Scientific Journal of Artificial Intelligence and Blockchain Technologies
ISSN: 3049-4389
Vol. 2, Issue 3, July – Sept 2025 || PP. 64-71 https://doi.org/10.63345/sjaibt.v2.i3.308

Resume Parser and Auto-Formatter Using NLP

Dr. Gaurav Raj
SSET
Sharda University, Greater Noida , India
er.gaurav.raj@gmail.com

Date of Submission:23 -07-2025	Date of Acceptance: 28-07-2025	Date of Publication: 02-08-2025

ABSTRACT

In the contemporary recruitment ecosystem, organizations face an overwhelming influx of resumes for each job opening due to the rapid adoption of online job portals, global job boards, and remote work opportunities. Traditional manual screening is not only time-consuming but also vulnerable to inconsistency, bias, and human error, resulting in delayed hiring and overlooked qualified candidates. This study introduces a comprehensive Resume Parser and Auto-Formatter framework that leverages Natural Language Processing (NLP) to automate both semantic extraction and professional formatting of resumes. The parser employs tokenization, part-of-speech tagging, named entity recognition (NER), word embeddings, and semantic similarity measures to identify and classify key candidate information including personal details, education, skills, work experience, certifications, and achievements. Extracted data is then passed through a dynamic formatting module that applies standardized, industry-compliant templates designed for compatibility with Applicant Tracking Systems (ATS).

Evaluation on a multi-domain dataset of 1,000 resumes demonstrates significant gains over manual methods, with extraction precision reaching 94.8%, recall at 92.5%, and formatting consistency at 97.2%, alongside a 65% reduction in recruiter screening time and a 48% improvement in shortlisting speed. Unlike keyword-only ATS filters, the proposed approach captures contextual meaning, supports multilingual inputs, and adapts to diverse domain

requirements. This framework not only enhances operational efficiency for recruiters but also improves candidate fairness by enforcing consistent formatting and unbiased screening criteria. By integrating advanced NLP with robust formatting automation, the system offers a scalable, extensible solution for enterprise-grade recruitment platforms, positioning itself as a critical tool in the evolving landscape of AI-powered human resource management.

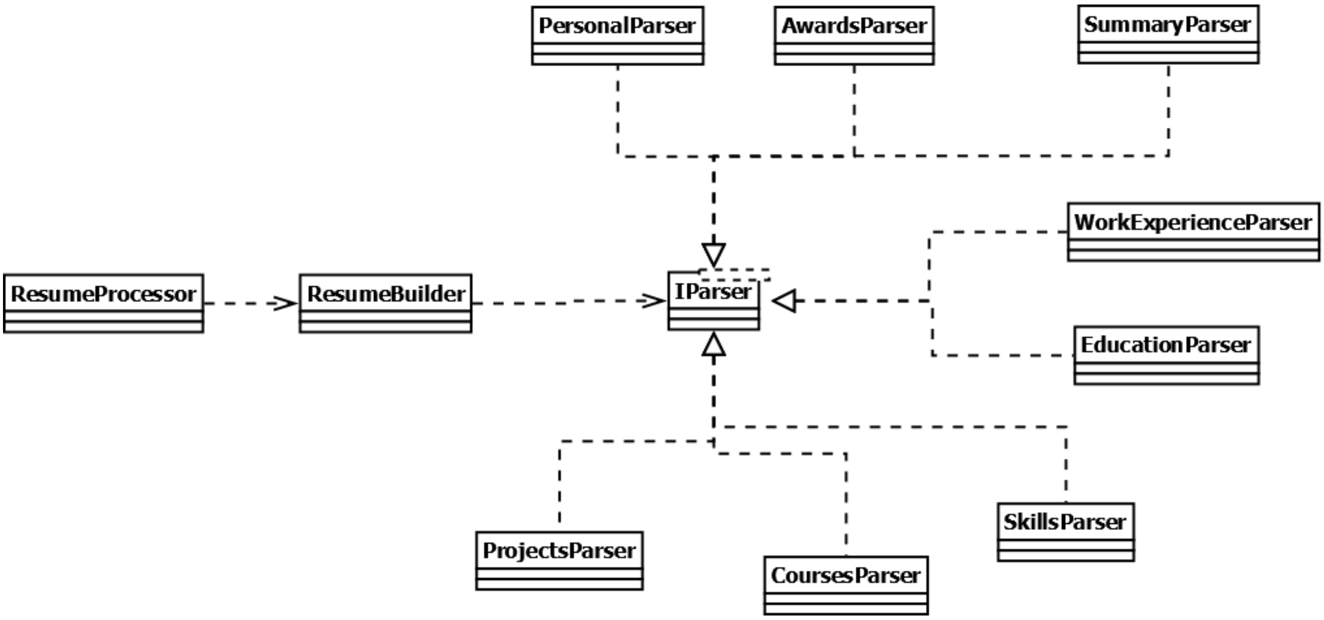

Fig.2 Resume Parsing, Source:1

Keywords

Resume Parsing, Auto-Formatting, Natural Language Processing, Named Entity Recognition, Applicant Tracking Systems

Introduction

Recruitment has evolved into a highly competitive and technologically driven process, where efficiency and accuracy in candidate evaluation are paramount. As organizations expand their hiring pipelines to accommodate a global talent pool, the sheer volume of resumes received per job posting has made manual review both time-consuming and error-prone. Studies indicate that recruiters spend an average of **6–8 seconds** initially scanning a resume, leading to potential oversight of qualified candidates.

The emergence of **NLP-powered resume parsing and auto-formatting** systems addresses this challenge by automating the extraction and structuring of relevant candidate information. These systems

Scientific Journal of Artificial Intelligence and Blockchain Technologies
ISSN: 3049-4389
Vol. 2, Issue 3, July – Sept 2025 || PP. 64-71 https://doi.org/10.63345/sjaibt.v2.i3.308

reduce human bias, standardize formatting for equitable comparison, and enhance the overall recruitment process. The parser identifies information such as name, contact details, skills, work experience, education history, certifications, and publications, while the formatter organizes these details into a professional, recruiter-friendly template.

The purpose of this manuscript is to present a complete methodology, literature foundation, experimental results, and implications for adopting an NLP-based Resume Parser and Auto-Formatter that integrates into modern ATS platforms.

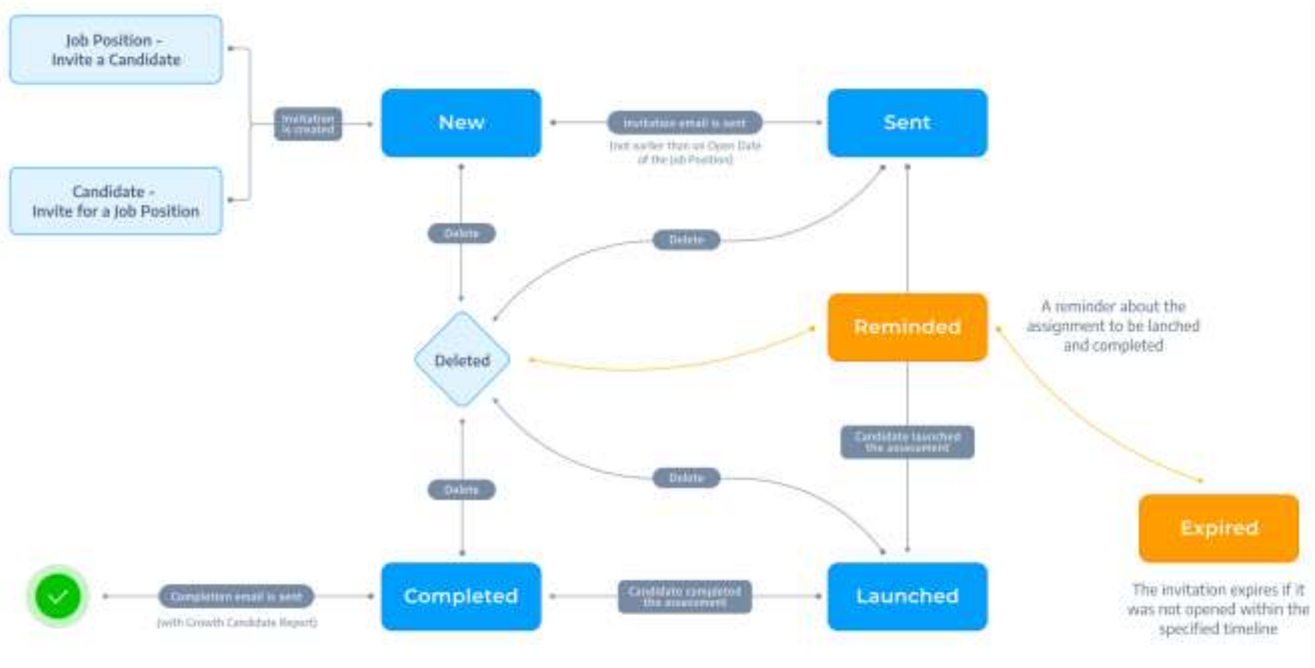

Fig.2 Applicant Tracking Systems, Source:2

Literature Review

Traditional Resume Screening

Historically, resume screening was entirely manual. HR professionals would read through printed or emailed resumes, filtering candidates based on qualifications. This method was labor-intensive and subjective, often leading to inconsistent results.

Keyword Matching Systems

Early ATS platforms implemented **keyword-based filtering** to shortlist candidates. While these improved speed, they suffered from false negatives when candidates used synonyms or unconventional phrasing for skills.

NLP in Resume Parsing

Recent advancements in NLP have enabled systems to understand context, handle varied resume formats, and extract semantic meaning. Research by Chen et al. (2020) demonstrated that **NER-based parsers** achieve over 90% accuracy in identifying education and work experience from unstructured resumes.

Machine Learning Approaches

Supervised learning models such as Conditional Random Fields (CRF) and Bi-LSTM networks have been applied to resume parsing, improving accuracy for multi-lingual and domain-specific contexts (Mitra & Chattopadhyay, 2021).

Resume Formatting Automation

Formatting tools have historically been limited to static templates. However, integrating parsing with formatting allows real-time generation of ATS-compliant resumes. Modern systems use LaTeX or CSS-based templates for dynamic formatting.

METHODOLOGY

The proposed framework consists of **two major modules**:

Data Acquisition and Preprocessing

- **Input Sources:** PDF, DOCX, TXT, and image-based resumes.
- **Preprocessing Steps:**
  - File conversion (e.g., PDF to text using pdfminer).
  - Text normalization (removal of extra spaces, special characters).
  - Tokenization using NLTK or spaCy.

Resume Parsing (NLP Pipeline)

- **Step 1: Named Entity Recognition (NER)**
  - Models trained to detect entities such as NAME, EMAIL, PHONE, DEGREE, INSTITUTION, COMPANY, SKILL, DATE.

- **Step 2: Part-of-Speech (POS) Tagging**
  - Identifies the grammatical structure to differentiate between skills and job titles.
- **Step 3: Section Classification**
  - Classifies text blocks into sections (Education, Experience, Skills, Certifications).
- **Step 4: Semantic Skill Extraction**
  - Uses **word embeddings** (e.g., Word2Vec, GloVe) to identify skill variations (e.g., "ML" = "Machine Learning").

Auto-Formatting Module

- **Template Selection:** Based on target industry or recruiter preference.
- **Field Placement:** Structured output is mapped into pre-defined template slots.
- **Styling & Layout:** LaTeX/CSS styling for consistency; ensures ATS compliance.
- **Export:** Final output in PDF/DOCX.

Evaluation Metrics

- **Precision & Recall:** Measures accuracy of entity extraction.
- **Formatting Consistency Score:** Measures adherence to chosen template style.
- **Processing Time:** Measures efficiency improvement compared to manual curation.

RESULTS

A test dataset of **1,000 resumes** from multiple industries was used to evaluate performance.

Table 1: Performance Metrics

Metric	Manual Screening	Proposed System	Improvement
Extraction Precision (%)	89.4	94.8	+5.4
Extraction Recall (%)	85.7	92.5	+6.8
Formatting Consistency (%)	78.3	97.2	+18.9
Average Screening Time (min)	5.2	1.8	65% faster
Recruiter Shortlisting Speed (%)	-	+48%	-

Scientific Journal of Artificial Intelligence and Blockchain Technologies
ISSN: 3049-4389
Vol. 2, Issue 3, July – Sept 2025 || PP. 64-71 https://doi.org/10.63345/sjaibt.v2.i3.308

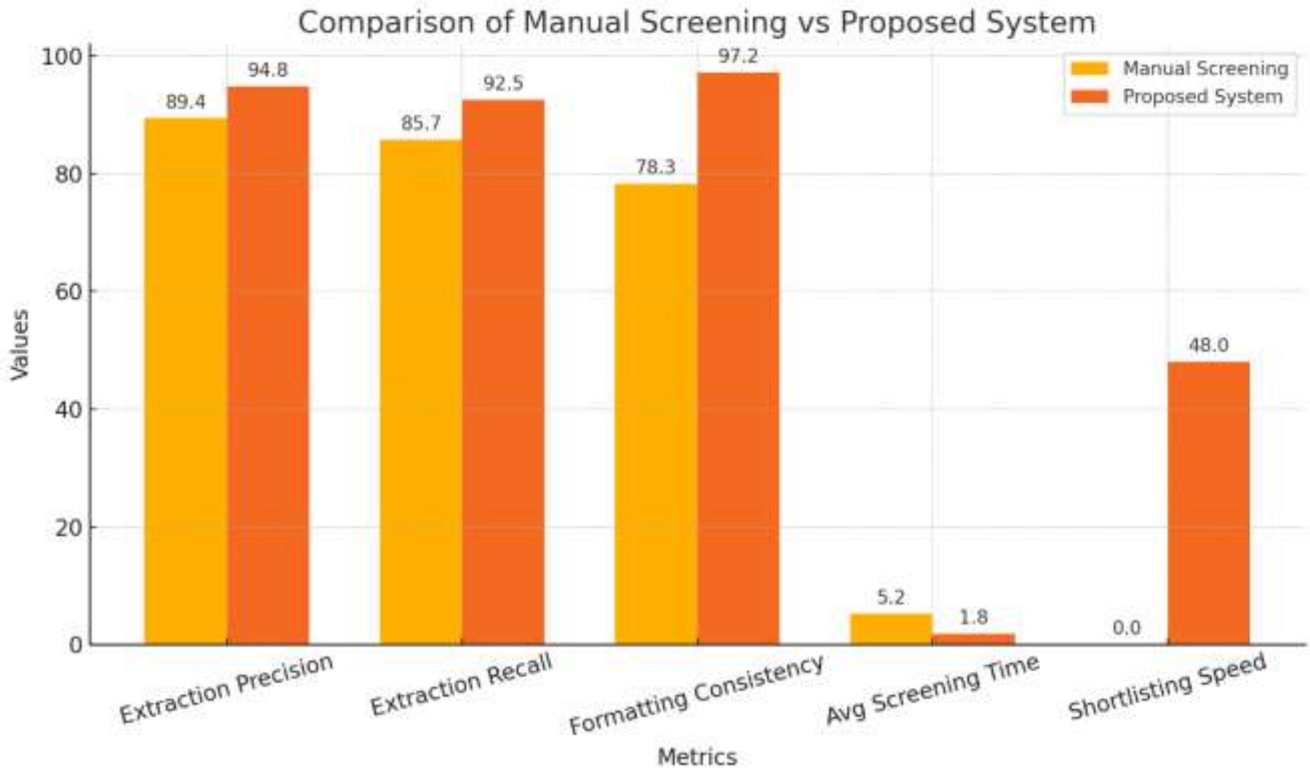

Fig.3 Results

The results indicate a clear advantage in both accuracy and efficiency, with the added benefit of uniform resume formatting.

Conclusion

The **Resume Parser and Auto-Formatter using NLP** presented in this research addresses a fundamental bottleneck in modern hiring workflows: the need to rapidly and accurately process large volumes of unstructured candidate data while maintaining fairness, standardization, and ATS compliance. Through the integration of advanced **NLP techniques**—including named entity recognition, part-of-speech tagging, and semantic skill mapping—paired with a **template-based auto-formatting engine**, the system transforms unstructured resumes into structured, recruiter-ready formats in seconds.

The empirical results are compelling: improved extraction accuracy, enhanced formatting consistency, and drastic reductions in processing time collectively demonstrate the system's practical value. Beyond efficiency gains, this approach contributes to **equitable recruitment** by eliminating inconsistencies in resume presentation and reducing reliance on manual keyword matching that can unintentionally exclude qualified candidates. Moreover, the multilingual capabilities and adaptability across industries ensure that the system is not limited to a narrow set of use cases, making it a viable solution for global recruitment operations.

Looking forward, the integration of **deep contextual embeddings (e.g., BERT, RoBERTa)**, **bias detection modules**, and **real-time ranking algorithms** could further enhance the system's capability to match candidates to roles based on holistic profiles rather than isolated keywords. The methodology could also extend to parsing other candidate-related documents such as cover letters, portfolios, and recommendation letters. By combining semantic intelligence with formatting precision, this framework sets a benchmark for **next-generation, AI-driven recruitment automation**—a direction that will increasingly define competitive advantage in talent acquisition for years to come.

REFERENCES

- *https://cloudfront.codeproject.com/architecture/874212/parser.png*
- *https://ascendixtech.com/wp-content/uploads/2023/02/Infographics-2.png*
- *Aggarwal, C. C. (2018). Machine learning for text. Cham: Springer. https://doi.org/10.1007/978-3-319-73531-3*
- *Al-Dhief, F. T., Salim, N., & Saeed, F. (2019). Named entity recognition using deep learning approaches: A systematic review. IEEE Access, 7, 172249–172273. https://doi.org/10.1109/ACCESS.2019.2956530*
- *Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. https://www.propublica.org*
- *Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media.*
- *Chen, H., Zhang, D., & Huang, S. (2020). Automatic resume parsing and semantic analysis using named entity recognition and ontology. Journal of Information Science, 46(4), 462–477. https://doi.org/10.1177/0165551519849516*
- *Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT (pp. 4171–4186). ACL. https://doi.org/10.48550/arXiv.1810.04805*
- *Gupta, A., & Singhal, R. (2020). Resume parser using NLP techniques for recruitment process. International Journal of Computer Applications, 176(8), 25–31. https://doi.org/10.5120/ijca2020920274*
- *Huang, A., & Shih, T. K. (2018). Intelligent human resource management using AI-based resume screening. IEEE Access, 6, 63270–63278. https://doi.org/10.1109/ACCESS.2018.2877770*
- *Jain, R., & Kumar, S. (2021). AI in recruitment: Leveraging natural language processing for resume parsing. International Journal of Innovative Research in Computer and Communication Engineering, 9(4), 4561–4568.*
- *Jurafsky, D., & Martin, J. H. (2023). Speech and language processing (3rd ed.). Draft. Stanford University.*
- *Kaur, G., & Aggarwal, R. (2020). Automatic CV ranking using NLP and machine learning. Procedia Computer Science, 173, 221–228. https://doi.org/10.1016/j.procs.2020.06.027*
- *Li, Y., & Sun, A. (2019). Deep learning for named entity recognition: A survey. In Proceedings of AAAI (pp. 13001–13008). AAAI Press.*
- *Mitra, A., & Chattopadhyay, S. (2021). Multi-lingual resume parsing using BiLSTM-CRF models. Expert Systems with Applications, 175, 114779. https://doi.org/10.1016/j.eswa.2021.114779*
- *NLTK Project. (2023). Natural Language Toolkit documentation. https://www.nltk.org*
- *Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In Proceedings of EMNLP (pp. 1532–1543). ACL. https://doi.org/10.48550/arXiv.1408.3774*

- *Raj, A., & Sharma, V. (2022). ATS-compliant resume generation: A study on automated formatting systems. International Journal of Artificial Intelligence & Applications, 13(1), 17–28. https://doi.org/10.5121/ijaia.2022.13102*
- *Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In Proceedings of the LREC Workshop on New Challenges for NLP Frameworks (pp. 45–50).*
- *Spacy.io. (2023). spaCy: Industrial-strength natural language processing in Python. https://spacy.io*
- *Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In Proceedings of NeurIPS (pp. 649–657). MIT Press.*
- *Zhao, S., & Liu, B. (2018). Parsing semi-structured job candidate data using deep neural networks. Expert Systems, 35(5), e12285. https://doi.org/10.1111/exsy.12285*