

# Smart Resume Parser - Final Project Report

Author: Manoj Reddy

Year: 2025

## 1. Objective

To develop an intelligent system that extracts structured information from unstructured resume files (PDF/DOCX) using Natural Language Processing.

## 2. Abstract

Recruiters often receive hundreds of resumes in varying formats. Manually extracting relevant data like skills, education, and experience is time-consuming. The Smart Resume Parser automates this process using Python, spaCy, and Streamlit, providing a fast and accurate way to organize resume information into structured formats such as JSON or CSV.

## 3. Tools & Technologies Used

- 1 Python 3.x - Core programming language
- 2 spaCy - NLP-based text processing
- 3 PyMuPDF (fitz) - PDF text extraction
- 4 python-docx - DOCX text extraction
- 5 regex - Pattern-based field extraction
- 6 pandas - Structuring and exporting data
- 7 Streamlit - Interactive UI for testing and visualization

## 4. System Architecture

Input: Resume files (.pdf, .docx) Processing Steps: 1. Extract text using PyMuPDF or python-docx 2. Preprocess (clean, normalize) 3. Apply spaCy NLP model for entity recognition 4. Use regex patterns to extract fields 5. Store structured output as JSON/CSV Output: Structured data displayed on UI

## 5. Workflow Diagram

Upload Resume → Extract Text → NLP & Regex Parsing → Display Results → Export (CSV/JSON)

## 6. Key Modules

- 1 parser.py - Handles extraction logic (NLP + regex)
- 2 utils.py - Helper functions for cleaning and text operations
- 3 app.py - Streamlit web interface

4 `make_resumes.py` - Generates sample resumes for testing

## 7. Sample Output

Example structured output extracted from a sample resume: Name: John Smith Email: john.smith@example.com Skills: Python, React, SQL, AWS Experience: Acme Corp (2020–Present) Education: B.Sc. Computer Science

## 8. Results & Observations

✓ Successfully parsed skills, experience, and education from multiple resume formats. ✓ High accuracy with structured resumes. ✓ Minor inconsistencies with unstructured resumes (scope for improvement).

## 9. Conclusion

The Smart Resume Parser demonstrates how NLP can automate resume screening, saving significant recruiter time. With enhancements like ML-based skill extraction and ATS integration, it can become a powerful HR analytics tool.

## 10. Future Scope

- 1 Improve accuracy using fine-tuned NER models
- 2 Add PDF layout understanding for complex designs
- 3 Build an API version for HR systems

## 11. References

- 1 spaCy Documentation - <https://spacy.io>
- 2 Streamlit Documentation - <https://docs.streamlit.io>
- 3 PyMuPDF Documentation - <https://pymupdf.readthedocs.io>