# TRIBHUWAN UNIVERSITY

# INSTITUTE OF ENGINEERING

# PURWANCHAL CAMPUS



A THIRD YEAR MINOR PROJECT PROPOSAL

ON
"RESUME PARSER"

SUBMITTED TO:

DEPARTMENT OF COMPUTER AND ELECTRONICS

SUBMITTED BY:

DINESH NEPAL (PUR077BCT033)

BISHWAS RAJ KOIRALA (PUR077BCT029)

BHUWAN OJHA(PUR077BCT020)

KUSHAL POUDEL (PUR077BCT043)

4 JANUARY,2024

# ACKNOWLEDGEMENTS

# ABSTRACT

The Resume Parser project represents a breakthrough in simplifying the resume screening process through advanced Natural Language Processing (NLP) techniques, specifically utilizing Spacy for Named Entity Recognition (NER).Spacy is an open-source software library for advanced natural language processing.

The primary goal of this project is to streamline the extraction of crucial information from resumes, providing a concise and structured summary.This technology aims to enhance the efficiency of resume evaluation, enabling quick and informed decision-making for recruiters and human resources professionals.

# Contents

## 8  EXPECTED OUTCOMES                                                     15

()

# List of Figures

()

# 1  INTRODUCTION

## 1.1  INTRODUCTION

In the ever-evolving landscape of talent acquisition and recruitment, the ability to swiftly and accurately evaluate a multitude of resumes is paramount. The Resume Parser project emerges as a technological milestone, introducing a groundbreaking solution to streamline the traditionally intricate process of resume screening.

Harnessing the power of advanced Natural Language Processing (NLP) techniques, with a specific emphasis on the formidable Spacy library for Named Entity Recognition (NER), this project seeks to redefine the efficiency and precision of extracting essential information from resumes. The essence of our endeavor lies in providing a comprehensive, yet concise, summary of resumes, covering critical facets such as names, skills, educational qualifications, and professional experiences.

This project's significance is underscored by its potential to revolutionize the hiring process, offering recruiters and human resources professionals a tool that not only expedites resume evaluation but also enhances decision-making through intelligent entity extraction. As we delve into the intricacies of the Resume Parser, we embark on a journey to reshape the landscape of resume analysis, setting new standards for efficiency and effectiveness in the realm of talent acquisition.

## 1.2   PROBLEM STATEMENT

The manual parsing of resumes is time-consuming and prone to errors. HR professionals often spend significant hours extracting relevant details from resumes, leading to delays in the hiring process. Our project seeks to eliminate this bottleneck by automating the extraction process.

## 1.3   OBJECTIVES

1. Develop a resume parser using ML NER to automate the extraction of key information from resumes.

2. Enhance recruitment efficiency by accurately identifying and categorizing entities, streamlining the candidate screening process

## 1.4   PROJECT SCOPE

**Resume Ranking System:**
Utilize the extracted summary data to develop a dynamic resume ranking system for HR recruiters. By incorporating relevant factors such as skills, education, and experiences, the system can intelligently prioritize and present resumes, aiding recruiters in identifying top candidates efficiently

**Job Recommendation System::**
Leverage the skillsets extracted from resumes to create a job recommendation system. This system can match candidate profiles with open positions, providing recruiters with tailored suggestions based on the candidate's expertise, thereby streamlining the job matching process.

**Customized Interview Questions:**
Utilize the extracted information to generate customized interview questions tailored to the specific skills and experiences of each candidate. This ensures that the interview process is focused and relevant, providing a deeper understanding of the candidate's suitability for the role.

**Career Pathway Recommendations:**

Develop a feature that suggests potential career pathways for candidates based on their skills and experiences. This could include recommending additional certifications, training programs, or lateral moves within the organization, promoting career development and growth.

# 2 LITERATURE REVIEW

## 2.1 RELATED THEORIES

**Natural Language Processing:**

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

**Named Entity Recognition**

Named-entity recognition (NER) also known as named entity identification, entity chunking, and entity extraction is a sub-task of information extraction that seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

**Spacy**

Spacy is an open-source natural language processing (NLP) library designed for efficient and practical implementation of various NLP tasks. It is written in Python and provides pre-trained models and tools for processing and analyzing human language.

## 2.2 RELATED WORK

1. Narendra G.O. and Hashwanth S. prepared a research paper on International Journal of Advanced Research in Science, Communication and Technology(IJARSCT) named "Named Entity Recognition based Resume Parser and Summarizer" on 1 March, 2022 [1]

2. Dipti Suhas Chavare and Archana Bhaskar Patil wrote a research paper on Grenze International Journal of Engineering & Technology (GIJET) named "Resume Parsing using Natural Languauge" on 2023 [2]

3. Vrinda Mittal, Priyanshu Mehta, Devanjali Relan and Goldie Gabrani prepared a research paper on Journal of Statistics and Management Systems named "Methodology for resume parsing and job domain prediction" on 2020 [3]

# 3 REQUIREMENT ANALYSIS

## 3.1 FUNCTIONAL REQUIREMENTS

Functional requirements describe what a system is supposed to accomplish. They define the specific features, capabilities, and functionalities that the system must have to meet the needs of its users. Functional requirements focus on "what" the system should do.These are our functional requirements of our projects.

1. **File Upload:**

   The system allows users to upload resumes in popular formats such as PDF and DOCX so that the user could get the summary of the resume.

2. **Resume Parsing**

   The system categorizes entities such as names, skills, educational qualifications, and experiences from the resume using the trained model that is trained from different datasets collected from various sources.

3. **Resume Summary Generation:**

   The system generates a structured summary for each uploaded resume, presenting extracted information in a clear and organized manner.

## 3.2 NON-FUNCTIONAL REQUIREMENTS

Non-functional requirements define the qualities or attributes that describe how the system should behave. They focus on characteristics such as performance, usability, security, and reliability. Non-functional requirements provide constraints and criteria for the system's overall behavior. Some of the non-functional requirements of our pojects are as follows:

1. **Availability**

   The system is available for use 24 hours without planned downtime.

2. **User Friendly**

   The system has easy to use user interface, requires minimal training for users.

3. **Compatibility:**

   The system is compatible with common browsers, operating systems, and devices.

4. **Data Backup and Recovery:** The system ensures data recovery in case of system failures or data loss.

# 4 METHODOLOGY

Methodology refers to a set of principles, practices, and procedures followed in a particular discipline, such as project management, software development, or research. It provides a structured approach to planning, executing, and controlling processes to achieve specific goals.

We have chosen to adopt the incremental model for our project, leveraging its distinct features to ensure a systematic and flexible development process.In this approach, we plan to break down the development of our resume parser into smaller, manageable parts called increments.

## 4.1 REQUIREMENT ANALYSIS

We're figuring out exactly what our resume parser needs to do. Talking to recruiters, we're finding out what info is important—like names and skills. We're making a detailed plan (like a to-do list) that prioritizes these features, so we know what to work on first.

## 4.2 DESIGN & DEVELOPMENT

Following requirement analysis, we're building the resume parser step by step. We're creating a design that can handle different types of resumes and making it easy for people to use. We start with the basics, like getting names and skills right, and then add more cool features based on what recruiters need. It's like building blocks—one piece at a time.

## 4.3 TESTING

The testing phase ensures the reliability and functionality of each increment developed. Comprehensive testing, including unit testing, integration testing, and user acceptance testing, is conducted to identify and address any defects or issues. Test cases and reports are generated, providing verified and validated increments ready for deployment.

## 4.4 IMPLEMENTATION

In the implementation phase, the resume parser is deployed into the production environment. Each increment is deployed systematically, and training materials are provided to end-users. Ongoing support and maintenance plans are established to address any post-implementation issues, ensuring a smooth transition and successful deployment of the resume parser. This phase marks the culmination of the project, with the delivered system meeting the outlined objectives and requirements.

# 5 TOOLS AND TECHNOLOGY USED:

To achieve the goals outlined in our resume parser project, we will leverage a combination of advanced tools and techniques. The integration of these elements will ensure the development of a robust and efficient system.

1. **Machine Learning and Named Entity Recognition (NER):**

   <u>Tool:</u> Python-based library Spacy

   <u>Technique:</u> Implementing pre-trained NER models for accurate extraction of entities such as names, skills, and experiences from resumes. Fine-tuning the models to enhance performance on specific resume data.

2. **Front-End Development:**

   <u>Tool:</u> React.js

   <u>Technique:</u> Creating an intuitive user interface that allows users to interact seamlessly with the resume parser. Incorporating features like drag-and-drop functionality for resume uploads and real-time feedback on extracted information.

3. **Back-End Development:**

   <u>Tool:</u>Django (Python web framework)

   <u>Technique:</u> Developing a robust backend that handles data processing, storage, and communication with the front end. Integrating the NER models into the backend to ensure efficient parsing of resume content.

4. **Database Management:**

   Tool: Github

   Technique: Storing parsed data in a relational database for easy retrieval and analysis. Ensuring data security and integrity through proper database design.

5. **Database Management:**

   Tool: Github

   Technique: Utilizing version control to manage collaborative development, track changes, and facilitate seamless integration of new features. GitHub or GitLab will be used as the hosting platform.

6. **Presentation and Report**

   Tool: Overleaf and Google Slides

   Technique: Overleaf is employed for creating and collaborating on report writing, leveraging its LaTeX capabilities for document structuring and version control. Simultaneously, Google Slides is used for creating and collaborating on presentations, providing a cloud-based platform for real-time editing and seamless teamwork.

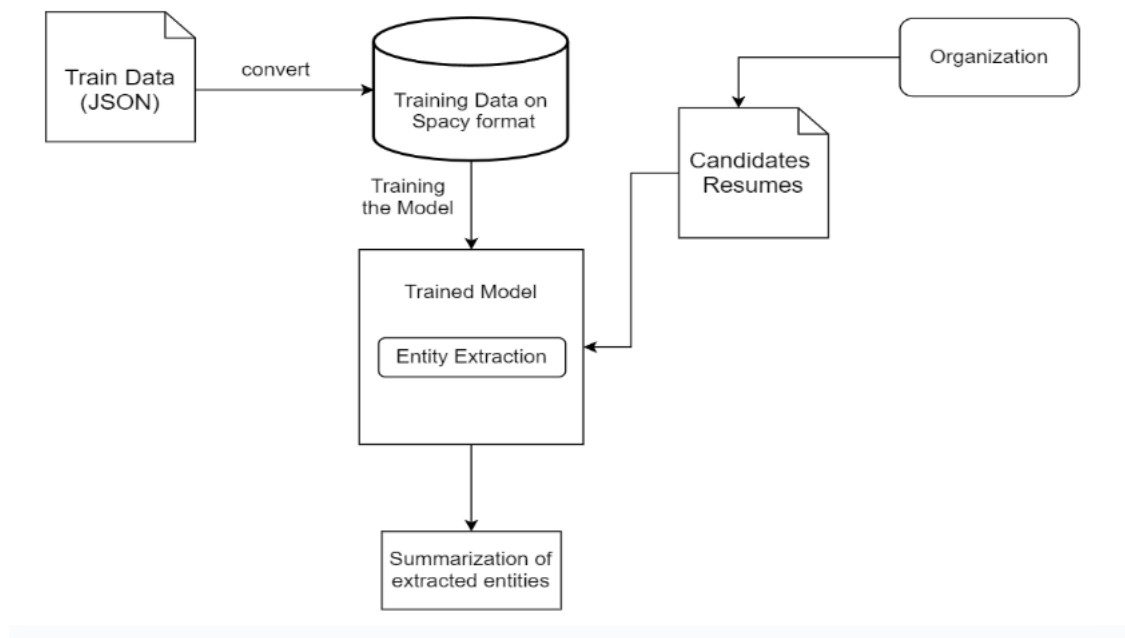# 6  SYSTEM DESIGN

## 6.1  SYSTEM OVERVIEW



Figure 6.1: System Diagram

## 6.2   USECASE DIAGRAM



Figure 6.2: Usecase Diagram
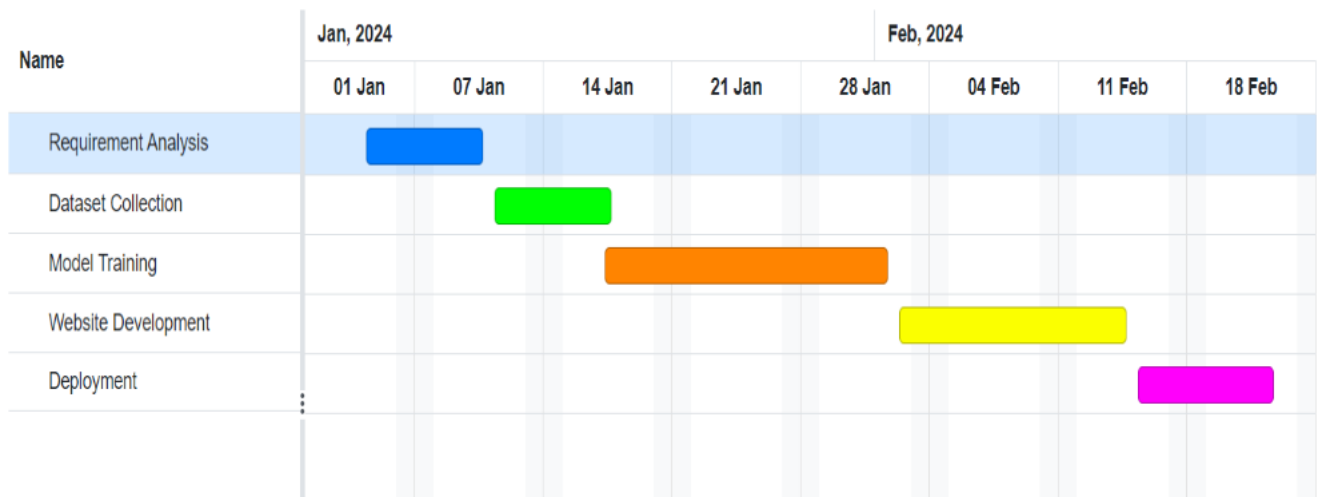
# 7    GANTT CHART



Figure 7.1: Gantt Chart

# 8  EXPECTED OUTCOMES

Nulla in ipsum. Praesent eros nulla, congue vitae, euismod ut, commodo a, wisi. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Aenean nonummy magna non leo. Sed felis erat, ullamcorper in, dictum non, ultricies ut, lectus. Proin vel arcu a odio lobortis euismod. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Proin ut est. Aliquam odio. Pellentesque massa turpis, cursus eu, euismod nec, tempor congue, nulla. Duis viverra gravida mauris. Cras tincidunt. Curabitur eros ligula, varius ut, pulvinar in, cursus faucibus, augue.

# Bibliography

[1] G. Narendra and S. Hashwanth, "Named entity recognition based resume parser and sum-marizer."

[2] D. S. Chavare and A. B. Patil, "Resume parsing using natural language processing." *Grenze International Journal of Engineering & Technology (GIJET)*, vol. 9, no. 1, 2023.

[3] V. Mittal, P. Mehta, D. Relan, and G. Gabrani, "Methodology for resume parsing and job domain prediction," *Journal of Statistics and Management Systems*, vol. 23, no. 7, pp. 1265–1274, 2020.