

**TRIBHUWAN UNIVERSITY
INSTITUTE OF ENGINEERING
PURWANCHAL CAMPUS**



**A THIRD YEAR MINOR PROJECT PROPOSAL
ON
"RESUME PARSER USING NAMED ENTITY RECOGNITION"**

SUBMITTED TO:

DEPARTMENT OF COMPUTER AND ELECTRONICS

SUBMITTED BY:

DINESH NEPAL (PUR077BCT033)

BISHWAS RAJ KOIRALA (PUR077BCT029)

BHUWAN OJHA(PUR077BCT020)

KUSHAL POUDEL (PUR077BCT043)

3 March,2024

DECLARATION

We hereby declare that the report of the project entitled “RESUME PARSER USING NAMED ENTITY RECOGNITION” which is being submitted to the Department of Electronics and Computer Engineering, IOE, Purwanchal Campus, Dharan in the partial fulfillment of the requirements for the award of the Degree of Bachelor of Engineering in Computer Engineering, is a bona fide report of the work carried out by us. The materials contained in this report have not been submitted to any University or Institution for the award of any degree.

Dinesh Nepal [PUR077BCT033]

Biswash Raj Koirala[PUR077BCT029]

Bhuwan Ojha [PUR077BCT020]

Kushal Poudel [PUR077BCT043]

CERTIFICATE OF APPROVAL

The undersigned certify that they have read and recommended to the Department of Electronics and Computer Engineering, IOE, Purwanchal Campus, a minor project work entitled “RESUME PARSER USING NAMED ENTITY RECOGNITION” submitted by Dinesh Nepal, Biswash Raj Koirala, Bhuwan Ojha and Kushal Poudel in partial fulfillment for the award of Bachelor’s Degree in Computer Engineering. The Project was carried out under special supervision and within the time frame prescribed by the syllabus. We found the students to be hardworking, skilled and ready to undertake any related work to their field of study and hence we recommend the award of partial fulfillment of Bachelor’s degree of Computer Engineering.

.....

Head of Department

Mr. Pravin Songraula

Department of Electronics and Computer Engineering

Purwanchal Campus,Dharan

ACKNOWLEDGEMENTS

We express our sincere appreciation to the Head of the Department (HOD), Mr. Pravin Sangroula and the Deputy Head of the Department, Mr. Pukar Karki for their invaluable guidance, insightful suggestions, constructive feedback, and continuous support throughout the preparation of this proposal. We would like to express our sincere gratitude to Mr. Kshitiz Regmi, a highly skilled Machine Learning Engineer, whose invaluable contributions significantly enhanced the success of this project.

We are deeply thankful to the esteemed teachers and faculty members whose valuable feedback, encouragement, and scholarly insights have enriched this proposal. Their dedication to academic excellence and commitment to fostering knowledge have been a source of inspiration. Their collective guidance have immensely contributed to the refinement and improvement of this proposal.

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

1	INTRODUCTION	1
1.1	INTRODUCTION	1
1.2	PROBLEM STATEMENT.....	2
1.3	OBJECTIVES.....	2
1.4	PROJECT SCOPE	2
2	LITERATURE REVIEW	4
2.1	RELATED THEORIES.....	4
2.2	RELATED WORK.....	5
3	REQUIREMENT ANALYSIS	6
3.1	FUNCTIONAL REQUIREMENTS	6
3.2	NON-FUNCTIONAL REQUIREMENTS	7
4	METHODOLOGY	8
5	TOOLS AND TECHNOLOGY USED:	10
6	SYSTEM DESIGN	12
6.1	SYSTEM OVERVIEW.....	12
6.2	USECASE DIAGRAM.....	13
6.3	SEQUENCE DIAGRAM.....	14
6.4	GANTT CHART	15
7	RESULTS	16
8	REFERENCES	17

List of Figures

6.1	System Diagram	12
6.2	Usecase Diagram	13
6.3	Sequence Diagram	14
6.4	Gantt Chart	15

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 Language" 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 projects 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 METHODOLOGY

The proposed system with Named Entity Recognition for Resume Parsing includes following blocks:

1. **Dataset:**

This refers to the original dataset that contains the resume of people .The datasets were secondary datasets which were available on different sites on internet such as Kaggle, Github and some datasets were manually created by us.

2. **Data Preprocessing:**

This step involves preparing the dataset for training and testing.In our case, we annotated the resumes .”Annotation” refers to the process of adding metadata or additional information to a resume document.This metadata typically includes details about the various sections and elements within the resume. Data preprocessing ensures that the data is in a suitable format for the machine learning model.

3. **Preprocessed Data:**

This is the dataset after applying the data preprocessing techniques. It contains the transformed dataset which is in JSON format ready for further analysis and model training.

4. **Train-Test Split:**

The preprocessed data is split into two sets: the training set and the testing set. The training set is used to train the machine learning model, while the testing set is used to evaluate the model’s performance on unseen data. The split is typically done randomly, allocating a certain percentage of the data to each set (e.g., 80% for training, 20% for testing).

5. Training Set:

This subset of the preprocessed data is used to train the NER model. It includes the annotated data set used to teach the model the underlying patterns and relationships.

6. Loading into Spacy:

The training datasets are loaded into spacy to obtain the docbin file so that these docbin file are used to train the NER model

7. Training NER model:

This step involves training the model to apply Named Entity Recognition such that it identifies the metadata or the entities from the text

8. Trained NER model:

After the NER model is trained, it becomes a trained model that can be used for entity recognition. It encapsulates the learned probabilities and distributions, enabling it to make predictions on unseen data with the accuracy score of 86 % .

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):

Tool: Python-based library Spacy

Technique: 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:

Tool: HTML, CSS, JS and Bootstrap

Technique: 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:

Tool: Django (Python web framework)

Technique: 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. **Version Control:**

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.

5. **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

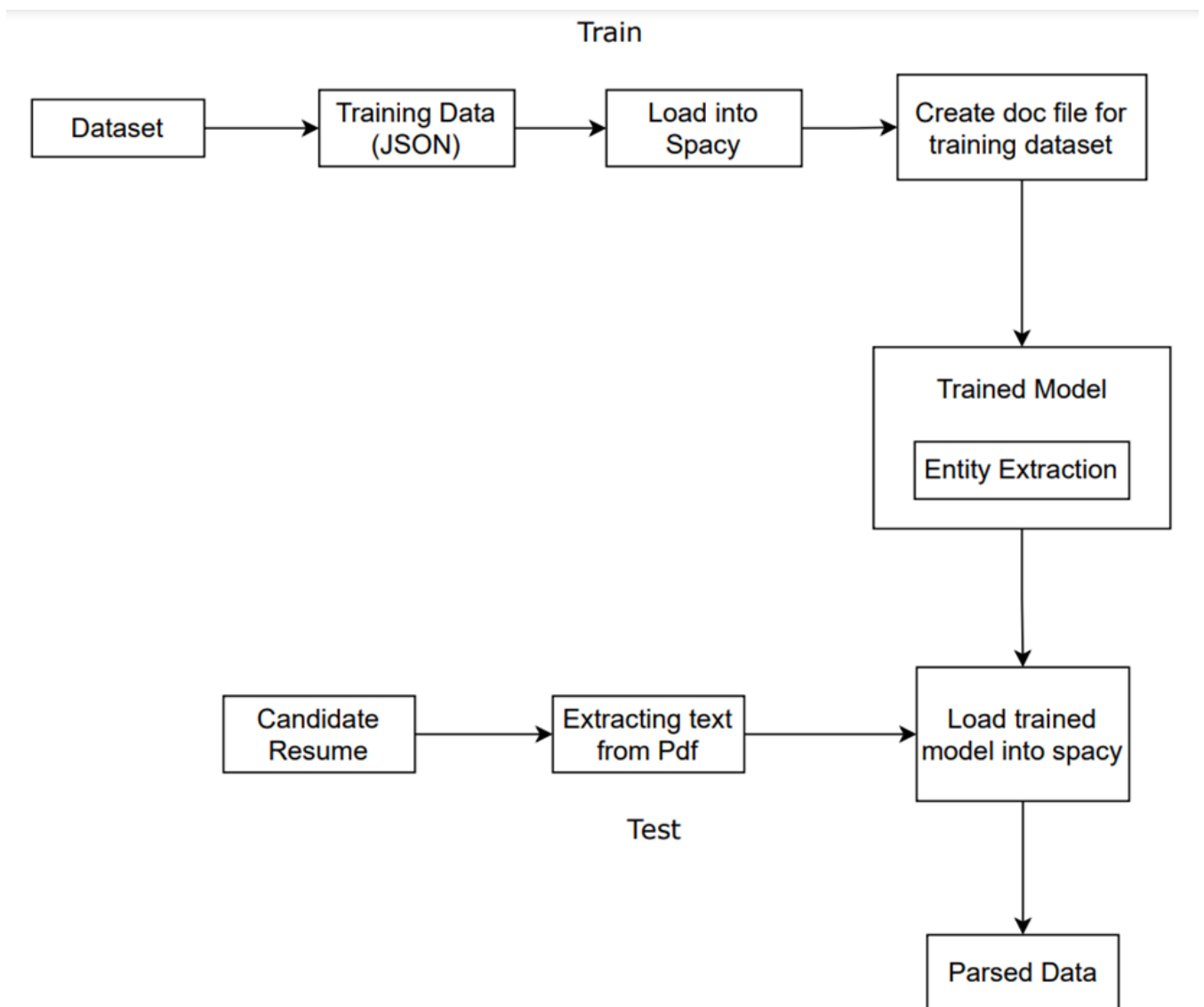


Figure 6.1: System Diagram

6.2 USECASE DIAGRAM

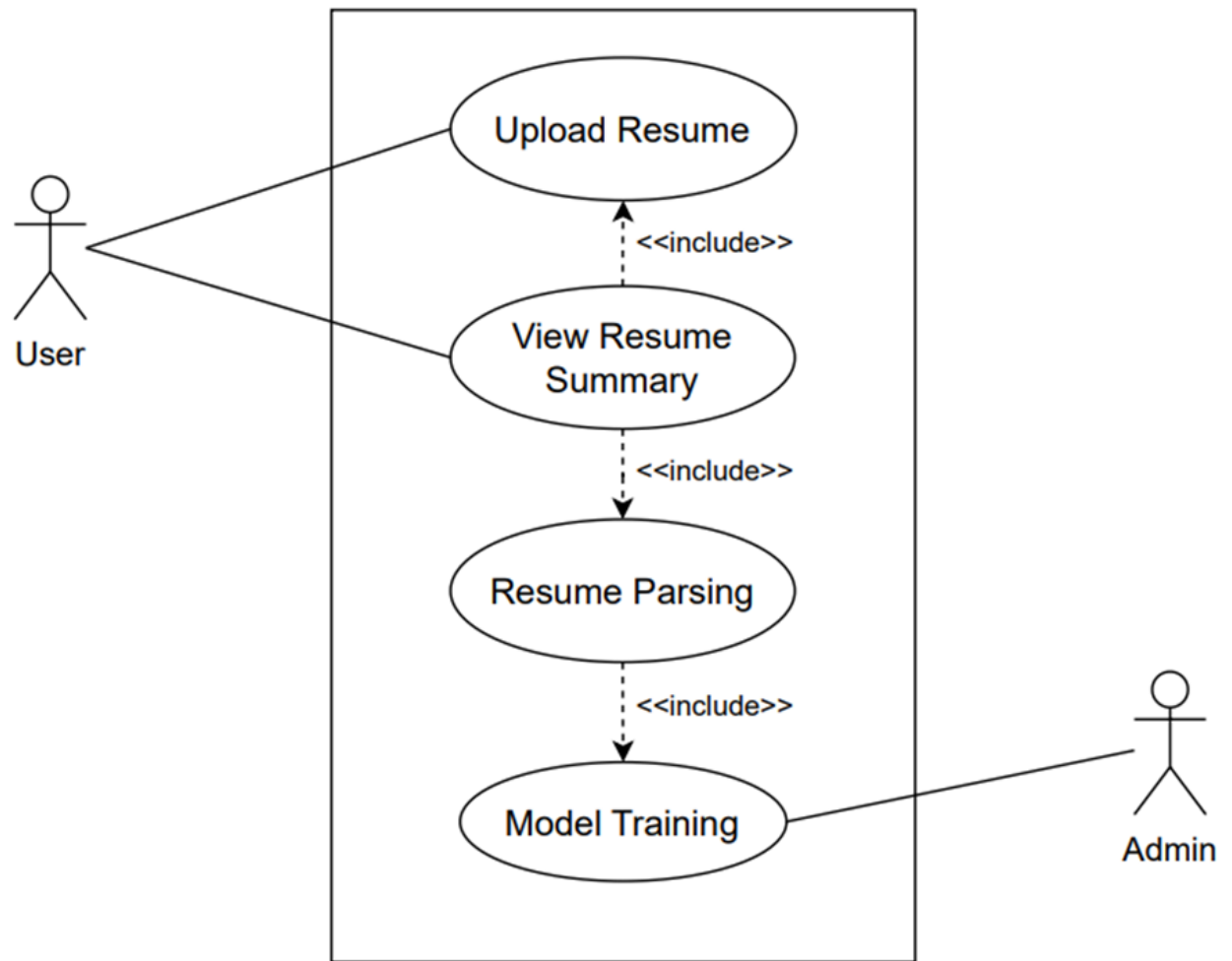


Figure 6.2: Usecase Diagram

6.3 SEQUENCE DIAGRAM

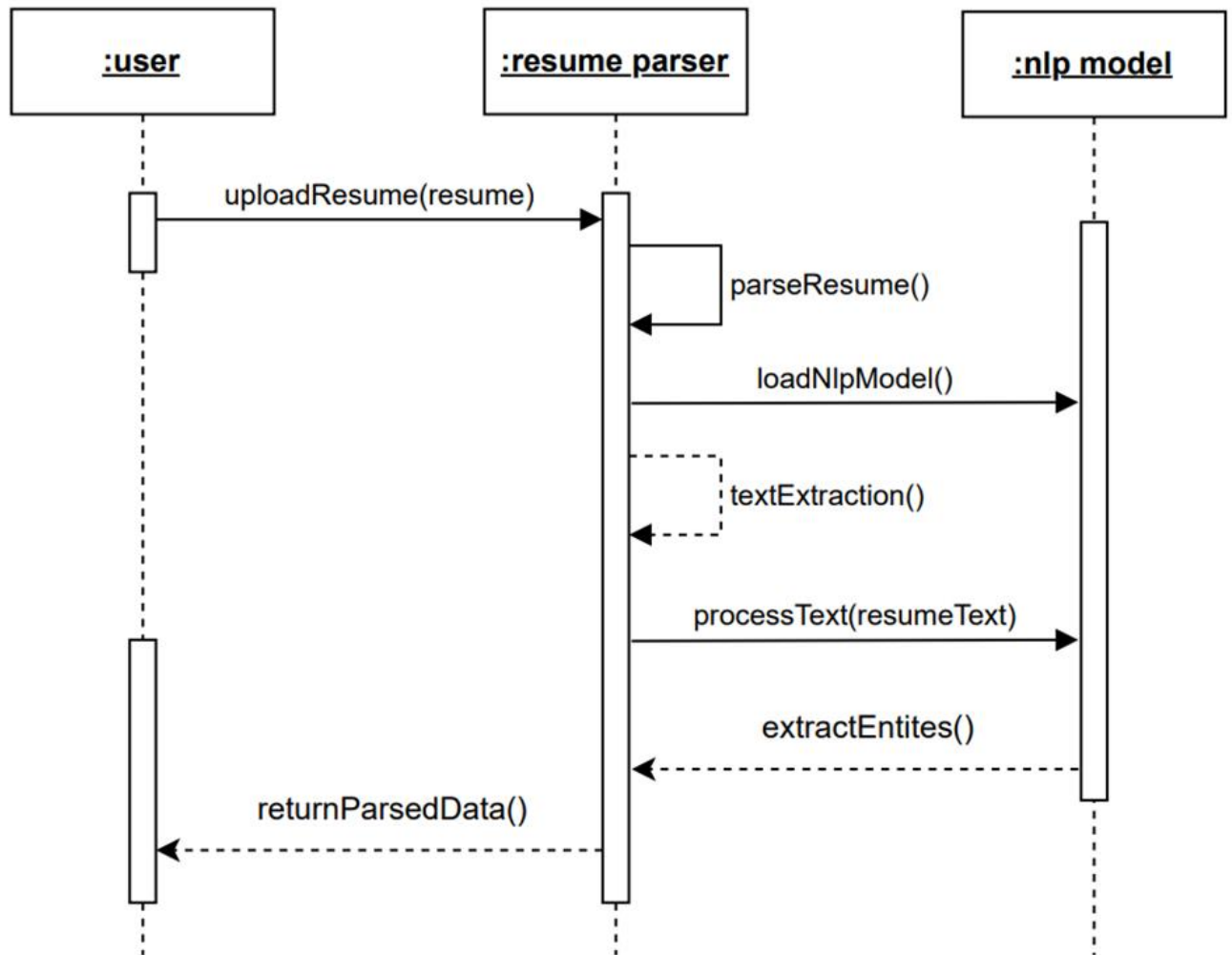


Figure 6.3: Sequence Diagram

6.4 GANTT CHART

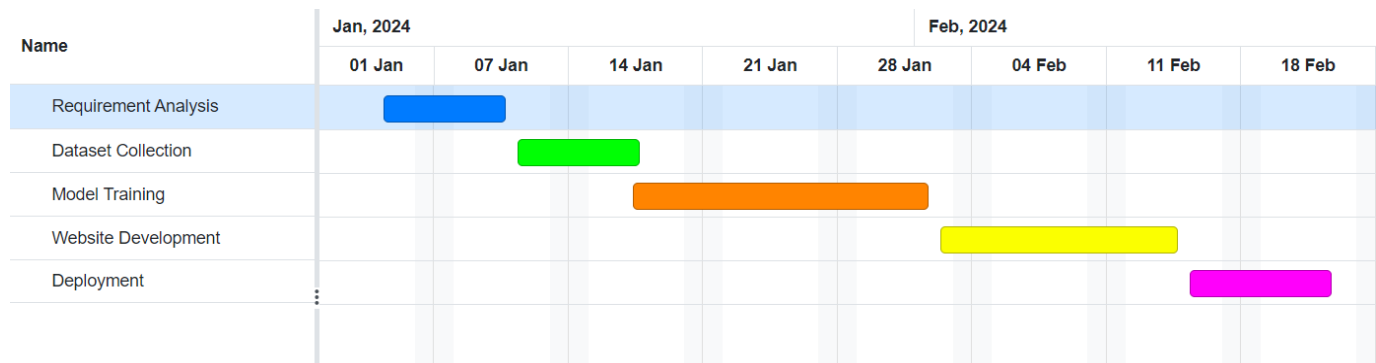
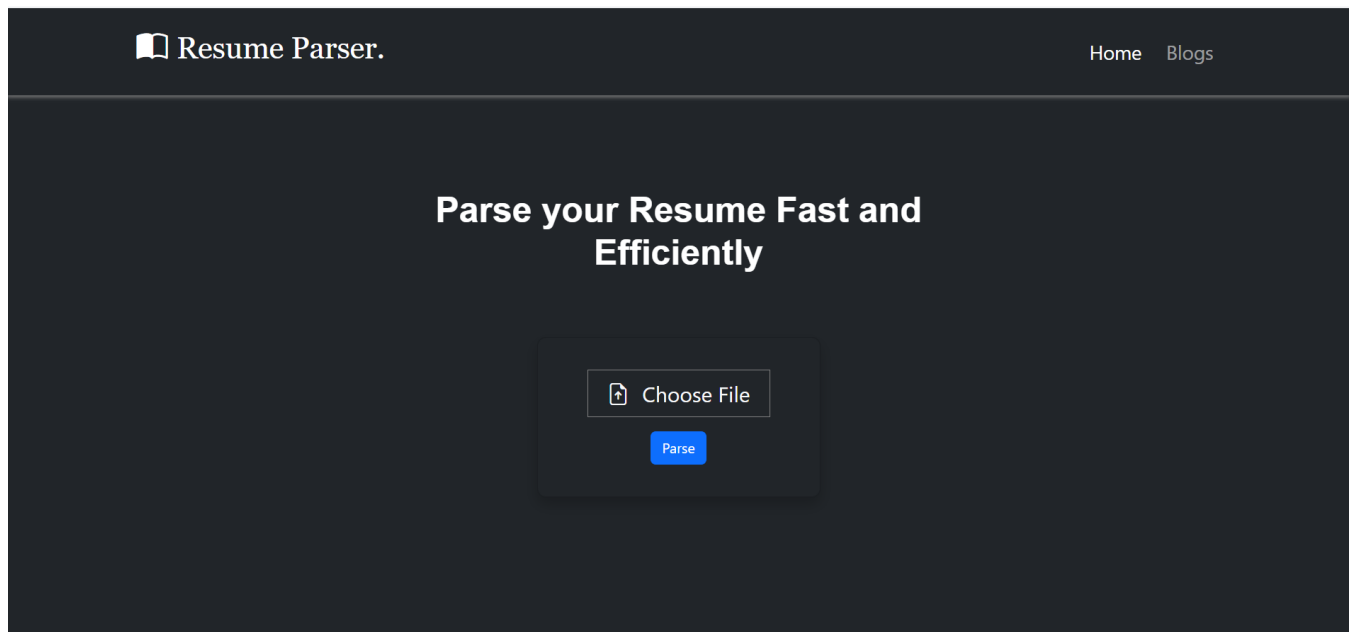


Figure 6.4: Gantt Chart

7 RESULTS



ALL		SKILLS	LANGUAGE	CERTIFICATION	WORKED AS	YEARS OF EXPERIENCE
Text						Label
Akhil Shah						NAME
www.linkedin.com/in/akhil-shah04(LinkedIn						LINKEDIN LINK
Java						SKILLS
SQL						SKILLS
C++						SKILLS
Hindi						LANGUAGE
Gujarati						LANGUAGE
Marathi						LANGUAGE
English						LANGUAGE
Kutchi						LANGUAGE
Hadoop Foundations - Level 1						CERTIFICATION
Big Data Foundations - Level 1						CERTIFICATION
Learn ORACLE 12C Complete ☐ Honors & Awards: ☐ Secured						CERTIFICATION
Database Administrator - TIAA July 2021 - Present						WORKED AS
1 year 1 month						YEARS OF EXPERIENCE

Bibliography

- [1] G. Narendra and S. Hashwanth, “Named entity recognition based resume parser and summarizer.”
- [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.

