Resume Skill Extractor

Submitted by: Akash Kumar Saroj

BSc CSDA, IIT Patna

Abstract

This project automates the process of extracting relevant skills from resumes using Python and simple NLP techniques. It helps recruiters quickly identify candidate skills and missing job-required skills, reducing manual effort and errors.

Problem Statement

Recruiters must read multiple resumes to check if candidates match job requirements. Manual scanning is slow, error-prone, and time-consuming.

Objectives

- Automatically extract skills from resumes (PDF/DOCX).
- Match extracted skills with predefined job requirements.
- Highlight missing skills.
- Provide a simple output table with match percentage.

Methodology

1. Upload Resume file.
2. Extract text using parser (PyMuPDF / docx2txt).
3. Compare extracted skills with job skills.
4. Generate output table (Excel/CSV).
5. Provide match percentage and missing skills alerts.

Functional Components

- Resume parser module
- Skill keyword database
- Matching and extraction logic
- Output skill list in table
- Optional match percentage calculator

Sample Dataset & Expected Output

Sample Resume Text: 'Experienced in Python, SQL, and data visualization using Power BI.'

Job Skills: Python, SQL, Tableau

Expected Output:

- Extracted Skills: Python, SQL
- Missing Skills: Tableau
- Match: 67%

Tools & Technologies Used

- Python 3.x
- Libraries: PyMuPDF, docx2txt, pandas, openpyxl, re
- Excel/CSV for output tables

Future Scope

- Integrate NLP for better skill recognition.
- Web application interface for recruiters.
- Analyze multiple resumes simultaneously for bulk hiring.
- Extend to experience and education extraction.