

RESUME PARSER FOR SUMMERIZATION USING NLP

Siddhi Kashid
Computer Science Technology
UMIT, SNTD
Mumbai, Maharashtra-400049
Email:siddhiyuvraj791@gmail.com

Advaita Kulkarni
Computer Science Technology
UMIT, SNTD
Mumbai, Maharashtra-400049
Email:imadvaita@gmail.com

Mansi Patil
Computer Science Technology
UMIT, SNTD
Mumbai, Maharashtra-400049
Email:mansip334@gmail.com

Prof. Iffat Kazi
Computer Science Technology
UMIT, SNTD
Mumbai, Maharashtra-400049

Abstract—Resume parsing is the process of extracting relevant information from a job seeker's resume. With the increasing number of job applications and resumes, there is a need for an automated system that can extract and analyse relevant information efficiently. Natural Language Processing (NLP) techniques are used to build such systems that can accurately identify and extract information from resumes. In this project, we propose a resume parser that uses NLP techniques to extract information from resumes. The system will analyse the text in the resume and extract relevant information such as name, email, phone number, education, skills. The system will also identify the different sections in the resume and extract information from each section separately. We will also use NLP techniques to improve the accuracy of the system in identifying and extracting information from resumes. The system will be built using Python and various NLP libraries such as NLTK. Our system will provide a simple and user-friendly interface that allows users to upload their resumes and obtain an organised report of the extracted information. The system will help recruiters to quickly screen and shortlist candidates based on their skills, education thus saving time and effort in the recruitment process.

Index Terms—resume parser; text classification; job title classification

I. INTRODUCTION

The objective of the Resume Parser project, which uses Natural Language Processing (NLP), is to simplify the recruitment process by automatically extracting important information from resumes without requiring manual review. By using an NLP model, it is possible to extract various types of information from resumes, including skills, education, contact details, and more, regardless of the resume's format or structure. To create an effective NLP model for resume parsing, it is essential to train it on a relevant dataset. Resumes are a great example of unstructured data. Each resume has its unique style of formatting, has its own data blocks, and has many forms of data formatting. This makes reading resumes hard, programmatically. Here is a simple resume parser used for extracting information from resumes. The foundation of this project is a resume automation system. Concerning the project, there will be an admin panel at first, into which the administrator must initially log in. Following that, there will be a section for uploading gathered CVs, which the admin will

manage. After that, Regex, NLTK, and Spacy's phrase matcher will extract necessary information such as Name, Address, Email, phone number, Nationality, Skills [Hard Skills, Soft Skills], Education, Experience, Experience Year, Languages etc. Finally, the extracted information is saved in the database by admin if necessary. People with diverse personalities come from a variety of fields and backgrounds. In the same way, their resume writing style varies. They've worked on a variety of projects, and each of them has unique writing style. As a result, each resume is unique. Some people work in the human resources department. They will have to go through hundreds of resumes on the internet. Executives summarize the resume, enter specific information into their database, and then call the applicant for job counselling after obtaining the resume. An executive spent around 10-15 minutes on each resume, summarizing it and entering the information into the database. This project will help in the automation of this procedure. Reason Behind Implementing NLP Natural language processing software screens information to determine the semantics behind the words on a page. In other words, it uses artificial intelligence to detect the true meaning behind what people have written. When job applicants apply for a position, they can describe exactly the same information in a number of ways. For example, the 'Work History' section of a CV could be called 'Job History', 'Previous Employment' or any of a number of other titles. Our natural language processing AI analyses what is written to determine what the writer meant and sort the data into the appropriate category. This means that we can gain accurate data even from information that has a high level of variability.

II. OBJECTIVES OF THE STUDY

- For information extraction, NLP model will be configured and reconfigured.
- The unstructured data will be transformed into structured form.
- To take the help of the cutting-edge and latest NLP technology to enhance their business processes.
- To extract the required information about candidates without having to go through each resume manually.

- To replace slow and expensive human processing of resumes with extremely fast and cost-effective software

III. RELATED WORK

[1] "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.

[2] "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

[3] "Combination of Neural Networks and Conditional Random Fields for Efficient Resume Parsing:" Abstract: 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

[4] "A CV Parser Model using Entity Extraction Process and Big Data Tools: " Abstract: 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.

[5] "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.

IV. EXISTING SYSTEM

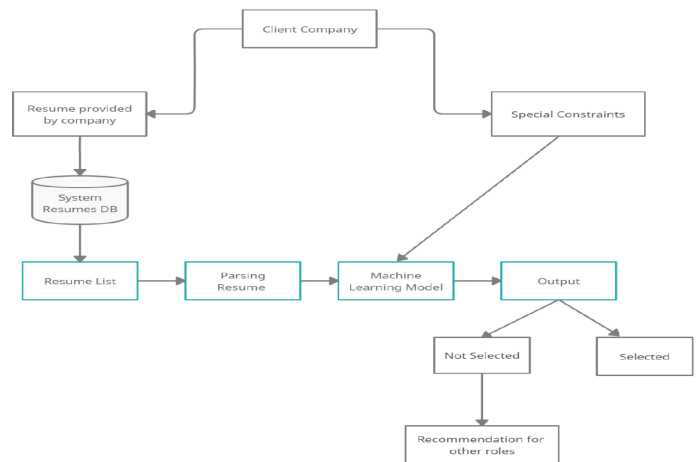


Fig. 1. Existing System

Humans are prone to errors and may have personal biases that skew results when performing repetitive tasks such as evaluating resumes each one individually and other textual data. In this situation, NLP-powered solutions can be taught to grasp the needs and requirements of the firm in just a few steps. As a result, once they are up and running, they perform better in terms of precision

V. PROPOSED MODEL

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

- **Get Resume from the user** Here, user will upload the resume Following the submission of the resume, there will have two sub-models: resume parsing and information extraction
- **File upload and Parsing system** This system is at the center of the overall project. In this sub system the admin will upload the resume and go for further processing. The

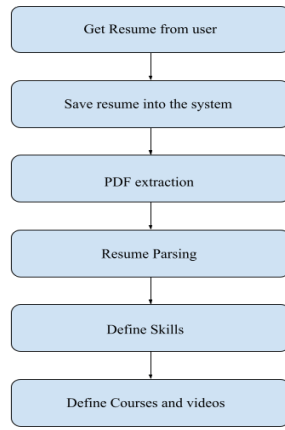


Fig. 2. Proposed System

parser subsystem will use different libraries to transform the submitted unstructured resume to a structured format. This will make it much easier to examine, analyze, and grasp the data.

- **Information extraction system** After parsing the data, we employ the NLTK, Spacy phrase matchers, regex to extract the essential information. The required extracted data will be saved in JSON format. Finally, the extracted data or dump JSON file is stored in a MySQL database for further usage if it is needed.

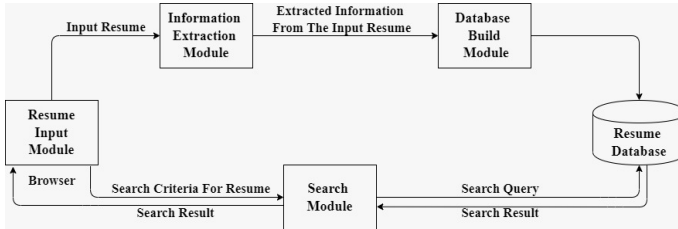


Fig. 3. Workflow of Proposed System

VI. SYSTEM DESIGN AND IMPLEMENTATION

Workflow of NLP

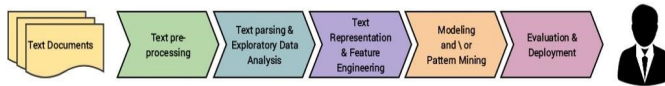


Fig. 4. NLP Workflow

Natural language processing comprises a wide variety of methods for analyzing human language, based on machine learning techniques as well as rules-based and computational approaches. Tokenization, lemmatization and stemming, parsing, part-of-speech tagging, language identification are some basic NLP tasks. NLP tasks, in general, break downs the

language into smaller, essential components, attempt to comprehend links between the pieces, and then examine how the components combine to form meaning.

The procedures given above represent the standard workflow for an NLP task. The initial stage is generally text wrangling and pre-processing on the collection of documents. Then there's parsing and some basic exploratory data analysis. The next stage is to represent text using word embeddings and then do feature engineering. Following that, we must select a model based on whether we are dealing with a supervised or unsupervised learning scenario. The final step of any ML workflow is model testing and deployment.

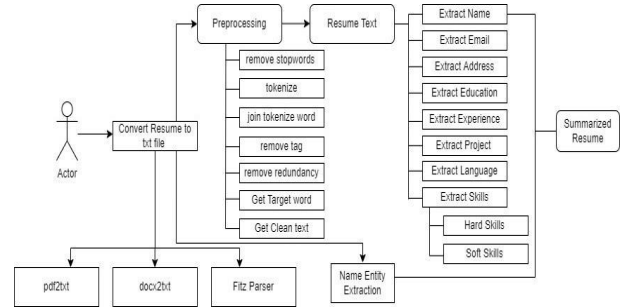


Fig. 5. System Architecture

VII. MODULES AND LIBRARIES

1) Pyresparser Features

- Extract name
- Extract email
- Extract mobile numbers
- Extract skills
- Extract total experience
- Extract college name
- Extract degree
- Extract designation
- Extract company names

2) **Pandas** : Open source data analysis and manipulation tool

3) **base64** : A binary to text encoding scheme

4) **streamlit** : It allows you to create a stunning-looking application with only a few lines of code

5) **time** : Provides various time-related functions

6) **datetime** : To work with dates and times

7) **pyresparser** : A simple resume parser used for extracting information from resumes

8) **pdfminer3** : Tool for extracting information from PDF documents

9) **io** : Allows us to manage the file-related input and output operations

10) **random** : Returns a random number between the given range

11) **pil** : Adds image processing capabilities to your Python interpreter

- 12) **pymysql** : Interface for connecting to a MySQL database server from Python
- 13) **pafy** : Python library to download YouTube content and retrieve metadata
- 14) **plotly** :Interactive, open-source plotting library

VIII. RESULT ANALYSIS

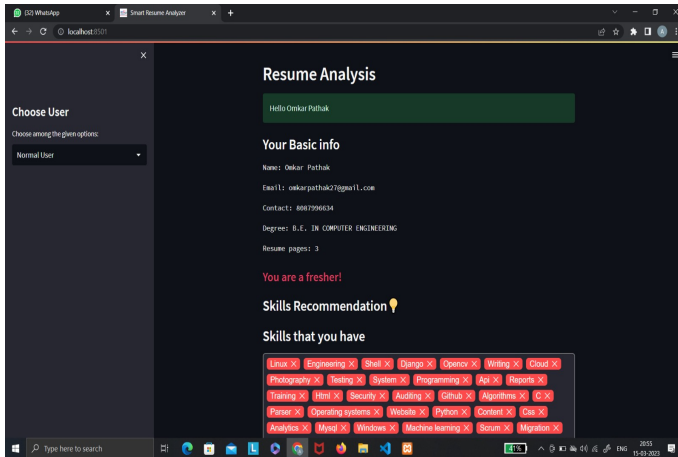


Fig. 6. User Side

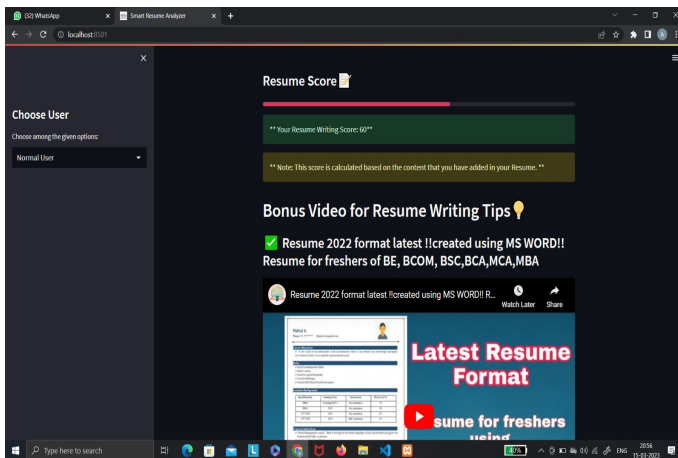


Fig. 7. Resume Analysis

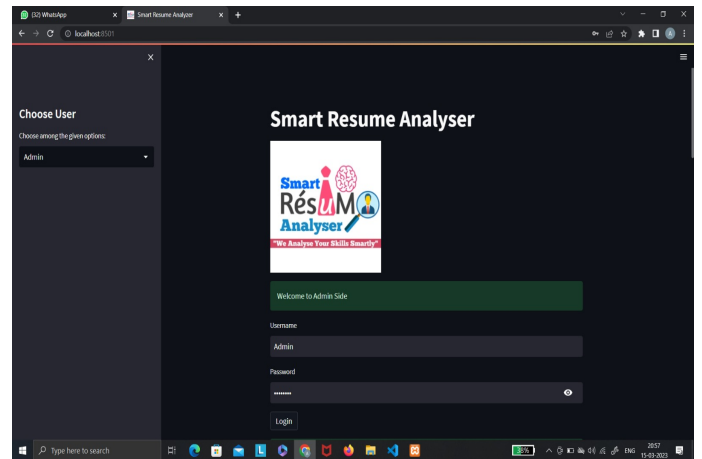


Fig. 8. Admin Side

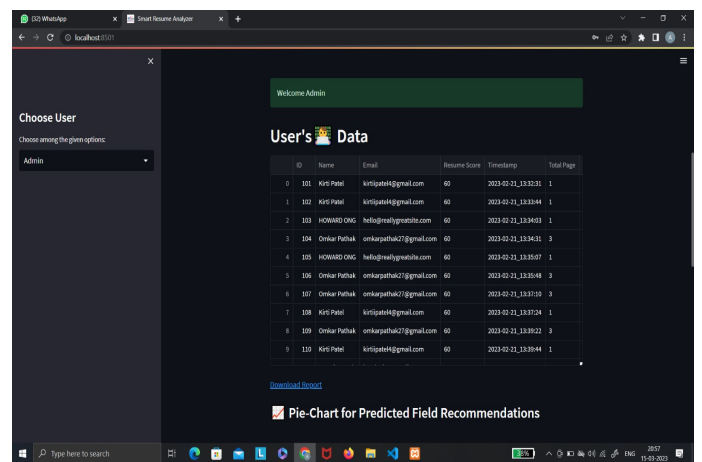


Fig. 9. User Data

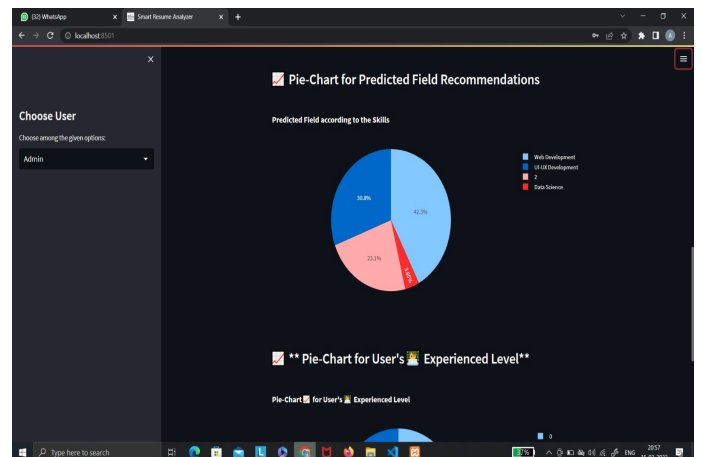


Fig. 10. Recommendation

IX. CONCLUSION

A normal resume is a compilation of information about a person's work experience, academic background, qualifica-

tions, and personal details. These elements might be present in a variety of ways or not at all. It's difficult to keep up with the jargon used in resumes. A resume is made up of corporate names, institutions, degrees, and other information that can be written in a variety of ways. It will take time to review all the resume by an individual.

Machine works faster than human and their accuracy to do any task was also good. Therefore, I have made a system which includes machine learning that extract the important information from resumes within a minute or less than a minute. The hiring individual can use this system for hiring any individual.

X. SCOPE / FUTURE WORK

- 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.
- Improve the accuracy of the model to extract all the data.
- Test the model further and make it work on resumes from all over the world.
- make the resume parser multi-language.
- Make Resume Parser available via API.
- Make resume parser having feature of audio to listen rather than reading and seeing it.

XI. ACKNOWLEDGEMENT

This project has helped us in acquiring tremendous knowledge in various aspects. We are highly indebted to our project guide Prof. Iffat Kazi for her guidance and constant supervision as well as for providing necessary information regarding the project. We would like to express our sincere gratitude to Dr. Shikha Nema, Principal (Usha Mittal Institute of Technology) for her valuable encouragement and insightful comments. We would also like to thank to all the teaching and non-teaching staff for their valuable support but not the least we would like to thank our parents and friends.

REFERENCES

- [1] Vinaya R. Kudatarkar, Manjula Ramannavar, Dr. Nandini S. Sidnal "An Unstructured Text Analytics Approach for Qualitative Evaluation of Resumes", 2015, IJIRAE.
- [2] Satyaki Sanyal, Neelanjana Ghosh, Souvik Hazra, 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