

Major Project I Mid Viva Report



TOPIC- Resume Analyzer/ Parser

Group Members-

Rachit - 21103089

Hrishita Singh - 21104007

Ujjawal Goel - 21103203

SUPERVISOR- Mr. Shivendra Singh

Abstract:

The Resume Analyzer project is designed to streamline and enhance the recruitment process by leveraging advanced Natural Language Processing (NLP) techniques and machine learning algorithms. This tool automates the extraction, evaluation, and comparison of key information from resumes, including skills, qualifications, experience, and educational background. By analyzing resumes against predefined job descriptions and requirements, the system generates a comprehensive score and detailed insights, facilitating more informed and efficient hiring decisions. This report includes a literature survey of existing systems and technologies that have been implemented. Further it includes key features of the application and methodologies to implement them.

Introduction:

The Resume Analyzer project is a comprehensive tool designed to enhance the resume-building process for users and provide insightful analytics for administrators. With distinct user and admin logins, users can upload their resumes to receive structured feedback on personal information, skills, and keywords. The platform evaluates the resume with an ATS (Applicant Tracking System) score and provides personalized recommendations on skills, courses and certifications to boost employability. Additionally, users receive job recommendations using web scraping of current openings and key tips to improve the overall structure of their resumes. To further assist users, the platform offers a bonus video to help them prepare for interviews, tailored to their recommended skills.

On the admin side, the project offers college Training and Placement (TNP) departments valuable insights into student data. Admins can access details like the total number of students and their most common job preferences, which helps in aligning campus recruitment efforts with student interests. Admins can also download all student data in CSV format for further analysis, making the platform a powerful tool for optimizing the placement process and preparing students for the job market.

Literature survey:

Literature Survey 1: Literature Survey for Resume Parsing Using NLP

The field of resume parsing has grown in importance with the rise of automation in recruitment, aiming to streamline the manual process of reviewing resumes and extracting relevant information for candidate shortlisting. This literature survey summarizes key research contributions to the development of automated resume parsers, focusing on the use of Natural Language Processing (NLP) and Machine Learning (ML) techniques to address the challenge of unstructured resume data.

1. Information Extraction Systems

Resume parsing falls under the broader category of Information Extraction (IE) systems, which are designed to extract structured data from unstructured sources such as resumes. **Ciravegna (2001)** emphasizes the need for adaptive information extraction systems that can generalize from rule-based approaches, highlighting early efforts to create models that learn from data rather than relying solely on hardcoded rules. These systems improve the efficiency of processing resumes in various formats and structures, a critical challenge in resume parsing.

2. Entity Recognition and Classification

An essential task in resume parsing is the accurate identification of key entities such as names, skills, degrees, and work experience. **Chandel et al. (2006)** propose efficient top-k search algorithms for dictionary-based entity recognition, an approach that has been widely adopted in resume parsing. By employing dictionaries of known entities, their system can quickly identify relevant information. However, modern systems have evolved to incorporate machine learning algorithms for more flexible and scalable entity recognition.

3. Machine Learning for Resume Parsing

The application of ML techniques, including neural networks and deep learning models, has improved the accuracy and robustness of resume parsers. **Wang & Zu (2019)** present a resume parsing pipeline that uses neural network-based classifiers and distributed word embeddings to automatically extract relevant information from resumes. This eliminates the need for handcrafted features and allows the model to generalize better across different resume formats and structures. **Kopparapu (2015)** also advocates for using ML in parsing both structured and unstructured resumes, providing an early framework for integrating machine learning with NLP techniques.

4. Natural Language Processing Techniques

NLP techniques form the core of modern resume parsing systems. Tokenization, part-of-speech (POS) tagging, and lemmatization are standard NLP tools used to process resumes. **Chakrabarti (2002)** and **Cafarella et al. (2005)** laid the groundwork for using NLP in extracting meaningful information from text, which directly influences the development of resume parsers. Tokenization helps break down the text into meaningful chunks, POS tagging helps identify the grammatical roles of each token, and lemmatization reduces words to their base forms to standardize the information.

5. Recent Advances: BERT and Other Models

Recent approaches have leveraged state-of-the-art NLP models like BERT for resume parsing. **Bhatia et al. (2019)** explore the use of BERT, a transformer-based language model, to improve resume parsing and job-candidate matching. BERT's ability to understand contextual word embeddings allows it to handle the nuances of resumes written in natural language, providing a more sophisticated method for entity recognition and information extraction.

6. Challenges in Resume Parsing

Several challenges remain in the field of resume parsing. **Vukadin et al. (2021)** point out the difficulties in parsing multilingual resumes and recognizing less standardized information such as personal names, especially in globalized contexts. Regular Expressions (Regex) and custom algorithms like those used in **SpaCy** can handle well-defined patterns like email addresses and phone numbers, but handling less structured information such as skills or job titles requires more advanced techniques.

7. Applications and Impact

The automation of resume parsing has a significant impact on the hiring process, especially for large corporations and recruitment agencies that deal with thousands of resumes daily. By reducing the time and effort needed for manual resume screening, automated parsers can improve recruitment efficiency. Additionally, these systems help reduce biases in hiring by focusing on relevant qualifications and skills, rather than superficial formatting or presentation.

Conclusion

The literature on resume parsing demonstrates a clear shift from manual, rule-based systems to more flexible and powerful machine learning and NLP-based systems. The continued development of advanced models like BERT has opened new possibilities for improving the accuracy and efficiency of resume parsing systems. While challenges remain, particularly in handling diverse formats and globalized resume

structures, current trends suggest that automated resume parsing will continue to play a critical role in modern recruitment processes.

Literature survey 2: Literature Survey for Resume Parser and Analyzer Using NLP

The paper on "Resume Parser and Analyzer Using NLP" explores a system designed to streamline the recruitment process by automating the analysis of resumes using Natural Language Processing (NLP). The system focuses on key aspects such as resume parsing, skill extraction, and ranking to provide recruiters with valuable insights. A detailed literature survey helps situate the project within the broader field of NLP and resume analysis technologies.

1. Natural Language Processing (almost) from Scratch

Ronan Collobert et al.

This paper introduces a neural network architecture capable of handling various NLP tasks, including Part-of-Speech (POS) tagging, Named Entity Recognition (NER), and Chunking (shallow parsing). The architecture demonstrates its effectiveness through benchmarks on tasks such as NER and Semantic Role Labeling (SRL), showcasing its versatility in syntactic and semantic language understanding. These tasks are integral to the resume parser, which needs to recognize entities like names, organisations, and roles in a resume. The use of a neural network architecture from this research offers a solid foundation for accurate information extraction in resume analysis.

2. Resume Information Extraction with a Novel Text Block Segmentation Algorithm

Shicheng Zu et al.

This paper fills a gap in applying neural networks to resume parsing, focusing on six key resume fields: personal information, education, work experience, projects, skills, and publications. The authors demonstrate that a BLSTM-CNN-CRF algorithm is highly effective in text block classification and named entity recognition (NER). This segmentation algorithm plays a crucial role in distinguishing important fields from less relevant information (e.g., hobbies), a feature that can be adopted in the resume analyzer to improve parsing accuracy and efficiency.

3. A Few Shot Approach to Resume Information Extraction via Prompts

Chengguang Gan et al.

This paper explores the use of prompt learning, a technique in NLP where few examples are used to guide language models in classifying text. It employs Masked Language Models (MLM), Pre-training Language Models (PLMs), and Seq2SeqPLMs to create prompts that improve resume parsing tasks. The

study's focus on prompt-based learning can enhance the system's ability to handle the dynamic nature of resume formats with minimal training data, which is vital in developing adaptable parsers that adjust to new resume structures.

4. Text Preprocessing in NLP

Agrawal R.

Text preprocessing techniques are foundational to any NLP system. In this survey, common methods such as tokenization, stop-word removal, lowercasing, and lemmatization are discussed. These techniques are essential in cleaning raw resume data to ensure accurate field extraction. The insights from this study support the use of standard text preprocessing steps before feeding resumes into the NLP-based parser for analysis.

5. End-to-End Resume Parsing and Finding Candidates for a Job Description using BERT

Bhatia V. et al.

This paper discusses an end-to-end system that parses resumes and matches them to job descriptions using BERT (Bidirectional Encoder Representations from Transformers). The use of transformers like BERT is noteworthy for the resume parser as they offer superior capabilities in understanding context, which can be applied to match resumes against predefined job requirements.

6. Automatic Extraction of Usable Information from Unstructured Resumes

Kopparapu S. K.

The study focuses on extracting structured information from unstructured resume formats. Using rule-based and machine learning techniques, this paper highlights the challenges of resume parsing and the necessity of accurately identifying key sections (e.g., work experience and education). The insights can guide the development of a resume parser that is both efficient and reliable, handling the diversity of resume formats encountered in real-world applications.

Conclusion

This literature survey reviews several foundational studies that support the development of an NLP-based resume parser and analyzer. Each referenced work contributes to different aspects of the project, such as entity recognition, information extraction, neural network models, and text preprocessing. Together, these studies offer a solid theoretical background for building an efficient and accurate resume analysis system that benefits both recruiters and job seekers.

Literature survey-3: Resume Parser Using Machine Learning

Overview

The paper "Resume Parser Using Machine Learning" by Yash Kanojia et al. addresses the challenge of efficiently processing and analyzing large volumes of resumes in the recruitment industry. The primary focus is on developing a system that leverages machine learning (ML) and natural language processing (NLP) techniques to automate the extraction of key information from resumes, such as educational background, work experience, skills, and contact details. This automation aims to streamline the recruitment process by converting unstructured resume data into a structured format that can be easily analyzed.

Key Contributions

1. Problem Definition and Motivation:

- **Manual Resume Screening Challenges:** The paper identifies the inefficiencies associated with manual resume screening, such as time consumption and the difficulty of handling diverse resume formats (.pdf, .docx, .jpeg). The proposed system seeks to address these issues by automating the extraction process.
- **Objective:** To develop a system that reduces manual effort by converting resumes into structured data, thus enabling faster and more accurate candidate selection.

2. System Design:

- **Modules:** The proposed system consists of three main modules:
 1. **Administration Management System:** Manages user access and oversees the resume processing workflow.
 2. **File Uploader and Analyzer System:** Handles the upload and initial processing of resumes.
 3. **Information Extraction System:** Uses NLP techniques to extract relevant data from resumes and convert it into a structured format.

3. Methodology:

- **Text Preprocessing:** Techniques include text cleaning, tokenization, stopword removal, stemming, and lemmatization. Libraries such as NLTK and Spacy are employed for these tasks.
- **Information Extraction:** The system utilizes Regular Expressions (Regex), NLTK, and Spacy's phrase matcher to identify and extract specific information (e.g., names, addresses, skills) from resumes.

4. Performance Metrics:

- **Efficiency:** The system is expected to significantly reduce the time required for resume screening by automating information extraction.
- **Accuracy:** The ML models are trained to accurately recognize and extract information from diverse resume formats. The system aims for high precision and recall in entity recognition.
- **Customization and Scalability:** The system can be tailored to specific requirements and is designed to handle large volumes of resumes efficiently.

5. Continuous Improvement:

- **Model Performance:** The paper discusses techniques for improving model accuracy, such as fine-tuning algorithms, optimizing hyperparameters, and leveraging state-of-the-art methods.
- **Data Quality:** Efforts to enhance data quality include cleaning, augmentation, and preprocessing of resumes.
- **Feature Engineering:** New features and representations are explored to capture more relevant information.
- **Feedback Loop:** Incorporates user feedback to iteratively improve the parsing model.

6. Impact:

- **Entity Recognition:** Advanced NER models improve the extraction of entities such as names and skills.
- **Parsing Accuracy:** Techniques like ensemble learning and feedback mechanisms contribute to higher accuracy and robustness.

Related Work:

1. **"Deep Resumes Parsing with Gated Recursive Networks"** by Sun et al. (2017) explores deep learning approaches for resume parsing using gated recursive networks, which improve the extraction of contextual information from resumes.
2. **"A Hybrid Deep Learning Model for Resume Information Extraction"** by Ren et al. (2018) discusses hybrid models combining deep learning with other techniques to enhance resume parsing capabilities.
3. **"An Ensemble Approach for Resume Parsing and Information Extraction"** by Prasad et al. (2019) highlights the benefits of ensemble methods in improving parsing accuracy and handling diverse resume formats.
4. **"BERT-Resumé: Leveraging BERT for Resumé Information Extraction"** by Mishra et al. (2020) demonstrates the effectiveness of BERT in understanding and extracting information from resumes, leveraging its contextual embeddings.
5. **"Building a Resume Parser Using Natural Language Processing"** by Patel (Medium) and **"How to Build a Resume Parsing Tool with Python"** by Prabhu (Towards Data Science) provide practical insights and tutorials on implementing resume parsing tools using NLP techniques.
6. **Libraries and Frameworks:**
 - **SpaCy:** Provides tools for text preprocessing and entity recognition.
 - **scikit-learn:** Offers machine learning algorithms and tools for classification tasks.
 - **TensorFlow and PyTorch:** Deep learning frameworks used for developing custom models for resume parsing.

Key Features:

1. **User and Admin Login:** Both users and administrators can securely log in to the system, each with specific access privileges.
2. **Resume Upload and Structured Insights:** Users can upload their resumes, which are analyzed to extract key details such as personal information, skills, and keywords in a well-organized format.
3. **Skill, Course, and Certification Recommendations:** The system provides personalized recommendations for skills, courses, and certifications that users can add to improve their resumes.

4. **ATS Score Evaluation:** Users receive an Applicant Tracking System (ATS) score that evaluates how well their resume aligns with modern recruitment standards.
5. **Resume Improvement Tips:** Users are provided with actionable tips to enhance their resume structure, formatting, and overall presentation.
6. **Job Recommendations via Web Scraping:** The system suggests relevant job openings by scraping online job portals, helping users find positions that match their profile.
7. **Interview Preparation Bonus:** At the end of the process, users are given access to a bonus video designed to help them prepare for interviews based on the recommendations and highlighted skills.
8. **Admin Data Analysis:** Admins can view and analyze data for all students, including the number of students, preferred job roles, and common job roles found in resumes.
9. **CSV Data Download:** Admins can download comprehensive student data in CSV format, useful for further analysis or reporting.

These admin functionalities are particularly beneficial for college Training and Placement (TNP) departments, as they can leverage the data to align recruitment efforts with students' skills and interests, helping to bring relevant companies to campus.

Methodology and Technologies:

1. **User and Admin Login:** Implement secure password based authentication using flask-login framework.
2. **Resume Upload and Structured Insights:** Use python file upload libraries to accept resumes in PDF format. NLP libraries such as NLTK can then be used to parse the resume content in a structured way.
3. **Skill, Course and Certification Recommendations:** Analyse the resume using NLP techniques to get the skills and identify the domain to give recommendations related to it.
4. **ATS Score Evaluation:** Develop a scoring algorithm that analyses the resume against industry standards and common ATS criteria.
5. **Resume Improvement Tips:** Implement a rule-based or machine learning system that analyses the structure of the resume (e.g., availability of resume sections) and offers suggestions for improvement. NLP can be used to detect common issues such as vague language or inconsistent formatting.
6. **Job Recommendations via Web Scraping:** Use web scraping tools like Selenium and WebDriver to gather job listings from popular LinkedIn. The extracted data can be matched with user profiles based on their parsed skills and experience using a similarity algorithm.

7. Interview Preparation Bonus: Embed relevant video content from YouTube to present interview preparation materials. Videos can be recommended based on the user's skill set and job interests.
8. Admin Data Analysis: Implement a dashboard using python libraries like pandas and matplotlib to analyse users data.
9. CSV Data Download: Provide an export feature where admins can download student data in CSV format using python libraries.
10. Streamlit: It is a python library to build and deploy python applications.
11. Database: SQL queries are used to implement the functionalities and store database.

Conclusion:

In conclusion, the resume analyzer project offers a comprehensive, automated solution to streamline the recruitment process for both candidates and administrators. By leveraging Natural Language Processing (NLP) and Machine Learning (ML) techniques, the platform efficiently extracts key information from resumes, provides valuable insights, and offers personalised recommendations to enhance the user's profile. The system also improves the chances of passing through Applicant Tracking Systems (ATS) by providing a score and targeted feedback. Job recommendations and interview preparation videos further equip candidates for success. On the administrative side, the platform enables college TNP departments to analyze student data, understand trends, and facilitate recruitment efforts by aligning student skills with industry requirements. Overall, the resume analyzer enhances efficiency, reduces bias, and optimizes the recruitment process for all stakeholders.

References:

- 1) Ronan Collobert, Jason Weston L'eon Bottou ,Michael Karlen, Koray Kavukcuoglu, Pavel Kuksa "Natural Language Processing (almost) from Scratch" arxiv.
- 2) "Deep Resumes Parsing with Gated Recursive Networks" by Sun, F. et al. (2017)
- 3) "A Hybrid Deep Learning Model for Resume Information Extraction" by Ren, Z. et al. (2018)
- 4) "An Ensemble Approach for Resume Parsing and Information Extraction" by Prasad, S. et al. (2019)
- 5) "BERT-Resumé: Leveraging BERT for Resumé Information Extraction" by Mishra, T. et al. (2020)

- 6) "Building a Resume Parser Using Natural Language Processing" by Patel, N. (Medium)
- 7) "How to Build a Resume Parsing Tool with Python" by Prabhu, V. (Towards Data Science)
- 8) "Resume Parsing with Python and NLP" by Poddar, S. (Analytics Vidhya)
- 9) Chavare, Dipti Suhas, and Archana Bhaskar Patil. "Resume Parsing using Natural Language Processing." *Grenze International Journal of Engineering & Technology (GIJET)* 9, no. 1 (2023).