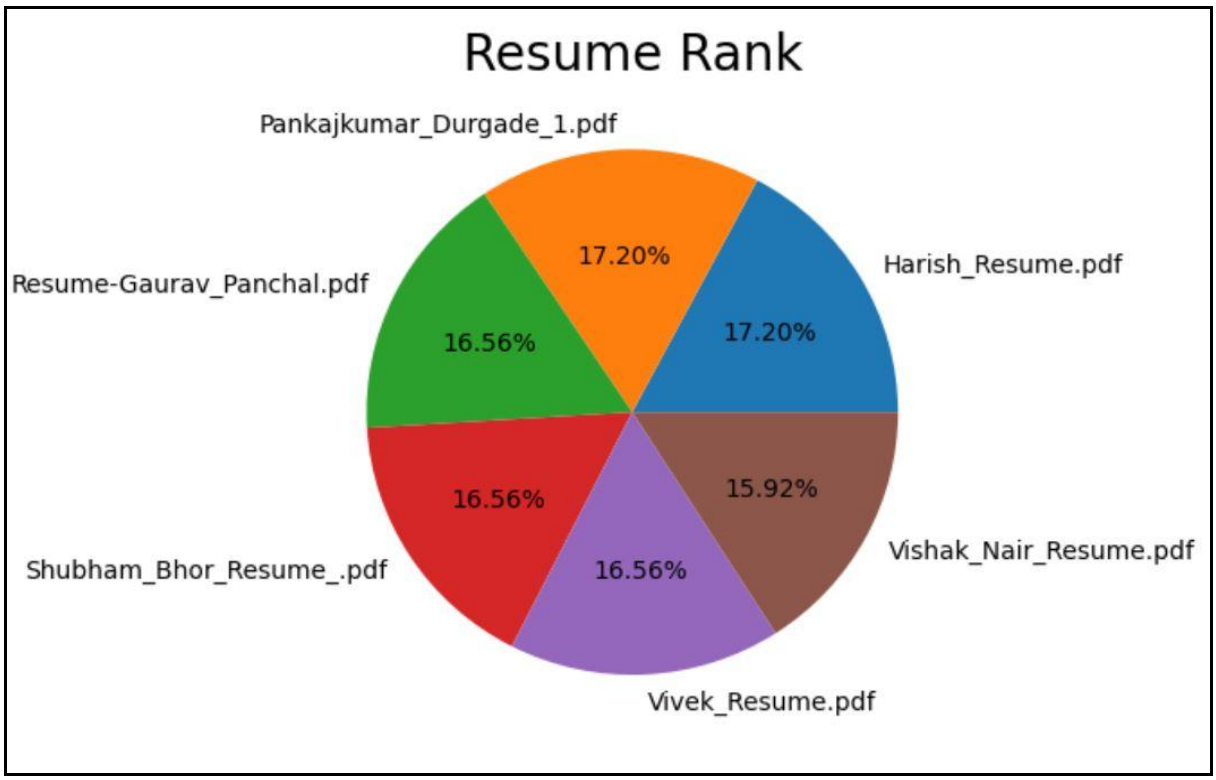

Fig. 4: Pie-Chart

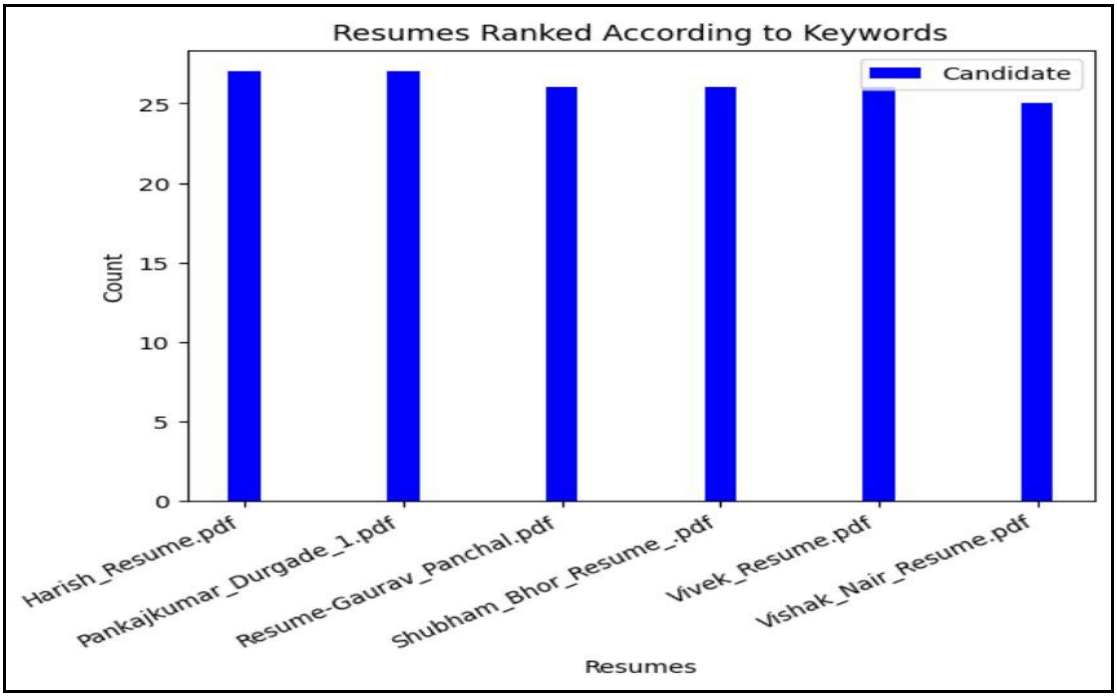

Fig. 5: Bar Graph

V. CONCLUSION

Our approach is to make the work of companies and candidates easier and effective. Basically our aim is to ease the recruitment process. The process will provide the quality of applicants for the companies. The unfair and discriminatory practice in the process will be dampened. Based on the information in the form of technical skills the resumes will be ranked in order.

VI. FUTURE SCOPE

The main future scope of our project is to parse resumes from different applications and websites like LinkedIn, GitHub, Naukri.com, etc. In future,this system can be made more versatile in which wide ranges of psychometric tests will be added. As a future work, we can enlarge the resume dataset and improve the performance of the proposed system.

VII.SUMMARY

Previously, we have studied the existing system architecture and also our proposed system architecture. The existing system architecture has its own unique features and ideas behind it with advantages and disadvantages.In our proposed system we are going to deal with these type of disadvantages.In this proposed methodology we are using Optical Character Recognition(OCR) to extract the data from Resume.The main technique used here is Natural Language Processing and Ranking Algorithm which is helpful for ranking the resume according to the particular companies. Additionally our other goal is to extract the data from Social Media like Linkedin for applying jobs which will make the recruitment process easier getting quality applicants from various regions by avoiding unfair and discriminatory practice.

REFERENCES

[1]. *Vinaya R. Kudatarkar, ManjulaRamannavar, Dr.Nandini S. Sidnal* "An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes", 2015, IJIRAE.

[2]. *Satyaki Sanyal, Neelanjan Ghosh, SouvikHazra, Soumyashree Adhikary,* "Resume Parser with Natural language Processing", 2007,IJESC.

[3]. *Papiya Das, Manjusha Pandey, Siddharth Swarup Rautaray,* "A CV parser Model using Entity Extraction Process and Big Data Tools ", 2018,IJITCS.

[4]. *Ayishathahira and Sreejith,,* "Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing",International CET Conference on Control, Communication and Computing(IC4),2018..

[5]. *Dr.Parkavi A,Pooja Pandey,Poornima J,Vaibhavi G S,Kaveri BW,* "E-Recruitment System Through Resume Parsing, Psychometric Test and Social Media Analysis", 2019,IJARBEST.

[6]. *Nirali Bhaliya, Jay Gandhi, Dheeraj Kumar Singh,* "NLP based Extraction of Relevant Resume using Machine Learning ",2020,IJITEE.