VIVEKANAND EDUCATION SOCIETY INSTITUTE OF TECHNOLOGY CHEMBUR, MUMBAI – 400 074

A report on

Resume parsing and Ranking using NER

Submitted by

(D16AD AIDS NLP GROUP)

Athrava Baheti(03)
Arpit Burley (09)
Ashish Patil (46)

Approved by

Prof. Dr. Anjali Yeole

October 18,2024

PREFACE

The rapid digitalization of industries and the increasing volume of job applications in today's recruitment landscape have introduced challenges in processing and analyzing resumes efficiently. Human Resources departments and recruiters often face an overwhelming number of resumes for each job opening, making it crucial to extract meaningful information and rank candidates quickly. In response to this challenge, Resume Parsing and Ranking using Named Entity Recognition (NER) has emerged as an innovative solution within the field of Natural Language Processing (NLP).

This project explores the domain of automated resume parsing, focusing on extracting key information such as names, contact details, educational qualifications, work experience, and skills from resumes using NER techniques. NER is a core NLP task that helps identify and classify named entities in text, such as people, organizations, dates, and more, making it highly suitable for understanding and structuring unstructured resume data.

Furthermore, this project delves into ranking the parsed resumes based on their relevance to a given job description. The ranking system is designed to compare extracted resume features with job requirements, prioritizing candidates who match the desired skill sets and qualifications. By utilizing a combination of machine learning, deep learning models, and NLP techniques, we aim to develop a system that not only parses resumes accurately but also ranks them effectively to streamline the recruitment process.

The goal of this project is to automate the time-consuming task of manual resume screening, providing HR professionals with a tool that can rapidly parse and rank candidates based on their qualifications and fit for a job. This solution aims to enhance the efficiency of hiring processes, reduce human bias, and ensure that organizations can focus on top-quality candidates.

By automating resume parsing and ranking, this project contributes to the growing field of NLP and underscores the importance of applying NER for real-world applications, particularly in recruitment and talent acquisition. The insights gained from this project offer valuable contributions to improving hiring processes and demonstrate the potential of NLP technologies in managing and understanding vast amounts of unstructured data in the workforce.

Sincerely,
NLP group, D16AD

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to everyone who has supported and guided me throughout the development of this project, "Resume Parsing and Ranking using Named Entity Recognition (NER)." Their valuable insights, encouragement, and expertise were essential in bringing this work to completion.

First and foremost, I extend my deepest thanks to my project supervisor, Prof. Dr. Anjali Yeole, for their continuous guidance, constructive feedback, and unwavering support throughout the various stages of this project. Their expertise in Natural Language Processing (NLP) has been invaluable in shaping the direction of this work, and their suggestions have significantly contributed to the quality of the final outcome.

I would also like to extend my appreciation to my colleagues and peers who provided valuable insights and suggestions during our discussions. Their input helped refine the methodologies and approaches used in this project and enriched the overall results.

Additionally, I am grateful for the vast resources, research papers, and datasets available from the academic community and open-source platforms, which made this project possible. The use of advanced tools and technologies, especially in the fields of machine learning and NLP, has been instrumental in the success of this work.

Lastly, I would like to thank my family and friends for their constant encouragement, understanding, and patience throughout this journey. Their belief in my abilities has been a constant source of motivation.

This project is the result of the combined efforts of many individuals, and I am truly grateful to everyone who contributed to its successful completion.

Sincerely,
NLP group, D16AD

TABLE OF CONTENTS

ABSTRACT

Project Title: Resume Parsing and Ranking using Named Entity Recognition (NER)

As the number of job applications and resumes continues to increase in the digital age, the need for efficient methods to automatically extract and evaluate key information from resumes has become crucial in the recruitment process. This project explores Resume Parsing and Ranking using Named Entity Recognition (NER), a Natural Language Processing (NLP) technique that automates the extraction of important information from resumes and ranks candidates based on their relevance to specific job requirements.

In this project, we focus on utilizing NER to identify and classify key entities in resumes, such as names, contact details, educational qualifications, work experience, and skills. The extracted entities are then matched against job descriptions, and resumes are ranked based on how closely they align with the job requirements. We employ both rule-based and machine learning-based NER techniques to perform resume parsing, ensuring that various resume formats and styles are effectively processed.

To rank the parsed resumes, we implement ranking algorithms that compare extracted features (skills, experience, etc.) with the desired qualifications for a job. The ranking system prioritizes candidates who possess the most relevant experience and qualifications for the role. We evaluate the performance of our system on a set of resumes and job descriptions, using accuracy and ranking metrics to measure the effectiveness of our approach.

The results of this project demonstrate the capability of NER-based systems to automate resume screening and ranking, offering a practical solution for streamlining recruitment processes. This work contributes to the growing application of NLP in human resource management, showcasing how advanced text processing techniques can improve the efficiency and accuracy of candidate evaluation, saving time and reducing manual effort for HR professionals.

INTRODUCTION

In today's fast-paced and information-driven world, the volume of job applications has surged, especially with the advent of online job portals and digital resumes. This growing influx of resumes presents a significant challenge for recruiters and Human Resources (HR) professionals who must sift through hundreds or even thousands of resumes to find the right candidates. The traditional manual approach to resume screening is time-consuming, prone to human error, and often inefficient. Resume Parsing and Ranking using Named Entity Recognition (NER) offers an automated solution that can streamline the recruitment process by accurately extracting and evaluating key information from resumes.

The primary goal of resume parsing is to automatically extract relevant information such as personal details, education, work experience, skills, and other qualifications from resumes. Named Entity Recognition (NER), a sub-task of Natural Language Processing (NLP), plays a central role in this process by identifying and classifying specific entities within the text, such as names, organizations, dates, and job titles. Once the resumes are parsed, they can be ranked based on how well they match the requirements of a given job description, making the process of shortlisting candidates more efficient and objective.

In this project, we explore the use of both rule-based and machine learning-based NER models for resume parsing. Rule-based approaches rely on predefined patterns and linguistic rules to extract information, while machine learning-based approaches, particularly deep learning models, learn from large datasets to identify patterns and relationships in text. Furthermore, we implement ranking algorithms to compare the parsed information with job descriptions and rank candidates based on their relevance to the job.

With advancements in NLP, particularly in deep learning, models such as Long Short-Term Memory (LSTM) networks and transformer-based architectures like BERT have proven to be highly effective in understanding and processing unstructured text. These models allow for more accurate extraction and classification of entities in resumes, even when they are formatted differently or contain complex sentence structures.

The significance of this project lies in its potential to revolutionize the recruitment process by automating the time-consuming task of resume screening. Resume parsing and ranking using NER can reduce the workload for HR professionals, improve the accuracy of candidate selection, and ensure that the best-suited candidates are prioritized for interviews. This project aims to demonstrate the effectiveness of NLP in real-world HR applications, showcasing how modern text processing techniques can be leveraged to enhance hiring processes and improve decision-making.

METHODOLOGY

The project titled **"Resume Parsing and Ranking using Named Entity Recognition (NER)"** involves a detailed, step-by-step process for automatically extracting, analyzing, and ranking resumes. A user-friendly interface is developed using Streamlit, allowing for seamless interaction with the system. The methodology encompasses several stages, each playing a key role in facilitating an efficient and automated recruitment process.

1. Resume Upload

Method: The system allows users to upload resumes in formats such as PDF, DOC, DOCX, and TEXT. The Streamlit interface provides a simple drag-and-drop functionality or file selection option for ease of use.

Purpose: This step serves as the entry point for all subsequent operations. Users can upload resumes directly through the interface, ensuring that data can be processed efficiently from various file types.

2. File Reader

Method: The file reader module extracts raw text from uploaded resumes. The system employs format-specific tools, such as PDF readers and text extraction libraries for DOCX and TXT files.

Purpose: This step converts resumes into plain text, which forms the basis for further processing. It ensures that all textual content is ready for the subsequent steps of analysis.

3. Data Extraction using Named Entity Recognition (NER)

Method: Once the text is extracted, a Named Entity Recognition (NER) model is used to identify and classify key resume information, such as:

- Personal Information: Name, contact details (email, phone number).
- Professional Information: Skills, job titles, work experience, and important dates (e.g., employment periods).
- Educational Qualifications: Degrees, certifications, and other academic achievements.

Purpose: This stage translates unstructured text into structured, meaningful data, isolating key information that recruiters typically focus on. NER ensures relevant resume content is captured.

4. Resume Scoring

Method: After extracting key information, the system applies a scoring algorithm based on predefined criteria, including:

- Skills match and relevance.
- Work experience (including years of experience and job relevance).
- Educational background.

Purpose: The resume scoring provides an objective measure, ranking candidates based on how well their resume aligns with the job description. This helps prioritize candidates for further review.

5. Ranking & Analysis

Method: The resumes are ranked based on their scores, and a detailed analysis is generated. The ranking algorithm takes into account various factors such as skills, experience, and qualifications. Detailed insights into strengths, weaknesses, and overall fit for the job are provided alongside each resume.

Purpose: This step enables efficient decision-making by providing recruiters with a clear, ranked list of candidates. The detailed analysis helps assess the suitability of each candidate and highlights areas for further consideration.

6. User Interface Development with Streamlit

Method: The Streamlit framework powers the user interface, offering a highly interactive and responsive platform. The interface is designed with the following features:

- File Upload: Users can upload resumes in various formats directly through the Streamlit app.
- Real-time Scoring and Ranking: Once resumes are uploaded, users can instantly see the parsing results, resume scores, and rankings without leaving the app.
- Detailed Insights: Users can explore individual resumes in detail by clicking on specific candidates to see additional data and suggestions.

Purpose: The Streamlit interface provides an intuitive and interactive platform that allows non-technical users (such as recruiters or HR professionals) to engage with the system effortlessly. By simplifying the user experience, it ensures that users can upload resumes, view ranked results, and make informed decisions without needing to understand the underlying technical processes.

ANALYSIS

The analysis of the Text Summarization project focuses on comparing two principal methodologies: extractive and abstractive summarization. Each approach has its unique strengths and limitations, impacting their performance across various criteria such as coherence, fluency, and factual accuracy.

1. Extractive Summarization
Strengths:
- Factual Accuracy: Extractive summarization excels in maintaining high factual accuracy by selecting sentences directly from the source text, ensuring the original meaning is preserved.
- Simplicity: Techniques like TF-IDF and TextRank are straightforward to implement and computationally less intensive, making them suitable for applications requiring real-time processing.

Limitations:
- Coherence and Fluency: Extracted sentences are not always arranged in a cohesive manner, leading to summaries that may appear disjointed and lack smooth transitions.
- Redundancy: Extractive models often select repetitive sentences or those with redundant information, diminishing the efficiency and overall quality of the summary.

2. Abstractive Summarization
Strengths:
- Coherence and Readability: Abstractive models, particularly those utilizing LSTM and transformer architectures (like BERT and GPT), produce summaries that mimic human writing, characterized by coherent sentence structures and logical flow.

- Condensation: These methods effectively condense information by paraphrasing, allowing for the generation of more concise and meaningful summaries.

Limitations:
- Factual Distortion: Abstractive models, while innovative, may introduce factual inaccuracies as they generate new sentences that can sometimes misrepresent the original content.
- Complexity: Training these models is resource-intensive, requiring substantial computational power and large datasets, which can be a barrier for practical applications.

CONCLUSION

The "Resume Parsing and Ranking using Named Entity Recognition (NER)" project aimed to automate and streamline the recruitment process by efficiently extracting and ranking key information from resumes. Given the volume of resumes recruiters handle, this system offers an efficient solution for parsing data and ranking candidates based on predefined criteria, improving speed and accuracy in shortlisting.

The project utilized Named Entity Recognition (NER) to extract relevant details such as names, contact information, skills, job titles, and employment dates from resumes. Handling multiple file formats (PDF, DOC, DOCX, TXT) and applying text preprocessing ensured the system extracted relevant content with high accuracy, regardless of format. NER proved invaluable in structuring unstructured resume data, making it easier to process and analyze.

The scoring and ranking mechanism was central to the system, providing a data-driven method to evaluate candidates based on job-related criteria. This feature enabled recruiters to make informed decisions while reducing the potential for human error or bias. Detailed analysis and feedback added an extra layer of insight, helping both recruiters and candidates better understand the results.

A user-friendly interface, developed using Streamlit, allowed non-technical users to interact with the system easily. It enabled users to upload resumes, visualize rankings, and review detailed feedback, making the tool practical and accessible for HR professionals.

In summary, this project successfully demonstrated the application of NER in automating resume parsing and ranking. It addressed key challenges in recruitment, enhancing both efficiency and decision-making. While the system proved effective, future improvements could include expanding its adaptability across different industries and integrating more advanced machine learning models for greater accuracy. This project showcases the potential of AI in transforming the hiring process.

LIST OF REFERENCES

1. **Gupta, R., & Kumar, M. (2020). "Intelligent Resume Parser using NLP"**
   Proceedings of the 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS 2020). Available at:
   https://ieeexplore.ieee.org/document/9120989

2. **Ravi, N., & Patra, B. (2021). "Resume Parsing using Named Entity Recognition."**
   Proceedings of the 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCIS 2021). Available at:
   https://ieeexplore.ieee.org/document/9397511

3. **Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). "Neural Architectures for Named Entity Recognition."**
   Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2016). Available at: https://arxiv.org/abs/1603.01360

4. **Skeppstedt, M., Kvist, M., Nilsson, G. H., & Dalianis, H. (2014). "Rule-based Entity Recognition."**
   Proceedings of the International Conference on Artificial Intelligence in Medicine. Available at: https://link.springer.com/chapter/10.1007/978-3-319-07563-8_5