e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

# AUTOMATED RESUME PARSING USING NAME ENTITY RECOGNITION

**Dr. S.A. Bhavsar[*1], Rajeshwari Shinde[*2], Vaishnavi Kharche[*3], Akanksha Ghotekar[*4]**

[1]Professor, Computer Engineering, Matoshri College of Engineering and Research Centre, Nashik, India.

[*2,3,4]Student, Computer Engineering, Matoshri College of Engineering and Research Centre, Nashik, India.

## ABSTRACT

Automated resume parsing is a crucial component in modern recruitment processes, enabling the efficient extraction of relevant candidate information from resumes. Traditional methods often rely on keyword matching, which can be imprecise and overlook contextual relevance. This paper explores the application of Named Entity Recognition (NER) for automated resume parsing, offering a more accurate and context-aware approach. NER, a subset of natural language processing (NLP), involves identifying and classifying entities in text into predefined categories such as names, locations, dates, and job titles. In the context of resume parsing, NER models can be trained to recognize and extract key information such as candidate names, contact details, educational qualifications, work experience, and skills. This approach not only improves the accuracy of data extraction but also reduces the need for manual intervention, thereby speeding up the recruitment process. The proposed NER-based resume parser leverages machine learning algorithms, particularly those designed for sequence labeling tasks, to automatically identify and categorize relevant information from various resume formats. By doing so, it addresses common challenges such as the variability in resume structures and the presence of unstructured text. The implementation of this system can significantly enhance the efficiency of recruitment pipelines, enabling organizations to quickly shortlist candidates based on precise criteria. This paper discusses the development, training, and evaluation of the NER model, demonstrating its potential to revolutionize resume parsing in the hiring process.

**Keywords:** Automated Resume Parsing, Named Entity Recognition (NER), Natural Language Processing (NLP), Machine Learning, Recruitment Automation, Information Extraction, Sequence Labeling.

## I. INTRODUCTION

In today's fast-paced job market, recruiters are inundated with a high volume of resumes, making the manual screening and parsing of these documents a daunting and time-consuming task. Traditional resume parsing techniques, which often rely on simple keyword matching, are not only inefficient but also prone to errors, as they fail to capture the context and relevance of the information presented. This can lead to the overlooking of qualified candidates or the misclassification of data, which ultimately hinders the effectiveness of the recruitment process. With the increasing need for organizations to quickly and accurately identify top talent, there is a growing demand for more sophisticated, automated solutions that can handle the complexity and variability of modern resumes.

Named Entity Recognition (NER), a powerful technique within the field of natural language processing (NLP), offers a promising solution to these challenges. NER is designed to automatically identify and classify entities within text, such as names, dates, locations, and job titles, making it ideally suited for extracting key information from resumes. By implementing an NER-based automated resume parsing system, organizations can achieve a higher level of precision in data extraction, ensuring that important details are not missed, regardless of the resume format. This system not only streamlines the recruitment process by reducing manual intervention but also enhances the overall accuracy and efficiency of candidate selection, enabling recruiters to make more informed decisions and ultimately build stronger teams.

Traditional resume screening involves manually reviewing each resume to extract key information, such as a candidate's education, work experience, skills, and contact details. Recruiters often have to go through hundreds, sometimes thousands, of applications for a single job posting. This process is time-consuming, labor-intensive, and prone to human errors. Inconsistent formats, incomplete information, and subjective judgments can also make it difficult to accurately evaluate candidates. Additionally, manual screening can lead to

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

unconscious bias, where decisions are influenced by personal preferences rather than job-related criteria. As a result, traditional methods can slow down hiring, increase costs, and risk missing out on qualified candidates.

Automated Resume Parsing powered by Named Entity Recognition (NER) offers a faster and more efficient way to screen resumes. NER-based tools use Natural Language Processing (NLP) to detect and classify key information automatically, regardless of how the resume is structured. These tools can handle large volumes of resumes within minutes, ensuring consistency and reducing the chance of errors. Automated parsing also helps remove bias by focusing only on relevant data points, such as experience and skills, rather than personal information. The structured data generated through parsing makes it easy to search, filter, and rank candidates based on job requirements. Additionally, automated systems can handle multilingual resumes, integrate with Applicant Tracking Systems (ATS), and learn from new data to improve over time. This streamlined process not only speeds up hiring decisions but also ensures better candidate matching and improves the overall recruitment experience for both employers and applicants.

## II. LITERATURE SURVEY

In this study [1], the author introduces a web application designed for efficient resume parsing, serving as a vital tool in today's competitive job market. With employers often receiving hundreds or thousands of applications for each job posting, the need for an effective and automated solution for analyzing resumes is increasingly critical. This web application leverages Natural Language Processing (NLP) techniques to streamline the resume evaluation process, significantly enhancing the speed and accuracy with which recruiters can assess candidate qualifications. Built on a robust web infrastructure, the application allows users to upload resumes in various formats, including PDF, Word, and plain text. Upon uploading, the system automatically analyzes the content, extracting key information such as skills, education, work experience, and contact details. This automated extraction reduces the manual effort required by recruiters, enabling them to focus on higher-value tasks like engaging with qualified candidates. To achieve high accuracy in parsing, the application employs advanced NLP libraries and methodologies equipped with sophisticated algorithms that understand the context and semantics of the text within resumes. By leveraging machine learning models trained on diverse datasets, the system effectively identifies and categorizes information, even in non-standard formats. This precision addresses a significant challenge in traditional recruitment practices, where manual screening can lead to inconsistencies and the oversight of qualified candidates. The user-friendly design of the web application ensures that recruiters with varying technical expertise can navigate and utilize the tool easily. The intuitive interface provides straightforward options for uploading resumes and viewing parsed results, enhancing user experience and encouraging adoption among organizations looking to improve their recruitment processes. Overall, this web application represents a significant advancement in technology for talent acquisition, streamlining candidate evaluation and improving hiring decisions in the ever-evolving job market.

In this work [2], the authors present an Automated Resume Parsing and Ranking System (ARRS) that leverages Natural Language Processing (NLP) to enhance recruitment processes significantly. By automating the screening process, ARRS improves talent acquisition efficiency by extracting key information from resumes and ranking candidates according to predefined criteria. This system addresses the challenges faced by recruiters, who often sift through numerous applications, ensuring that the most qualified candidates are identified quickly and effectively. ARRS allows customization tailored to specific job requirements, enabling recruiters to define the criteria that matter most for each position. It assesses various factors, such as skills, experience, and education, to generate prioritized candidate lists, ensuring that only the most relevant applications make it to the next stage of the hiring process. This targeted approach not only saves time but also enhances the quality of candidate selection, reducing the risk of overlooking qualified individuals. The user-friendly design of ARRS is another significant advantage, ensuring that recruiters with varying levels of technical expertise can easily navigate and utilize the system. The interface simplifies the process of uploading resumes and reviewing parsed results, making it accessible to all users. Additionally, ARRS integrates seamlessly with Applicant Tracking Systems (ATS), allowing for smooth data transfer and management within existing recruitment workflows.

In this study [3], the author introduces a resume parser that integrates Named Entity Recognition (NER) with Keyword and Pattern Matching using Regular Expressions (Regex) to enhance the efficiency and accuracy of

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
( Peer-Reviewed, Open Access, Fully Refereed International Journal )
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

resume processing. The NER model leverages Natural Language Processing (NLP) libraries to identify, classify, and organize key entities such as names, contact details, skills, education, and work experience. By systematically categorizing this information, the parser streamlines candidate evaluation, making it easier for recruiters to access relevant data from resumes. Complementing NER, the Keyword and Pattern Matching component uses Regex to extract specific details like job titles, company names, and other structured information. This dual approach ensures that even nuanced details are captured accurately, improving the parser's performance across diverse resume formats. Whether resumes are submitted as PDFs, Word documents, or plain text, the system can handle varying structures and terminologies, reducing the inconsistencies and errors often encountered in manual resume screening. The combination of NER and Regex makes this resume parser a powerful tool for processing large volumes of resumes efficiently. Its ability to extract and organize critical information with high precision addresses a significant challenge in the recruitment process, where manual screening can be time-consuming and prone to human error. As a result, recruiters can focus more on engaging with qualified candidates rather than sorting through resumes. This system demonstrates its practical value by enhancing performance and accuracy, particularly in scenarios where companies need to process high volumes of applications. It provides a reliable solution for organizations of all sizes, helping streamline recruitment workflows and improve candidate shortlisting. With its capability to handle diverse formats and deliver consistent results, the parser offers an effective way to meet the growing demand for automated solutions in modern talent acquisition.

## III. SYSTEM ARCHITECTURE AND METHODOLOGY

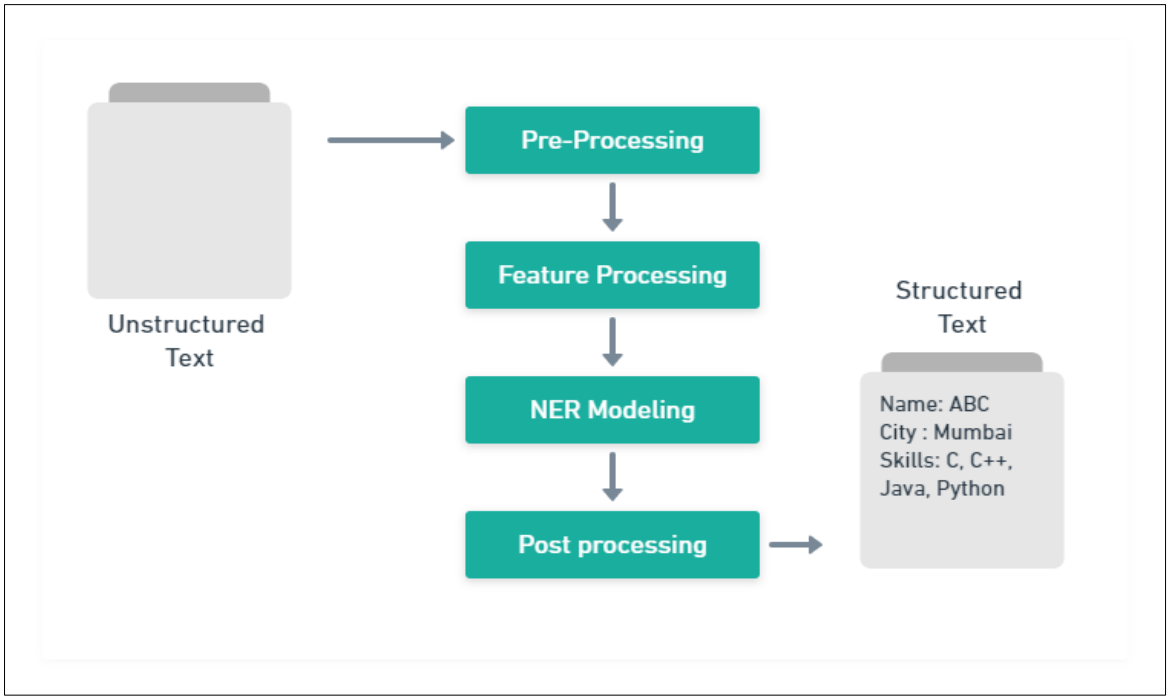


**Figure 1** - System Architecture

The methodology for developing the automated resume parsing system begins with the collection and preprocessing of a diverse set of resumes to train and validate the Named Entity Recognition (NER) model. These resumes, which come in various formats and structures, are first converted into a standardized text format to ensure consistency in processing. Preprocessing steps include tokenization, where the text is broken down into individual words or tokens, and the removal of irrelevant content such as headers, footers, and non-informative elements. This cleaned and structured data forms the basis for training the NER model to accurately identify and classify key entities such as names, contact information, educational qualifications, work experience, and skills.

The next phase involves training the NER model using machine learning algorithms. The model is trained on labelled datasets where the entities of interest are annotated manually, allowing the model to learn the patterns and context in which these entities typically appear. After training, the model undergoes rigorous testing and

validation to ensure its accuracy and robustness across different resume formats. Once validated, the NER model is integrated into a larger pipeline that automates the parsing process, extracting the relevant entities from incoming resumes and populating them into structured data formats, such as JSON or XML, that can be easily utilized by Applicant Tracking Systems (ATS) or other recruitment tools. This methodology ensures that the system is not only accurate but also adaptable to the wide variety of resume formats encountered in real-world applications.

## IV. PROJECT MODULES

The four modules involved in resume parsing using Named Entity Recognition (NER):

a. **Document Conversion -** The first step in resume parsing involves extracting text from resumes provided in various formats, such as PDFs, DOCX, or scanned images. Different tools handle various formats: PyMuPDF or pdfminer for PDFs, python-docx for Word documents, and Tesseract OCR for images. The goal is to convert the document into a plain text format to ensure it can be processed further by the NER pipeline. Handling multiple file types ensures flexibility, as job applicants may submit resumes in any of these formats.

b. **Text Preprocessing -** Text extracted from resumes may contain unnecessary elements such as formatting tags, punctuation, or irrelevant symbols. Preprocessing includes cleaning (removing noise and symbols), tokenization (splitting text into words or sentences), stopword removal (ignoring common words), and lemmatization (reducing words to their root forms). The purpose of this step is to normalize the data to ensure consistency and improve the NER model's accuracy. For example, "developing," "developed," and "development" would all be reduced to "develop."

c. **Named Entity Recognition (NER) -** NER models use machine learning or deep learning techniques to identify specific entities from the resume, such as names, email addresses, phone numbers, skills, qualifications, and job titles. Pre-trained NER models like spaCy, BERT, or custom-trained models are often employed. These models classify the extracted text into predefined categories like PERSON (for candidate names), SKILLS, ORGANIZATION, EXPERIENCE, etc. Accurate entity extraction enables structured information retrieval for further processing and analysis.

d. **Information Extraction -** The extracted entities are mapped into relevant fields or categories (e.g., name, education, contact information, and work experience). This structured data can then be used for automated processing, such as populating applicant tracking systems (ATS), generating summaries, or matching candidates with jobs. Post-processing might also include validating extracted information (e.g., checking if an email follows valid formatting) and handling missing or ambiguous data. The goal is to transform unstructured text into structured data for easy access, reporting, or decision-making.

These modules work together in a pipeline to efficiently convert raw resumes into a structured format, facilitating automated analysis and decision-making in recruitment processes.

## V. CONCLUSION

The development of an automated resume parsing system using Named Entity Recognition (NER) offers significant advancements in the recruitment process, addressing the inefficiencies and inaccuracies associated with traditional parsing methods. By leveraging NER, the system can accurately extract and categorize key information from diverse resume formats, reducing manual effort and enhancing the speed and precision of candidate selection. This not only streamlines recruitment workflows but also ensures that organizations can more effectively identify and engage with the best talent available. As recruitment continues to evolve, such technologies will play a crucial role in enabling data-driven, efficient, and fair hiring practices.

## VI. REFERENCES

[1] K. Gawhankar, A. Deorukhkar, A. Miniyar, H. Kapure and B. Ivin, "NLP-Driven ML for Resume Information Extraction," 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-6, doi: 10.1109/I2CT61223.2024.10543861.

[2] B. Nisha, V. Manobharathi, B. Jeyarajanandhini and G. Sivakamasundari, "HR Tech Analyst: Automated Resume Parsing and Ranking System through Natural Language Processing," 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2023, pp. 1681-1686, doi: 10.1109/ICACRS58579.2023.10404426.

[3] T. G. Sougandh, S. S. K, N. S. Reddy and M. Belwal, "Automated Resume Parsing: A Natural Language Processing Approach," 2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/CSITSS60515.2023.10334236.

[4] J. Zhang, X. Tan, J. Liu and Z. Liu, "Research on Named Entity Recognition Models for Cybersecurity," 2024 2nd International Conference on Signal Processing and Intelligent Computing (SPIC), Guangzhou, China, 2024, pp. 232-237, doi: 10.1109/SPIC62469.2024.10691446.

[5] W. Fulun and Z. Yonghua, "BERT-based Named Entity Recognition Method for Chinese Recipe Text," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 2021, pp. 543-547, doi: 10.1109/ICBAIE52039.2021.9390072.

[6] K. S, P. S. M, P. C and M. K, "Enhancing Named Entity Recognition using Deep Learning Approaches," 2024 5th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2024, pp. 1733-1737, doi: 10.1109/ICESC60852.2024.10690015.

[7] V. Khedkar, D. Desai, S. K. Tidke, C. Fernandes and M. R, "Chemical Named Entity Recognition for Ovarian Cancer's Drug Discovery," 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 2022, pp. 884-889, doi: 10.1109/DASA54658.2022.9765001.

[8] Z. Liu, K. Jiang, Z. Liu and T. Qin, "A Cybersecurity Named Entity Recognition Model Based on Active Learning and Self-learning," 2024 36th Chinese Control and Decision Conference (CCDC), Xi'an, China, 2024, pp. 4505-4510, doi: 10.1109/CCDC62350.2024.10587887.

[9] N. Laosen, K. Laosen and T. Paklao, "Named Entity Recognition for Thai Historical Data," 2024 21st International Joint Conference on Computer Science and Software Engineering (JCSSE), Phuket, Thailand, 2024, pp. 528-533, doi: 10.1109/JCSSE61278.2024.10613644.