Resume Parser Using Natural Language Processing Techniques

Shubham Bhor, Harish Shinde, Vivek Gupta, Vishak Nair, and Prof. Manasi Kulkarni

Department of Computer Engineering, PCE, Navi Mumbai, India - 410206

Abstract*— Job Requirement is considered as a major activity of human resource which is a very difficult task to find productive talent. Recruiters require large financial expenditure, time , and effort to select the best candidates. Similarly it is difficult for the job seekers to get the information about the opening or the requirements by the companies. This model proposes to extract information from the resume and ranking according to the preference of the associated company and requirements using Natural Language Processing(NLP) technique. It makes the hiring process easy and efficient. A resume contains various blocks within it and any respectable parser needs to pull out these blocks such as education,experience,project,address etc. Therefore, basically we will build a job portal where the users or the employee or job seekers would upload their resume and using the NLP technique the parser will pass the necessary information and generate a structured information resume according to the company needs also will rank the resumes according to the skills of the job seeker and requirements of the company. The job portal will consist of three modules i.e employee, admin and company and the admin. The admin module will be managing the company and employee modules.

Keywords—Extract Information,Resumes,NLP,Parser

1. 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. We aid are our purpose in the survey and we reviewed multiple research papers and we have narrowed down to these six research papers mentioned below.

2. 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, first stage was to get the data (resume) and convert them into structured format and then perform the analysis using deep learning technique. 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 helps 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 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. Resume Parser with Natural Language Processing:

In this their aim was to convert different formats of resumes to text and parse relevant information from there.They were able to scrape keywords from different social networking sites including Stack Overflow, LinkedIn, etc and find the similarity between them with which we could determine the genre of the resume.

F. 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 assessed for two qualitative aspects, coverage and comprehensibility; and these ratings are transformed into a comprehensive quality rating. All the three parameters were collectively measured 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.

2.1 Summary of Related Work

Table 1. Summary of literature survey

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 W 2019, IJARBEST	E-Recruitment System Through Resume Parsing,Psycho metric 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 Networks 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 languages ● Semantic mapping for skills, job boards,recruiters
5.	Satyaki Sanyal,Neelanjan Ghosh,SouvikHazra, Soumyashree Adhikary 2017,IJESC	Scrape keywords Conversion of different formats of resumes to text	● Scrape keywords ● Conversion of different formats of resumes to text
6.	Vinaya R. Kudatarkar, ManjulaRamannavar, Dr.Nandini S. Sidnal 2015 IJIRAE	An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes	● A simple text analytic ● Comprehensive quality rating ● Quality metric for resumes

3. 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 ease getting quality applications from various regions by avoiding unfair and discriminatory practice.

3.1 System Architecture

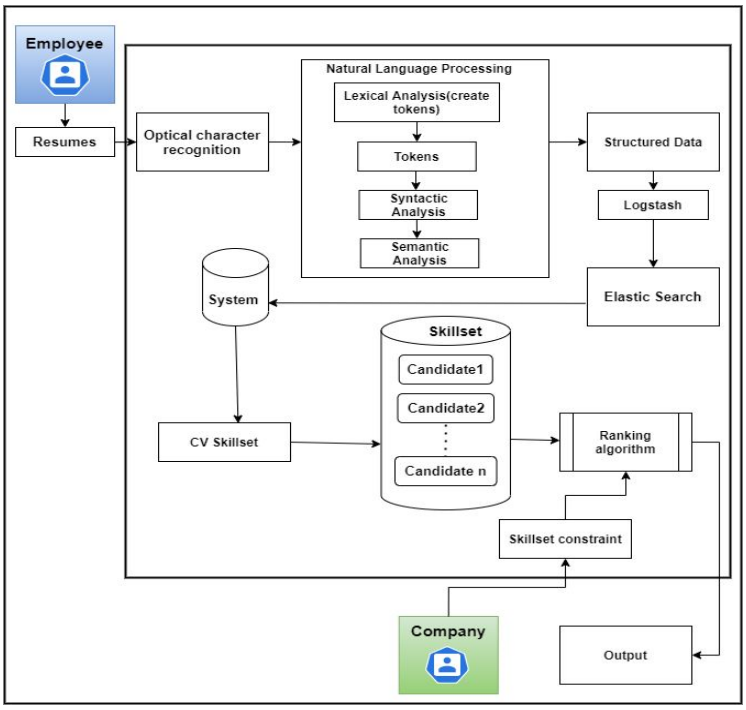

Fig 1. System Architecture

To use the Resume Parser system both the employee and company needs to provide the details 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:**Lexical analysis is the first phase of NLP parsing,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) **Semantic Analysis:**Semantic analysis checks the exact meaning of the text,
   sentences like “Sunny is raining” will be rejected by the English semantic analysis.
   4) **Named Entity Recognition (NER):**One of the problems with using the same NLP module for all the companies is the jargons 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 word. With NER, we can fine tune our NLP module to understand the real word objects from a domain[2][4][5]. For example, if a company wants to hire developers, they can use our system to differentiate between people who love “Python”- the programming language and the people who love “python”- the snake, based on the context in which the word is used.
5. The data obtained will be used for a dashboard which will contain graphs and pie charts based on the data in the resumes. An HR can use this dashboard to prepare his/her query based on the requirement that the respective company has. In order to make this dashboard, all the data will be fed into ElasticSearch. To make the system indifferent to many formats, the data to be fed will be fed by using LogStash.
6. Using queries inbuilt in ElasticSearch, the resumes will be scored and then they will be sorted according to the constraints that were provided by the HR.
7. Using queries inbuilt in ElasticSearch, the resumes will be scored and then they will be sorted according to the constraints that were provided by the HR. The individual traits of an applicant will be provided with

a proportionate boost based on the priority of the trait.

8. At the end, the cumulative score will be used to sort the applicants. A final sorted list of applicants will be displayed to the HR.

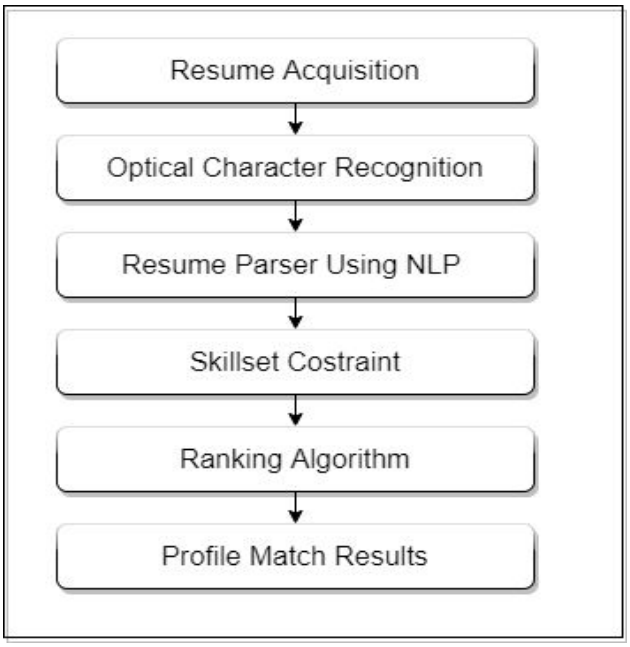

Fig 2. Overview

4. Conclusion and Future Scope

4.1 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.

4.2 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.

4.3 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 ease getting quality applicants from various regions by avoiding unfair and discriminatory practice.

ACKNOWLEDGMENT

We deeply express our sincere thanks to our Principal Dr. Sandeep M. Joshi and Dr. Sharvari Govilkar Head of the Department for encouraging and allowing us to present this work. It is our privilege to express our sincerest regards to our supervisor Prof. Manasi S. Kulkarni for the valuable inputs, able guidance, encouragement, whole-hearted cooperation and constructive criticism throughout the duration of this work.

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",Iternational 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.