

A Review Paper on Resume Parser Using AI

Dr. Mahesh B. Gunjal¹, Tejal P. Thorat², Khushi S. Muttha³, Vaishnavi C. Shete⁴, and Pragati D. Sagar⁵

^{1,2,3,4,5}Amrutvahini College of Engineering, Sangamner

Abstract—This review paper explores advancements in automated resume parsing and its implications for recruitment processes. It focuses on integrating Natural Language Processing (NLP) and machine learning techniques. With increasing volumes of resumes, efficient and accurate parsing is critical. This paper synthesizes methodologies, such as Named Entity Recognition (NER), regular expressions, and layout-aware parsing, and assesses their potential improvements for future applications.

Index Terms—Named Entity Recognition (NER), Natural Language Processing (NLP), Pattern Matching, Regular Expressions (Regex), Resume Parser

I. INTRODUCTION

In today's competitive job market, automated resume parsing has become a valuable tool. This process, which extracts structured data from resumes, can speed up candidate evaluation, enabling faster decision-making for recruiters. AI techniques, such as NLP, allow these systems to efficiently process vast amounts of information, offering improvements in candidate-job matching and screening processes. This paper reviews the methodologies employed in recent resume parsing systems.

II. LITERATURE REVIEW

1. Automated Resume Parsing: A Natural Language Processing Approach

Sougandh et al. (2023) developed a resume parser that uses NER and Regex for accurate data extraction from resumes. This system efficiently processes varied resume formats by recognizing entities like names and skills and extracting structured information like job titles and experience using regex[1].

2. On-Demand Job-Based Recruitment for Organizations Using Artificial Intelligence

Jayakumar et al. (2023) developed an AI-based resume parser that automates candidate screening using Natural Language Processing (NLP) and machine learning. The system extracts key resume details, including personal information and skills, and ranks candidates based on job suitability using a Random Forest algorithm. It achieved 92.9% accuracy

in predicting employee attrition and streamlines recruitment by automating resume parsing and candidate matching (Jayakumar et al., 2023)[2]

3. Procedural Justice and Fairness in Automated Resume Parsers

Vaishampayan et al. (2023) highlighted the ethical concerns around bias in resume parsing systems. Their study explores the candidates' perspectives on fairness, equality, and transparency in AI-based resume screening. They find that while AI can reduce human bias, it often lacks fairness for minority groups. The authors propose guidelines for enhancing fairness and inclusivity in these systems[3].

4. Layout-Aware Resume Parsing Using NLP and Rule-Based Techniques

Warusawithana et al. (2023) introduce a layout-aware parsing system that extracts section-wise content from resumes, rather than just prominent entities. This approach improves the accuracy of information extraction by considering resume layouts, and it integrates NER and rule-based techniques for better handling of structured data like personal information, education, and skills.[4]

5. Resume Parser and Job Recommendation System Using Machine Learning

Chandak et al. (2024) presented a combined resume parser and job recommendation engine, using techniques like Tf-Idf and cosine similarity to match candidates with job openings. This model significantly enhances recruitment processes by automating both resume parsing and job recommendations.[5]

6. Semi-structured Chinese Document Analysis in Resume Parsing

Zhang et al. (2009) introduced a resume parsing system tailored to semi-structured Chinese documents, focusing on the unique challenges presented by Chinese semantics and the complex structure of resumes. The study identifies the need for advanced methods like block analysis, pattern matching, and multi-level information identification to effectively

parse semi-structured resumes. Their implementation for ChinaHR, one of China's largest recruitment websites, highlights the system's ability to analyze, retrieve, and store resume data accurately and efficiently. The research contributes significantly to the field by demonstrating the successful application of these methods in real-world recruitment, offering insights for future work in semi-structured document analysis beyond resumes. Their approach is particularly relevant in dealing with the specific linguistic characteristics of Chinese, setting it apart from typical Western-based resume parsers (Zhang et al., 2009).[6]

7. Automated Navigation Systems for Job Portals Using Multilingual Chatbots and Resume Parsers

Shruti et al. (2024) conducted a comprehensive survey focusing on the application of automated navigation systems in job portals, emphasizing features such as multilingual chatbots and resume parsers. The research identifies challenges faced by job seekers, including overwhelming volumes of job listings, language barriers, and uneven formatting, which complicate interactions with application tracking systems (ATS). To address these issues, the authors suggest the integration of multilingual chatbots to improve accessibility, assist with navigation, and provide real-time guidance. Additionally, resume parsers and builders are proposed as essential tools to streamline the application process, ensuring compatibility with ATS despite varied resume formats. This paper highlights the importance of automation and multilingual support in making job portals more inclusive and user-friendly, contributing to improved user experiences in diverse global markets (Shruti et al., 2024).[7]

8. Job Applications Selection Using Natural Language Processing and Machine Learning

Pimpalkar et al. (2023) explored the application of natural language processing (NLP) and machine learning (ML) in enhancing the efficiency of resume selection for job applications. The study emphasizes the challenges recruiters face when processing large volumes of resumes and the need for automated solutions. By applying NLP and ML, the system can analyze unstructured data, extract relevant candidate information, and align it with job requirements more effectively. This approach helps expedite the recruitment process, reducing manual effort while maintaining accuracy in identifying candidates with the desired skills and qualifications. The study also

identifies research gaps in current methodologies, suggesting improvements in analyzing writing styles, word choices, and grammar. This paper offers significant insights into the future of resume analysis, proposing advancements that would allow machines to interpret resumes similarly to how humans do (Pimpalkar et al., 2023).[8]

9. Automated Resume Parsing and Ranking Using NLP and Deep Learning

Thangaramya et al. (2024) introduced an advanced resume parsing and ranking system that leverages natural language processing (NLP) and deep learning (DL) techniques to streamline the recruitment process. The system, referred to as the "Resume Parser and Ranker," is designed to extract key information from unstructured resumes, making the recruitment process more efficient. The paper highlights the use of Named Entity Recognition (NER) powered by deep learning, achieving a high accuracy rate of 93% in information extraction, close to human-level performance. This system incorporates heuristic calculations to rank candidates based on their resume content, which saves time and reduces manual effort in shortlisting candidates for interviews. The study underscores the benefits of automating resume parsing and ranking in improving recruitment accuracy and efficiency, providing an essential tool for companies to handle large volumes of resumes in a competitive job market (Thangaramya et al., 2024)[9]

10. AI-Driven Resume Parsing and Recruitment with NLP and Generative AI

Abisha et al. (2024) introduced "Resspar," an AI-driven resume parsing and recruitment system that integrates Natural Language Processing (NLP) and Generative AI techniques to automate the hiring process. The system employs a Language Model (LLM) for parsing resumes and uses Google's GenAI API to extract structured data from images of resumes, converting visual information into meaningful text. Resspar's capabilities extend beyond traditional parsing; it provides a web-based interface that allows recruiters to upload resumes, which are then parsed to extract key details such as names, emails, phone numbers, and professional skills. Additionally, the system features a powerful filtering mechanism to match candidates' skills with job requirements. By utilizing advanced AI technologies, including Prompt Engineering, Resspar aims to reduce the manual labor involved in resume screening while promoting fairness in recruitment by objectively assessing

candidates based on their qualifications. This approach offers a comprehensive solution to streamline the recruitment process, making it more efficient and less prone to human error (Abisha et al., 2024)[10]

11. Hybrid Resume Parsing and Matching Using RegEx and NER

Murthy et al. (2023) proposed a hybrid approach to resume parsing and matching that combines rule-based methods and Named Entity Recognition (NER) for more efficient candidate selection in the recruitment process. The system utilizes Regular Expressions (RegEx) and pre-trained models from SpaCy to extract key candidate details such as name, phone number, email, experience, and skills. This hybrid approach significantly improves the accuracy of information extraction from resumes. The scoring mechanism, based on cosine similarity between the candidate's resume and the job description, helps rank candidates in terms of suitability. The use of SBERT for vectorizing resumes and job descriptions further enhances the system's performance, achieving a parsing accuracy of 70%. This method allows recruiters to prioritize applicants based on relevance, thus saving time and improving the recruitment process by automating initial screening efforts (Murthy et al., 2023)[11]

12. Resume Analyzer and Skill Enhancement Recommender System Using CNN and Machine Learning

Prashanth et al. (2024) introduced a resume analyzer system that not only parses resumes but also predicts suitable job titles based on the candidate's resume content. The system employs a custom Convolutional Neural Network (CNN) model, which outperforms other machine learning models such as Random Forest and Support Vector Machines (SVM), as well as neural network models like BERT. Additionally, the system integrates word2vec vectorization and cosine similarity to match resumes with job descriptions, providing an efficient way to assess compatibility between applicants and job roles. A unique aspect of the system is its feedback mechanism, which scores resumes based on their contents—such as education, skills, and experience—and provides personalized feedback to applicants. This dual functionality of analyzing resumes and recommending improvements makes the system particularly valuable in enhancing both recruitment efficiency and candidate preparedness (Prashanth et al., 2024) [12]

13. Syntax Error Repair in LR-based Parsers with Dynamic Valid Length

Kiyama and Ashihara (2008) proposed an innovative approach to handling syntax errors in LR-based parsers through dynamic valid length error repair. Traditional syntax error repair methods rely on fixed valid lengths to decide whether an error is spurious, which can lead to inadequate error handling. This paper introduces a dynamic valid length mechanism, allowing the parser to evaluate all possible error repair candidates and select the longest valid sequence. The experimental results show that this approach significantly improves error correction, reducing non-correcting errors by approximately 90% and increasing correcting errors by 23.1% compared to conventional methods. This technique enhances the efficiency of error repair in compilers, allowing continuous parsing and improving the overall quality of error messages during compilation (Kiyama & Ashihara, 2008) [13]

14. Resume Parsing Framework for E-recruitment

Sajid et al. (2022) proposed a resume parsing framework to address the challenges posed by the diverse formats of resumes in e-recruitment systems. The study identifies the limitations of traditional methods—such as rule-based and supervised approaches—in handling the variety of resume styles, which often feature different fonts, colors, and table formats. These variations negatively impact the accuracy of data mining operations, including information extraction and applicant ranking. The proposed framework tackles these issues by employing text block classification to separate content and named entity recognition (NER) enriched with ontology to extract key information. By improving the accuracy of resume parsing, this system directly contributes to selecting the best candidates for a given job role. This framework enhances the efficiency of the recruitment process by overcoming the constraints of data annotation and knowledge incompleteness that affect previous techniques (Sajid et al., 2022).[14]

15. Ontology-Based Resume Parser for Job Matching

Çelik et al. (2013) proposed an Ontology-based Resume Parser (ORP) designed to transform free-form resumes into a structured ontological model. This project, a collaboration between Kariyer.net and TUBITAK, addresses the challenge of diverse resume formats that complicate data storage in relational databases. By employing an ontological structure, the system aims to enhance the efficiency of matching job

seekers with suitable employment opportunities. The ORP is tested on various Turkish and English resumes and is built on a Semantic Web approach, facilitating more effective expert finding for companies.[15]

16. A Review on Text Analytics Process with a CV Parser Model

Das et al. (2018) presented a comprehensive overview of the text analytics process with a focus on a CV parser model. The study highlights the challenges posed by the vast amounts of data generated today, particularly emphasizing the characteristics of big data, which include velocity, volume, and variety. The authors explain that a significant portion of this data is text-based and discuss the importance of deriving high-quality structured data from unstructured text through text analytics. The proposed CV parser model integrates with recruitment workflows, automating the processing of incoming CVs and extracting essential entities required for recruitment. By converting free-form CV documents into a structured XML format, the model enhances the efficiency of the recruitment process, allowing for better storage and evaluation of candidate data. The paper outlines the role of text analytics in enriching data and the utility of CV parsing tools in recruitment, ultimately contributing to more effective candidate selection in companies (Das et al., 2018). [16]

17. Intelligent Recruitment System Using NLP

Sharma et al. (2021) addresses the pressing issue of unemployment among the youth in India, highlighting the disparity between the abundance of job opportunities across various sectors and the overwhelming number of applications received for limited vacancies. The study emphasizes the need for efficient recruitment processes, as manual screening of resumes can be time-consuming and may result in organizations missing out on qualified candidates. To tackle this problem, the authors propose an intelligent recruitment system that utilizes Natural Language Processing (NLP) techniques to extract and analyze data from resumes. The proposed Resume Parser is designed to transform the extracted data into valuable insights, enabling recruiters to quickly identify the most relevant candidates. By streamlining the recruitment process, this system not only saves time and effort for recruiters but also enhances the quality of candidate selection, thereby contributing to more effective hiring practices in the competitive job market (Sharma et al., 2021).[17]

18. Improving Incremental Construction of Knowledge Bases by Using Terminological Logic Resources

Julia et al. (1998) proposed a novel system aimed at the incremental construction of a knowledge base (KB) utilizing terminological logic (TL) subsumption criteria. The system focuses on the integration of assertions derived from a syntactic and semantic parser, specifically targeting natural language sentences pertinent to the medical domain. Each parsed sentence is converted into a predicate calculus (PC) formula, with its variables annotated using TL expressions. These formulas are subsequently stored in the KB in accordance with TL subsumption criteria, which help prevent the introduction of redundancies and contradictions. By employing these criteria, the proposed system enhances the simplicity of the KB and optimizes the proof tree during information retrieval. This approach is particularly beneficial for managing complex medical information, facilitating more efficient knowledge management and retrieval processes (Julia et al., 1998).[18]

III. METHODOLOGY

A. Overview of Technologies

The reviewed papers employ several key technologies in resume parsing systems:

1. NER: Used to categorize entities, enhancing data extraction accuracy [3][4].
2. Regular Expressions (Regex): Applied to extract structured data across various resume formats.[3][4]
3. Machine Learning Algorithms: Utilized for job matching and recommendation systems. Techniques like Tf-Idf vectorization and cosine similarity were commonly employed [3].
4. Layout-Aware Parsing: A system that identifies text sections in resumes, improving the contextual accuracy of the data extracted by considering layout formatting [4].

B. Steps of Resume Parsing

1. Data Extraction: Extracting text from PDF resumes using frameworks like PDFMiner.[3][4]
- 2.NLP Techniques: Tokenization and entity recognition using tools like spaCy to identify key resume information.[3][4]
- 3.Layout Awareness: Extracting section-based content from resumes based on layout analysis, addressing gaps in traditional content extraction methods.[4]
4. Job Recommendation Integration: Connecting parsed resume data with machine learning models to recommend suitable job openings [2][5].

IV. CHALLENGES AND SOLUTIONS

A. Handling Multiple Resume Formats

Traditional parsers struggle with the diversity of resume formats. The integration of layout-aware techniques resolves these challenges by parsing section-wise content instead of extracting only key entities[4].

B. Fairness and Bias in Parsing

Ethical issues, such as bias in AI-driven parsing, remain a significant concern. Vaishampayan et al. propose enhancing procedural justice by implementing fairness checks in AI algorithms to ensure inclusivity [3].

C. Scalability

The need for scalable systems is critical, especially for handling large volumes of resumes. The combination of layout-aware parsing and machine learning-driven job recommendations facilitates efficient processing of applications [4][3].

V. EVALUATION

A. Performance Metrics

Performance of resume parsers is often measured using accuracy, precision, recall, and F1-score. The layout-aware system by Warusawithana et al. achieved high accuracy, especially in extracting structured data like education and work experience, showcasing a superior F1-score compared to traditional parsers[4].

B. Real-World Applications

Real-world applications of these technologies have shown their effectiveness in improving recruitment processes by streamlining resume parsing and job recommendations. Systems incorporating layout-aware techniques further enhance the accuracy of information extraction[3][4].

VI. FUTURE DIRECTION

A. Enhancing Parsing Techniques

Future research could focus on improving parsing accuracy by integrating more advanced machine learning models, such as deep learning approaches for better contextual understanding[4].

B. Bias Mitigation

Addressing bias in AI models, particularly in resume parsing, remains an important challenge. Regular audits and fairness checks could help reduce biases in the recruitment process[3].

VII. CONCLUSION

Automated resume parsing, powered by AI, has shown great promise in improving the efficiency of recruitment processes. The integration of layout-aware techniques, NER, and rule-based approaches offers significant advancements in the accuracy of data extraction. Future improvements in fairness and inclusivity, as well as leveraging machine learning, will further enhance the capabilities of these systems.

ACKNOWLEDGMENT

We would like to express our gratitude to all individuals and organizations who contributed to the completion of this research. Special thanks to the Department of Computer Engineering at Amrutvahini College of Engineering, Sangamner, India, for their support and resources. We also extend our appreciation to the research participants and collaborators whose valuable insights and contributions were instrumental in shaping this study.

REFERENCES

- [1] Thatavarthi Giri Sougandh,Sai Snehith K,Nithish Sagar Reddy,Meena Belwal."Automated Resume Parsing: A Natural Language Processing Approach", DOI: 10.1109/CSITSS60515.2023.10334236
- [2] Nithya Jayakumar,A. K. Maheshwaran,P. Saurav Arvind,G. Vijayaragavan."On-Demand Job-Based Recruitment For Organisations Using Artificial Intelligence" DOI: 10.1109/ICNWC57852.2023.10127551
- [3] Swanand Vaishampayan, Sahar Farzanehpour, Chris Brown."Procedural Justice and Fairness in Automated Resume Parsers for Tech Hiring: Insights from Candidate Perspectives" DOI: 10.1109/VL-HCC57772.2023.00019
- [4] S.P.Warusawithana ,N.N.Perera, R.L.Weerasinghe, T.M.Hindakaraldeniya, G.U.Ganegoda."Layout Aware Resume Parsing Using NLP and Rule-based Techniques" DOI: 10.1109/ICITR61062.2023.10382773
- [5] Ashish Virendra Chandak, Hardik Pandey,Gourav Rushiya,Harsh Sharma."Resume Parser and Job Recommendation System using Machine Learning" DOI:10.1109/ESIC60604.2024.10481635
- [6] Zhang Chuang, Wu Ming, Li Chun Guang, Xiao Bo, Lin Zhi-qing. "Resume Parser: Semi-structured Chinese Document Analysis." 2009 WRI World Congress on Computer Science and Information Engineering, Los Angeles,

- CA, USA, March 31 - April 2, 2009. IEEE, DOI: [10.1109/CSIE.2009.562] (<https://doi.org/10.1109/CSIE.2009.562>).
- [7] Shruti, G. R., Ajith, R., Bhat, S. P., Ranganath, L., & Menon, M. S. (2024). "Automated Navigation Systems for Job Portals – A Survey on Multilingual Chatbots and Resume Builders and Parsers." 2024 International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS), Bengaluru, India, April 22-23, 2024. IEEE, DOI: 10.1109/ICETCS61022.2024.10544358.
- [8] Pimpalkar, A., Lalwani, A., Chaudhari, R., Inshall, M., Dalwani, M., & Saluja, T. (2023). "Job Applications Selection and Identification: Study of Resumes with Natural Language Processing and Machine Learning." 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, February 18-19, 2023. IEEE, DOI: 10.1109/SCEECS57921.2023.10063010.
- [9] Thangaramya, K., Logeswari, G., Gajendran, S., Roselind, J. D., & Ahirwar, N. (2024). "Automated Resume Parsing and Ranking using Natural Language Processing." 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT), Vellore, India, May 3-4, 2024. IEEE, DOI: 10.1109/AIIoT58432.2024.10574696.
- [10] Abisha, D., Keerthana, S., Evanjaline, N., Kavitha, K., Mary, J. S., & Ramya, R. (2024). "Respar: AI-Driven Resume Parsing and Recruitment System using NLP and Generative AI." 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, August 28-30, 2024. IEEE, DOI: 10.1109/ICoICI62503.2024.10696451.
- [11] Murthy, G. G., Abhi, S., & Agarwal, R. (2023). "A Hybrid Resume Parser and Matcher using RegEx and NER." 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT), Faridabad, India, November 23-24, 2023. IEEE, DOI: 10.1109/ICAICCIT60255.2023.10466097.
- [12] Prashanth, V. J., Jathin, S. N. P., Gopinath, S., Udith, S., & Kavitha, C. R. (2024). "Resume Analyzer and Skill Enhancement Recommender System." 2024 Asia Pacific Conference on Innovation in Technology (APCIT), Mysore, India, July 26-27, 2024. IEEE, DOI: 10.1109/APCIT62007.2024.10673530.
- [13] Kiyama, M., & Ashihara, H. (2008). "Syntax Error Repair with Dynamic Valid Length in LR-based Parsers." 2008 International Conference on Control, Automation and Systems, Seoul, Korea, October 14-17, 2008. IEEE, DOI: 10.1109/ICCAS.2008.4694256.
- [14] Sajid, H., Kanwal, J., Bhatti, S. U. R., Qureshi, S. A., Basharat, A., & Hussain, S. (2022). "Resume Parsing Framework for E-recruitment." 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Korea, January 3-5, 2022. IEEE, DOI: 10.1109/IMCOM53663.2022.9721762.
- [15] D. Çelik, A. Karakas, G. Bal, C. Gültunca, A. Elçi, B. Buluz, "Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs," in 2013 IEEE 37th Annual Computer Software and Applications Conference Workshops, 22-26 July 2013, Japan. DOI: 10.1109/COMPSACW.2013.60.
- [16] Das, P., Sahoo, B., & Pandey, M. (2018). "A Review on Text Analytics Process with a CV Parser Model." 2018 3rd International Conference for Convergence in Technology (I2CT), Pune, India, April 6-8, 2018. IEEE, DOI: 10.1109/I2CT.2018.8529772.
- [17] Sharma, A., Singhal, S., & Ajudia, D. (2021). "Intelligent Recruitment System Using NLP." 2021 International Conference on Artificial Intelligence and Machine Vision (AIMV), Gandhinagar, India, September 24-26, 2021. IEEE, DOI: 10.1109/AIMV53313.2021.9670958.
- [18] Julia, R. M. da S., Pereira, A. E. C., Arantes, W. M., & Guillen, A. M. S. (1998). "Improving Incremental Construction of Knowledge Bases by Using Terminological Logic Resources." SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, USA, October 14-14, 1998. IEEE, DOI: 10.1109/ICSMC.1998.728108.