IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 6, June 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006117 |

Resume Parser Using Natural Language Processing

Dr.N. Lakshmi Kanthan[1], Praveen Kumar.S[2], A.R.Tufail Ahamed[3], Vykunth .S [4], Yuvankarrtik A.B[5]

Assistant Professor, Dept. of C.S.E, Saranathan College of Engineering, Trichy, India[1]

U.G Students, Dept. of C.S.E, Saranathan College of Engineering, Trichy, India[2,3,4,5]

ABSTRACT: Everyday many firms go through a large number of resumes. This is no easy task for a human. An intelligent system is needed to extract all of the critical information from the unstructured resumes and convert them all to a similar structured format. They are ranked for a specific job. Our suggested strategy entails extracting information and data from a resume and rating it depending on the job description given by the employer. Natural Language Processing (NLP) is used for the extraction. The recruiting process is made easier and more efficient by parsing and rating resumes efficiently. A resume contains a lot of data, which any decent parser should be able to extract info and necessary data such as degree, experience, project, and address. A job portal has been built to facilitate both the applicants and employers. It accepts resumes from applicants and job descriptions from employers.

KEYWORDS: Resume Parsing, CV Parsing, Natural Language Processing, Skill Recognition, Resume Sectioning, Information Extraction, Text Processing, Job Description Matching

I. INTRODUCTION

The processing of human language by machines is known as Natural Language Processing (NLP) (Copestake, 2004). The study of human speech and text interpretation and translation dates back to the late half of the twentieth century. The 1954 IBM-Georgetown demonstration was the very first public demonstration of Machine Translation (MT) (Jones, 1994).

From then, NLP has expanded to include grammar, spelling correction, text categorization, question answering, summarization, speech recognition, and information extraction, among other applications. However, it is not only technology that has evolved over time, but so have the application areas and the quality of the analyzers.

The extraction of data from resumes is a difficult task. To start with, resumes are quite varied: employers receive various types of resumes from many countries. These resumes differ in terms of language, style, length, structure. In the United States, it is normal practice to condense all material onto a single page, however in India, applicants often give a multi-paged document including lifetime accomplishments. Second, resumes frequently contain unfinished phrases, like skills or talents. Because such sentences are syntactically different from sentences in ordinary texts, NLP must deal with them separately (Derczynski, Maynard, Rizzo, et al., 2015).

The main focus of this thesis is on extracting information from software engineering resumes, including competence and job experience. The thesis will examine the performance, applicability, and support of five tokenizers, as tokenization is a crucial stage in the information extraction process. It will also look into contemporary open-source solutions, their technology, and their benefits and drawbacks. Finally, it will offer a new approach to address current faults

II. RELATED WORK

[1] Resume parsing is a method of sorting and storing data from resumes using software. This allows hiring managers to quickly search for keywords, create precise search parameters to locate the best candidates, and organize resumes. Resume parsing can help ensure that you find the best candidate for the position. Benefits of a Good resume parsing application saves time, keeps resume organized, maintains accuracy, sets clear standards, spares cost, accepts many formats and encourages applicants. Therefore the system aims at providing an efficient resume parsing application

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006117 |

using natural language processing to ease the process for employers and employees [2] NLP based Extraction of Resume Using Machine Learning. Resume parsing is one of the main substance examination strategies. CV parser combines the candidate's resume with selection flow and thus systems moving toward CV's. This paper proposes a CV parser adjustment of the usage of artistic substance examination. The proposed CV parser interpretation isolates substances required in the enlistment methodology inside the associations. However this method is able to successfully extract the skills, the main problem with this approach is that it is more time consuming for handling lots of unstructured text and its ambiguity.[3]E- recruitment system through resume parsing, psychometric test and social media analysisThe first stage of this process is to collect data (resumes) and transform them into preferred format, followed by analysis utilizing deep learning techniques. The psychometric test is the second step, and text mining is utilized to generate scores for each candidate. In the third step, web scraping is used to gather extra information about candidates from various social media sites. The system then suggests relevant jobs to the candidates. In the fourth stage, the algorithm suggests abilities and prerequisites that were required in order to be hired in the desired company.The main merits of this application is that it extracts information from online giving extra information about the employees ,But this application includes many manual work lacking automation. [4] e - combination of neural networks and conditional random fields for resume parsing. The convolutional neural network (CNN), Bi-LSTM (Bidirectional Long Short-Term Memory), and Conditional Random Field are used in this paper to develop a system for resume parsing (CRF). The CNN Model is used to categorize various chunks of a resume. To tag various items, CRF and Bi-LSTM-CNN models were utilized for sequence tagging. For word embedding, a pre-trained Glove model is employed. A resume might be divided into three segments and 23 fields extracted using the proposed system.This application was able to give more accuracy in that period, but the demerits of this application turns out to be High Time Consuming Process than the other applications out there.[5] intelligent recruitment system using NLP. The study focuses on extracting data from resumes and doing the necessary analysis to convert the data into information that recruiters can use. As a result, the Resume Parser would assist recruiters in quickly identifying the most qualified prospects, saving them time and effort. Here the author uses Natural Language Processing Techniques to Parse the resumes, Using this method, This application could give more accuracy, but the main Demerits of this application is that it cannot handle Large Volumes of Resumessegmentation in data mining are: clustering and subgroup discovery. Because of some limitations and scope of the clustering techniques, it leads to further refinements in methodology in data mining.

III. PROPOSED WORK

In this proposed methodology we are using Natural Language Processing technique for parsing the resume and getting a structured resume. A common job portal for employers as well as employees is provided to apply for a specific job role and the employers can see all the ranked resumes. The resumes received would be parsed and ranked according to job requirements. Additionally our other goal is to provide a filtering feature in the admin side / employer's side, which can filter out all the resumes based on the job role and the job score.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006117 |

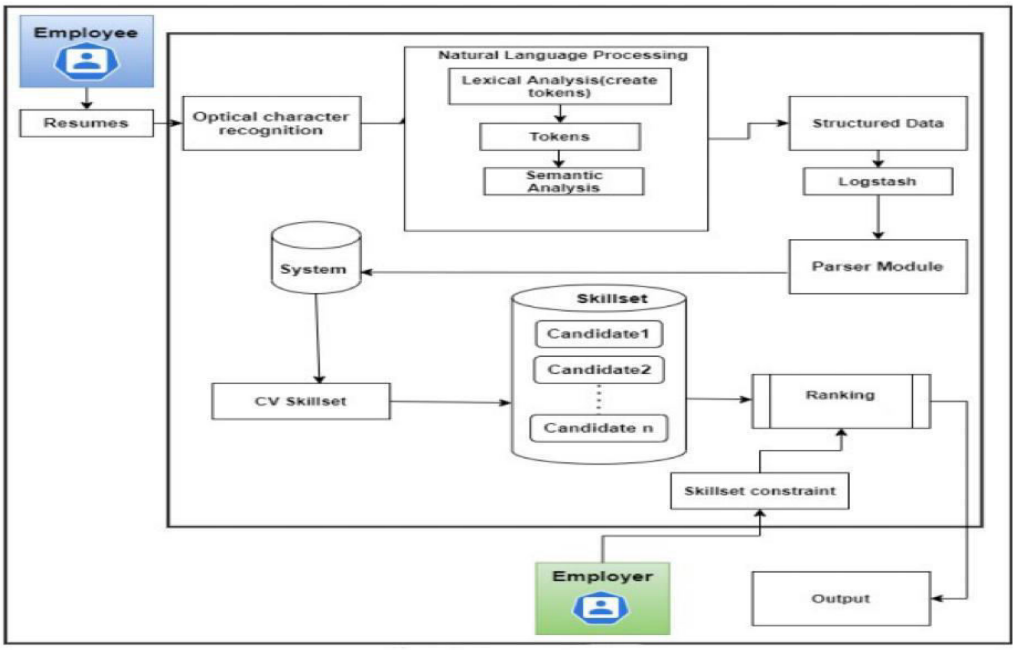

Figure 1: System architecture

To use the Resume Parser application both the employee needs to provide the detailed resumes while the employer needs to provide a detailed job description. The working of the system architecture is as follows:

1. A Web Portal is provided to the admin / employers side where they can be able to upload their job description and let the candidate choose the job role.
2. The Candidate also needs to use the web portal to create his account and upload the resume and choose the job role.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. The Scoring scheme is 50% weightage to the Skills Matching the Job Description, 25% for the Attributes matching, and the last 25% for the Recruiter Findings.
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.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006117 |

IV. RESULTS AND DISCUSSION

The result involves parsing the resume from /pdf/doc/rtf format into plain document and tokenized data entities. Resumes were ranked based on the Job Score. The Job score is calculated with 3 entities, 50% of Skills matching with the job description, 25% for the Attributes matching, and the last 25% for the recruiter findings. At last the Resumes are visualized using Tables, Pie chart, Barchart.

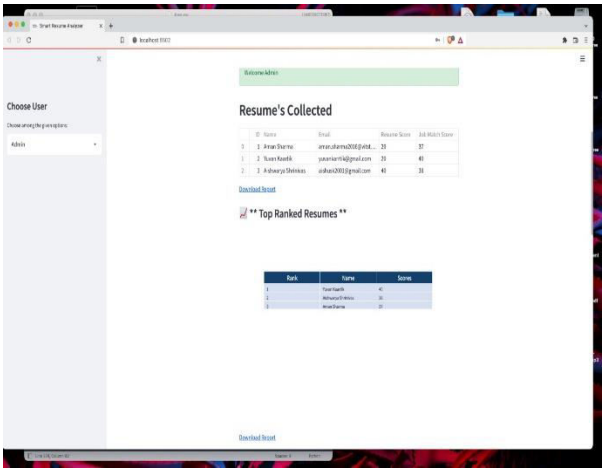

Figure 2: Employer's Page

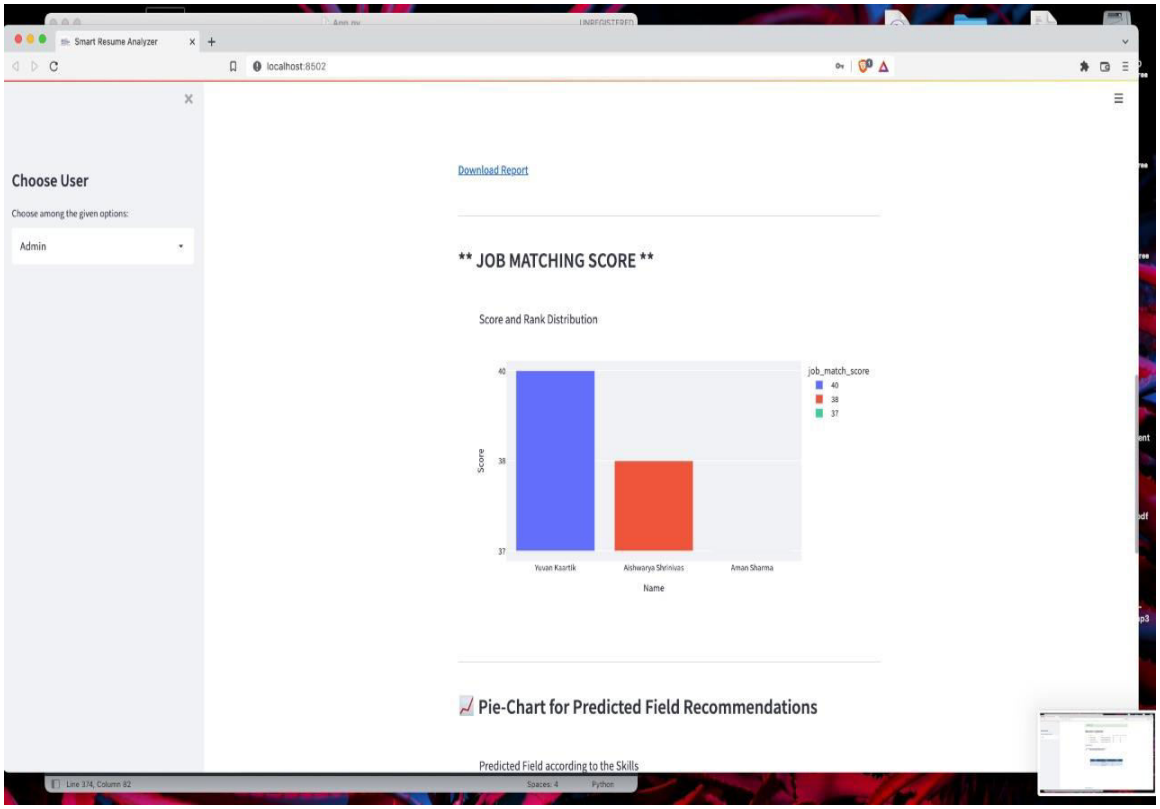

Figure 3: Bar chart of Ranked Resumes

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165 |

|| Volume 10, Issue 6, June 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1006117 |

V. CONCLUSION AND FUTURE WORK

In our Project, 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 the discriminatory practice in the process will be dampened. Based on the information in the form of technical skills, attributes, recruiter findings, the skills will be ranked in order and shown to user. The future scope of the project are as follows,

1. Improve the accuracy of the Model to extract more data from the resumes
2. Test The model with more resumes from all over the world
3. Make the resume Multi-lingual
4. Improve the Skills Dataset, to cover more in IT INDUSTRIES
5. To Make Resume Parser available as API.

REFERENCES

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

[2]. 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.

[3]. 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..

[4] Kulkarni A., Shivananda A., Kulkarni A. (2021) Creating a Résumé Parsing, Screening and Shortlisting System. In: Natural Language Processing Projects. Apress, Berkeley, CA.

[5] A. Sharma, S. Singhal and D. Ajudia, "Intelligent Recruitment System Using NLP," 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), 2021, pp. 1-5,

[6] Gugnani, Akshay, Vinay Kumar Reddy Kasireddy, and Karthikeyan Ponnalagu. "Generating unified candidate skill graph for career path recommendation." 2018 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2018. 66

[7] Jyoti Pandey, Abhijit Asati and Meetha Shenoy, Computational Operations and Hardware Resource Estimation in a Convolutional Neural Network Architecture

[8] Karan K V, Vedant Bahel, Ranjana R and Subha T, Transfer learning approach for analyzing attentiveness of students in an online classroom environment with emotion detection

[9] Kishore Ganesh, Sunita Singhal and Hemlata Goyal, Combining Statistical and Semantic approaches for Effective Information Retrieval.

[10] Kulkarni A., Shivananda A., Kulkarni A. (2021) Creating a Résumé Parsing, Screening and Shortlisting System. In: Natural Language Processing Projects. Apress, Berkeley, CA.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details