Project Synopsis

on

Empowering Careers with NLP-Driven Resume Analysis

Submitted as a part of course curriculum for

Bachelor of Technology

in

Computer Science

Submitted by

Sanyam Bansal (2000290100135)

Shweta Kushwaha (2000290100155)

Tanisha Porwal (2000290100166)

Tanzeem Irfan (2000290100169)

Under the Supervision of

Ms. Swati Sharma

Assistant Professor in Computer Science & Engineering Department

KIET Group of Institutions, Ghaziabad

Department of Computer Science

Dr. A.P.J. Abdul Kalam Technical University

2023-2024

TABLE OF CONTENTS

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature of Students
Name: Sanyam Bansal, Shweta Kushwaha, Tanisha Porwal, Tanzeem Irfan
Roll No: 2000290100135, 2000290100155,2000290100166,2000290100169
Date: 05/03/24

CERTIFICATE

This is to certify that Project Report entitled "Empowering Careers with NLP-Driven Resume Analysis" which is submitted by Tanisha Porwal, Sanyam Bansal, Shweta Kushwaha, Tanzeem Irfan in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Date: 05/03/24
Ms.Swati Sharma

Assistant Professor
Department of Computer Science & Engineering

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the synopsis of the B.Tech. Major Project undertaken during B.Tech. Fourth Year. We owe a special debt of gratitude to Assistant Professor Ms. Swati Sharma, Department of Computer Science & Engineering, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for her constant support and guidance throughout the course of our work. We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Head of the Department of Computer Science & Engineering, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature:
Date: 05/03/24
Name: Sanyam Bansal, Shweta Kushwaha, Tanisha Porwal, Tanzeem Irfan
Roll No: 2000290100135, 2000290100155,2000290100166,2000290100169

ABSTRACT

In a company's or organization's recruiting process, there are typically multiple candidates applying for a single opening. However, manually reviewing each resume can be time-consuming and challenging for recruiters due to the unique formats and sections of each applicant's resume. To streamline this process and minimize recruiter efforts, we propose a system that employs a CV parsing algorithm using Natural Language Processing (NLP).

The objective of this project is to develop an algorithm that parses resumes, extracts relevant fields, and creates candidate profiles based on the skills mentioned in the resume. The system consists of two main components: Job Seekers and Recruiters. Job seekers upload their resumes in various formats, such as PDF or DOC, and the parser analyzes these resumes for field extraction. After analysis, the system assigns a ranking to each resume and provides suggestions to improve the resume's rank, such as adding required skills, courses, or additional fields. The rankings are stored in a database, allowing recruiters to identify the most suitable candidates for a particular role efficiently. By utilizing AI, we aim to automate the manual tasks involved in the recruiting process, benefiting both the industry and the individuals involved.

Introduction

1.1 INTRODUCTION

Organization has to deal with thousands or sometimes lakhs of resumes for single job profile. So, it is not possible for recruiters to analyze them one by one. Also, we know that these resumes do not have proper formats. Each resume might be different from others in sections like format, fields, patterns, etc. Though only some information is common or required for eligibility like individual's name, email, education, skills, year of experience, projects, internships, etc. Recruiters only want to look upon this fields. It would be ideal if a computerised intelligent machine could extract all the important information from unstructured resumes and redesign it in a regularly structured layout that could subsequently be ranked for a particular function.

Hundreds of emails from people who send their resumes to apply for a job are received by multinational corporations. The current challenge is knowing which resumes should be sorted and shortlisted based on the limits. This resume scanner saves you time and reduces the amount of research you must do. Each set includes information about a person's background, professional experience, or educational background. Second user of this system are job seekers. They will apply for role they want. How would they know that where their resume stand. So, our system will analyze the field which are extracted from resumes and according to them system will suggest some areas where resume can be improve like required skills to be added in resume, quality of resume, enhancements required.

1.2 PROBLEM STATEMENT:

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.

1.3 OBJECTIVES:

• **Saves time:** Resume parsing software can help you quickly sort resumes by automatically searching for relevant keywords and qualifications instead of manually searching each resume.

• **Keeps resumes organized:** Resume parsing software can also store resumes for easy retrieval, keeping them neat and accessible whenever you need them. This allows a hiring manager to stay organized and efficient, which is crucial when a job posting attracts hundreds or thousands of candidates.

• **Maintains accuracy**: Using software to sort resumes can ensure you do not miss any desired keywords in applications. While manually scanning resumes for these words and qualifications can be effective, resume parsing software can review multiple resumes and offer quick results with little chance of error.

• **Sets clear standards:** Because parsing software highlights a qualification you specify, your hiring team can quickly identify the skills, qualifications and characteristics they want in an ideal candidate and maintain that standard throughout the hiring process.

• **Spares cost:** Using resume parsing software instead of manually scanning resumes can conserve company resources. Employees can spend their time more efficiently on tasks that can't be automated, such as

interviewing applicants with qualified resumes.

• **Encourages applicants:** With resume parsing, your company can feasibly sort through hundreds of applications per job post, enabling your company to handle numerous applications and giving hiring managers more choices in candidates.

• **Accepts many formats:** Professionals can use resume parsing software on several different types of resume formats, which allows your applicants more flexibility when crafting their resumes and avoids strict guidelines.

1.4 SCOPE:

The project entails building a sophisticated resume parsing and analyzing software utilizing Natural Language Processing (NLP) techniques. Its core features include parsing essential details from resumes, analyzing them to offer improvement suggestions, ranking parsed resumes, and managing a database for easy recruiter access. User-friendly interfaces cater to job seekers and recruiters alike. The system's scalability and continuous algorithm refinement ensure efficiency in handling diverse resumes. Ultimately, it aims to automate recruitment tasks, minimize manual efforts, and offer valuable insights for enhanced decision-making. The software's scalability ensures it can efficiently manage large volumes of resumes, catering to organizations with high recruitment demands. Furthermore, continuous algorithm refinement and updates guarantee the system's adaptability to evolving resume formats and content structures, ensuring its accuracy and reliability in parsing and analyzing resumes over time.

Literature Review

[1] discusses using content analysis techniques to extract important information from resumes for recruitment purposes. The system proposed in the paper aims to automate the process of parsing and extracting data from CVs, ultimately saving time for employers and making the recruitment process more efficient. However, the paper does not provide detailed information about how the system is implemented or evaluated, suggesting that there is room for future research to explore these aspects further.

[2] presents a new method for parsing and analyzing resumes by leveraging big data tools for entity extraction. The system utilizes natural language processing (NLP) techniques in conjunction with the R programming language for preprocessing tasks such as cleaning, tokenization, part-of-speech (POS) tagging, and transformation of the data. It emphasizes the use of Hadoop MapReduce for handling large datasets efficiently.
The paper underscores the significance of accurate models and statistical techniques in ensuring effective text analysis and entity extraction from resumes. By employing NLP and big data tools, the proposed approach aims to enhance the parsing and analysis of resumes, potentially improving the efficiency of recruitment processes.

[3] describes a system that employs machine learning to automatically extract detailed segment information from resumes. It involves several steps including feature extraction, data preprocessing, heading prediction, and rule-based segmentation. These processes result in the creation of a dictionary to organize and store section information. Manual data labeling by experts is used to verify and ensure the accuracy of the segmentation process.

[4] presents a text mining and machine learning-based system designed to rank candidates according to their suitability for a specific job position. By categorizing resumes based on rating scores, the system aims to enhance the efficiency of the recruitment process. It assists companies in identifying and shortlisting candidates whose skills and qualifications closely match the requirements of the job, thereby streamlining the candidate selection process.

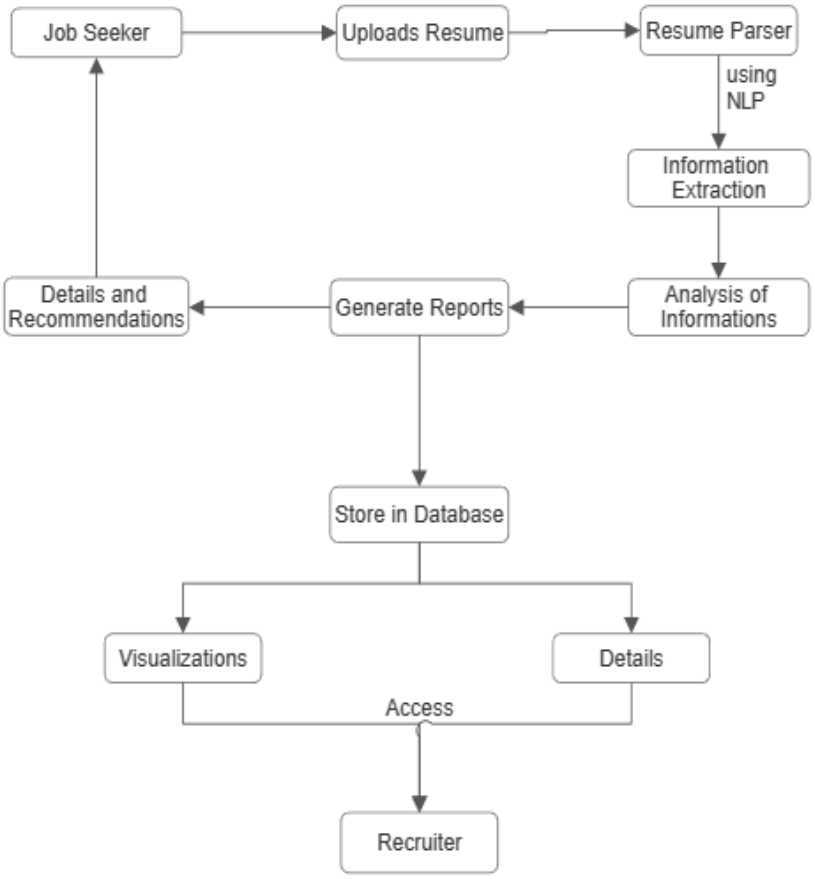
[5] discusses a system that utilizes natural language processing (NLP) to parse information from resumes. It then clusters these resumes into sectors based on keywords. The system's goal is to convert unstructured resumes into a structured format, presenting relevant information to employers according to the job description. This approach streamlines the resume evaluation process by presenting candidates' qualifications and experiences in a format that aligns with the specific requirements of the job

[6] introduces a Named Entity Recognition (NER) model designed specifically for classifying text within the context of resume parsing. This model utilizes natural language processing (NLP) techniques along with a hybrid approach combining the K-means algorithm and the Firefly algorithm.

By leveraging semantic relations between entities extracted from resumes, the system efficiently matches resume profiles. This approach aims to improve the accuracy and effectiveness of resume parsing by considering the relationships between entities, thereby enhancing the overall quality of candidate matching and selection processes.

Proposed Methodology

3.1 Flowchart



3.2 Pre-processing steps

• Removing Stop words
• Lower casing
• **Tokenization**: Tokenization is the process of breaking down text into smaller parts, called tokens. The goal is to represent text in a way that's meaningful for machines without losing its context.
• **Lemmatization**: Lemmatization is a text pre-processing technique used in natural language processing (NLP) models to break a word down to its root meaning to identify similarities.

Steps in Parsing:

a) Extract the text from pdf
b) Extract Text from Doc files
c) Detect the file extension
d) Extract the entities
e) Extract the email
f) Extract Name from Resume
g) Extract Mobile Number
h) Extract Skills from the resume
i) Extract Technical Skills
j) Extract Education and Year
k) Extract Experience

Technology Used

Natural Language Processing (NLP): Parsing and analyzing resumes for field extraction, identifying skills, and suggesting improvements**.**

Python: Developing the web application and implementing NLP algorithms.

NLTK and Spacy Libraries: These are used for text processing and NLP tasks such as stop word removal, tokenization, and part-of-speech tagging.

Regular Expressions: Text cleaning and extraction of specific information from resumes.

Regular Expressions: Text cleaning and extraction of specific information from resumes.

Database Management: Storing parsed resume data for recruiters' access and decision-making.

Conclusion

Our system is able to parse most of the resumes in accurate way. Using of Natural Language Processing (NLP) for this system was best suitable option. As here we are dealing with raw data and for the processing of this data NLP is suitable. The power of computers is increasing day by day, hence most of the computers are now able to perform machine learning/deep learning tasks. Therefore, there is rapid increase in automation in every field. Our system fulfills one of areas where automation requires. When there are lakhs of applicants applying for same profile in an organization, there will be need of such software systems. As automation continues to revolutionize various industries, our system represents a crucial advancement in streamlining the recruitment process and meeting the demands of modern organizations.

References:

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

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

[3] B. Gunaseelan, Supriya Mandal, and V. Rajagopalan, "Automatic Extraction of Segments from Resumes using Machine Learning", IEEE 2021.

[4] B. Kelkar, R. Shedbale, D. Khade, P. Pol, A. Damame, "Resume Analyzer Using Text Processing", May 2020 ISSN NO: 0377-9254.

[5] Satyaki Sanyal, Souvik Hazra, Soumyashree Adhikary, Neelanjan Ghosh , Resume Parser with Natural Language Processing" ,March 2017.

[6] Dr. Naveenkumar Jayakumar, Akshay Ramchandra Patil, Dr. Shashank

Joshi, Dr. Prasanth Narayanan, Dr. Saurab Saoji ,A Approach to parse resume and recommend Job Modified K-mers and Firefly algorithm (FFA) in Resume Parsing, 2023 DOI: 10.5281/SJIS.7759409.