Industry Project / IBM Project Report

On

Resume Data Extraction and Company FAQs with AI Integration

Developed By: -
DharmagnaVyas
(20162122006)
Sakshi Shah
(20162121027)
Yashvi Akola
(20162121001)

Guided By:-
Prof. Neha Sisodiya (Internal)
Mr. Anoj Dixit (External)

Submitted to
Department of Computer Science & Engineering
Institute of Computer Technology

Ganpat University | **Institute of Computer Technology**

|| विद्या समाजोत्कर्षं: ||

Year: 2024

CERTIFICATE

This is to certify that the **Industry** Project work entitled **"Resume Data Extraction and Company FAQs with AI Integration"** by Dharmagna Vyas (EnrolmentNo.21162122006), Sakshi Shah(Enrolment No.20162121027), and Yashvi Akola(Enrolment No.20162121001) of Ganpat University, towards the partial fulfillment of requirements of the degree of Bachelor of Technology – Computer Science and Engineering, carried out by them in the CSE(BDA) Department at Elearners 365. The results/findings contained in this Project have not been submitted in part or full to any other University / Institute for award of any other Degree/Diploma.

Name & Signature of Internal Guide

Name & Signature of Head

Place: ICT - GUNI

Date:

ACKNOWLEDGEMENT

Dharmagna Vyas (20162122006)

Sakshi Shah (20162121027)

Yashvi Akola (20162121001)

ABSTRACT

In today's competitive recruitment environment, the efficient extraction and validation of essential details from resumes are crucial for enhancing hiring processes. Our proposed system offers a comprehensive solution designed to accurately extract structured information, such as names, emails, contact numbers, skills, and locations, from resumes available in PDF/TXT formats. Employing a blend of rule-based extraction methods for structured data and advanced Language Model (LLM) techniques for unstructured data, our system ensures a thorough analysis of resume content. Noteworthy features include robust text and data validation mechanisms, which meticulously verify the accuracy and integrity of the extracted information, thus minimizing errors and boosting reliability. User experience is prioritized through support for resume uploads in diverse formats and the implementation of sophisticated text extraction tools, facilitating a seamless processing workflow. By effectively extracting and validating resume information, our system aims to streamline recruitment procedures and enhance overall data processing efficiency. This project signifies a significant advancement in leveraging technology to optimize talent acquisition strategies, ultimately leading to more informed hiring decisions and improved organizational outcomes. Additionally, the integration of two new projects, Resume Management and Company FAQs with AI Integration, further enriches the user experience and support capabilities, reinforcing our commitment to revolutionizing recruitment practices through innovative technological solutions. The system also supports chatting with multiple PDFs simultaneously, providing a robust, flexible tool for managing various resume inputs efficiently.

INDEX

CHAPTER: 1 INTRODUCTION

CHAPTER 1 INTRODUCTION

In the competitive landscape of talent acquisition, the ability to swiftly and accurately parse resumes is a critical capability for any organization aiming to attract and retain top talent. The advent of digital technologies has paved the way for innovative approaches to resume data extraction, among which rule-based systems, particularly those implemented using Python, have shown significant promise. This report presents an in-depth analysis of the development, implementation, and outcomes of a rule-based resume data extraction project, utilizing Python's powerful programming capabilities. Additionally, it explores the integration of two new projects focused on enhancing user engagement and support: Resume Management and Company FAQs with AI Integration.

In the rapidly evolving landscape of recruitment and talent acquisition, the efficient and accurate extraction of information from resumes is paramount. As organizations strive to streamline their hiring processes, the need for advanced technologies that can parse and analyze resumes with precision has become increasingly evident. This report delves into the methodology, tools, and outcomes of a comprehensive study on resume data extraction, aiming to provide valuable insights into the current state of the art and its integration with cutting-edge projects aimed at revolutionizing user experience and support in the digital recruitment realm.

The primary objective of this report is to assess the effectiveness and reliability of various resume data extraction methods and tools. By examining the strengths, limitations, and emerging trends in this field, we aim to offer a comprehensive overview that will aid both recruitment professionals and technology enthusiasts alike in navigating the complexities of modern talent acquisition processes. The system also supports chatting with multiple resumes in PDF format simultaneously, enhancing processing efficiency. Moreover, users have the capability to save the history of their chats, providing a detailed record of interactions, with the added feature to clear this history as needed for privacy or organizational purposes.

CHAPTER: 2 PROJECT SCOPE

This report covers the systematic approach to extracting key information from resumes using a rule-based system coded in Python, as well as the integration of an open-source LLM framework. It includes an examination of the types of data targeted for extraction, such as personal information, educational background, professional experience, and skill sets. The scope extends to the consideration of resume format (e.g., PDF) and the adaptability of the rule-based system to accurately parse data from these diverse sources. Additionally, the report addresses the challenges encountered during the extraction process, including handling unstructured data, maintaining data privacy, and seamless integration with the chatbot interface for interactive querying.

Furthermore, the system has been enhanced with a capability for users to interact with multiple resumes in PDF format simultaneously. This feature allows for efficient management and processing of large volumes of resume data, enhancing the tool's utility in high-volume recruitment scenarios. This capability supports real-time data extraction and provides an interactive user experience, where recruiters can swiftly navigate through multiple candidate profiles, ensuring that no critical information is overlooked during the hiring process.

CHAPTER: 3 SOFTWARE AND HARDWARE REQUIREMENTS

CHAPTER 3 SOFTWARE AND HARDWARE REQUIREMENTS

Minimum Hardware Requirements

Processor	2.0 GHz
RAM	12GB
HDD	100GB

Table 3.1 Minimum Hardware Requirements

Minimum Software Requirements

Operating System	Any operating system which can support an internet browser.
Programming language	Python
Other tools & tech	Google Colab, Azure

Table 3.2 Minimum Software Requirement

CHAPTER: 4 PROJECT PLAN

4.1 List of Major Activities

Task 1: Define Project Objectives and Requirements
Task 2: Gather Raw Resume Data from Company
Task 3: Preprocess Raw Resume Data (PDF/TXT)
Task 4: Explore and Understand the Structure of Resume Data
Task 5: Develop Rule-Based Extraction Methods
Task 6: Implement LLM Techniques
Task 7: Implement LLM framework for PDF resume handling.
Task 8: Develop custom chatbot interface for querying resume content.
Task 9: Design user-friendly interface for resume preview and chatbot interaction.
Task 10: Integrate backend systems for data flow management.
Task 11: Conduct testing and optimization for seamless functionality.
Task 12: Select suitable browser-based AI chatbot builder.
Task 13: Develop comprehensive FAQ database covering various company topics.
Task 14: Test and optimize chatbot functionality across platforms.
Task 15: Integrate chatbot into company website for enhanced user engagement.
Task 16: Continuously refine solution based on user feedback and evolving needs.

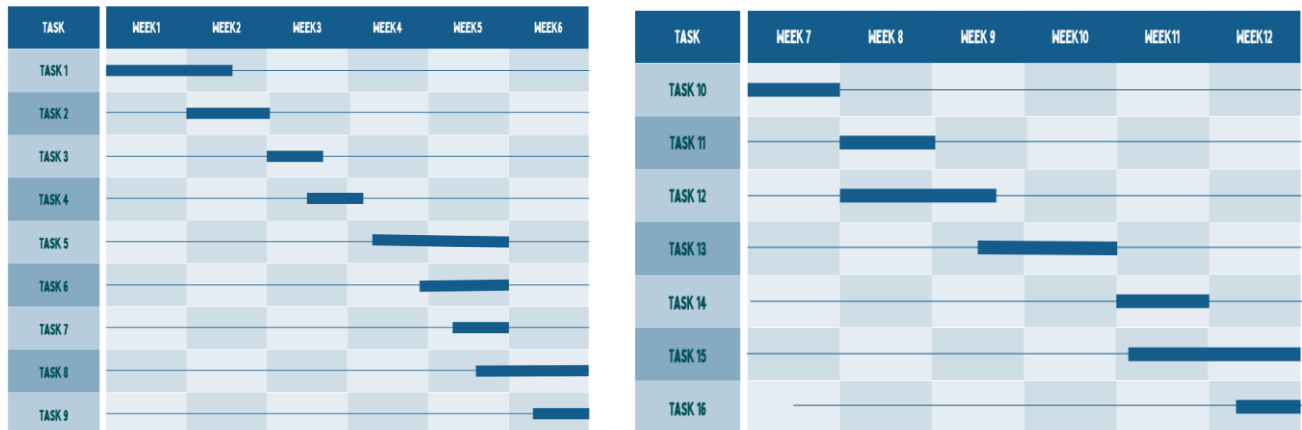## 4.1 Estimated Time Duration week wise

Figure 4.1 Task Completion Estimated Time Duration in Days

4.1.1 Data Collection

Our project relies on raw data or testing resume data provided by our company. This dataset serves as the foundation for our system's development and testing phases, enabling us to refine and optimize our extraction and validation techniques. By utilizing this dataset, we aim to create a robust solution that accurately extracts key details from resumes, ultimately improving recruitment processes and data processing efficiency.

CHAPTER: 5 IMPLEMENTATION DETAILS

CHAPTER 5 IMPLEMENTATION DETAILS

5.1. LLM Framework Integration

Technology Selection: We will meticulously evaluate open-source LLM frameworks like SpaCy, NLTK, based on key criteria. These criteria include accuracy in named entity recognition (NER) for identifying crucial resume sections (skills, experience), text extraction efficiencyfrom PDFs, and overall framework support. The chosen framework will be integrated with our platform's workflow

PDF Processing Pipeline: To ensure efficient handling of PDF uploads, a dedicated processing pipeline will be developed. This pipeline might utilize libraries like PyPDF2 or PDFMiner. These libraries can effectively extract text content while preserving formatting to the greatest extent possible. This extracted text will be fed into the chosen LLM framework for further analysis.

LLM Model Integration: The selected LLM framework will be seamlessly integrated with the text processing pipeline. Once text is extracted from the uploaded PDF, the LLM can perform advanced tasks like named entity recognition (NER). This allows the LLM to identify key resume sections and extract relevant information such as skills, experience, education, and contact details.

5.2.Chatbot Development:

Natural Language Processing (NLP): We will leverage Natural Language Processing (NLP) techniques to empower the chatbot to understand user queries related to their resumes. This will involve tasks like:
Intent Recognition: The chatbot will be trained to identify the user's intent behind their question. For example, a user might ask "What are my most relevant skills based on this resume?" or "Can you highlight my work experience in software development?". Intent recognition ensures the chatbot understands the user's objective and provides tailored responses.
Sentiment Analysis: The chatbot can analyze the sentiment of user queries to detect frustration or confusion. Based on sentiment analysis, the chatbot can adjust its responses to be more empathetic and helpful.
Chatbot Framework Selection: Considering factors like ease of development, customization options, and seamless integration with the chosen LLM framework, a suitable chatbot framework will be selected. Popular options include Rasa, Dialogflow, and Botpress. This framework will provide the foundation for building our custom chatbot interface.
Knowledge Base Construction: A comprehensive knowledge base will be constructed to store information extracted from the LLM analysis of user resumes. This knowledge base acts as a central repository for the chatbot to access relevant details and answer user queries accurately. It will be populated with structured information extracted from the LLM, such as skills, experience, and other resume sections.
Chat Interface Design: A user-friendly chat interface will be designed to facilitate seamless interaction between users and the chatbot. This interface might include features like:
Query History: A history section that allows users to review past interactions with the chatbot, potentially aiding in recalling previous responses or resuming conversations.
Context Awareness: The chatbot will be designed to maintain context across conversations. This allows the chatbot to understand follow-up questions based on previous interactions, providing a more natural and engaging experience.

Clear Response Presentation: Responses from the chatbot will be presented in a clear and concise format, potentially highlighting key information extracted from the resume or offering additional resources if needed.

1. User Interface Design:

Resume Preview: We will design a visually appealing and informative resume preview section. This section should present the uploaded resume with clear formatting and potentially highlight key extracted information based on LLM analysis. This could involve color-coding relevant skills, displaying experience durations visually, or summarizing key achievements.

Chatbot Integration: The chatbot interface will be seamlessly integrated within the resume preview section. This might involve a dedicated chat window that appears alongside the resume or a toggle option that allows users to switch between preview and chatbot interaction for ease of use.

Intuitive Navigation: A user interface will be designed that prioritizes intuitive navigation between the resume preview and chatbot interaction. This can involve clear buttons labeled for functionality (e.g., "Preview Resume," "Chat with Bot"), a breadcrumb navigation system for context awareness, and a well-organized layout that avoids information overload.

Backend Integration

The backbone of the platform's seamless operation lies in the integration of backend systems. This integration ensures a smooth flow of data between the resume uploads, chatbot interactions, and data storage solutions. The backend architecture was designed for scalability and security, capable of handling a large number of concurrent users and storing sensitive personal data securely. This involved setting up cloud-based storage solutions, developing APIs for the efficient transfer of data between the front end and back end, and implementing security measures such as encryption and secure access protocols. The backend also supports the chatbot's AI functionalities, allowing for real-time processing and analysis of resume data.

CHAPTER: 6 CONCLUSION AND FUTURE WORK

CHAPTER 6 CONCLUSION AND FUTURE WORK

Conclusion

In conclusion, Our project represents a significant step forward in revolutionizing resume management and fostering user engagement through the strategic application of advanced AI technologies. By seamlessly integrating a robust LLM framework with an intelligent chatbot interface, we've addressed the evolving demands of the digital recruitment landscape. This empowers users with:

Automated Resume Analysis: The LLM framework automates resume analysis, extracting key information and saving users valuable time compared to manual review.
Interactive Resume Exploration: The chatbot interface facilitates an interactive experience, allowing users to explore their resumes through natural language queries. This personalized approach unlocks deeper insights into their skillsets and qualifications.
Streamlined FAQ Support: The inclusion of a browser-based AI chatbot builder empowers companies to streamline FAQ support. Users can quickly access answers to commonly asked questions, improving overall user experience and productivity.

Future work

LLM Model Refinement: We will continuously evaluate and refine the chosen LLM framework based on advancements in the field. This ensures our platform leverages the latest capabilities for accurate resume analysis and information extraction.
Chatbot Training Expansion: We plan to expand the training data used for the chatbot, allowing it to understand a wider range of user queries and respond with even greater accuracy and nuance. This personalization fosters a more natural and engaging user experience.
Integration with Employer Platforms: Future iterations may explore integrating our platform with employer platforms, allowing for smoother resume submission and analysis within existing workflows, creating a more efficient recruitment process.
User Feedback and Feature Development: We will actively solicit user feedback to identify areas for improvement and prioritize the development of new features that further enhance the platform's value proposition for both candidates and employers.

CHAPTER: 7 REFERENCE

1. https://www.chatpdf.com/c/0fQVbn1tL66XqXERSNza7
2. https://www.lg.com/in/