Task: Intelligent Resume Parser

Name: Sethumadhavan V

Phone. No: 9159299878

Email: Sethumadhavanvelu2002@gmail.com

Problem Statement

Recruiters and hiring managers face challenges when screening large volumes of resumes. Manually extracting relevant information such as skills, experience, and qualifications is time-consuming, error-prone, and inconsistent.

Aim

To automate the extraction of structured metadata from resumes (PDF/DOCX), such as name, email, phone number, skills, experience, education, and certifications, using both traditional methods (regex and keyword search) and modern AI-powered models like Google Gemini.

Solution

Develop a web application using Flask that:

- Allows users to upload resumes in .pdf or .docx format.
- Extracts plain text content from resumes.
- Uses two methods to extract metadata:
  - **Primary**: Gemini AI (for intelligent, context-aware extraction).
  - **Fall-back**: Regex and keyword-based extraction (if Gemini fails).
- Returns structured information (like JSON) including skills, contact info, and qualifications.

Advantages

- **Automated Parsing**: Reduces manual effort and improves accuracy.
- **Fast & Scalable**: Processes multiple resumes quickly.
- **AI-Enhanced Extraction**: Leverages Gemini for context-aware information extraction.
- **Fall-back Strategy**: Ensures resilience using regex when AI fails or is unavailable.
- **Format Agnostic**: Supports both .pdf and .docx formats.
- **Web Interface**: Easy to use via a browser interface.

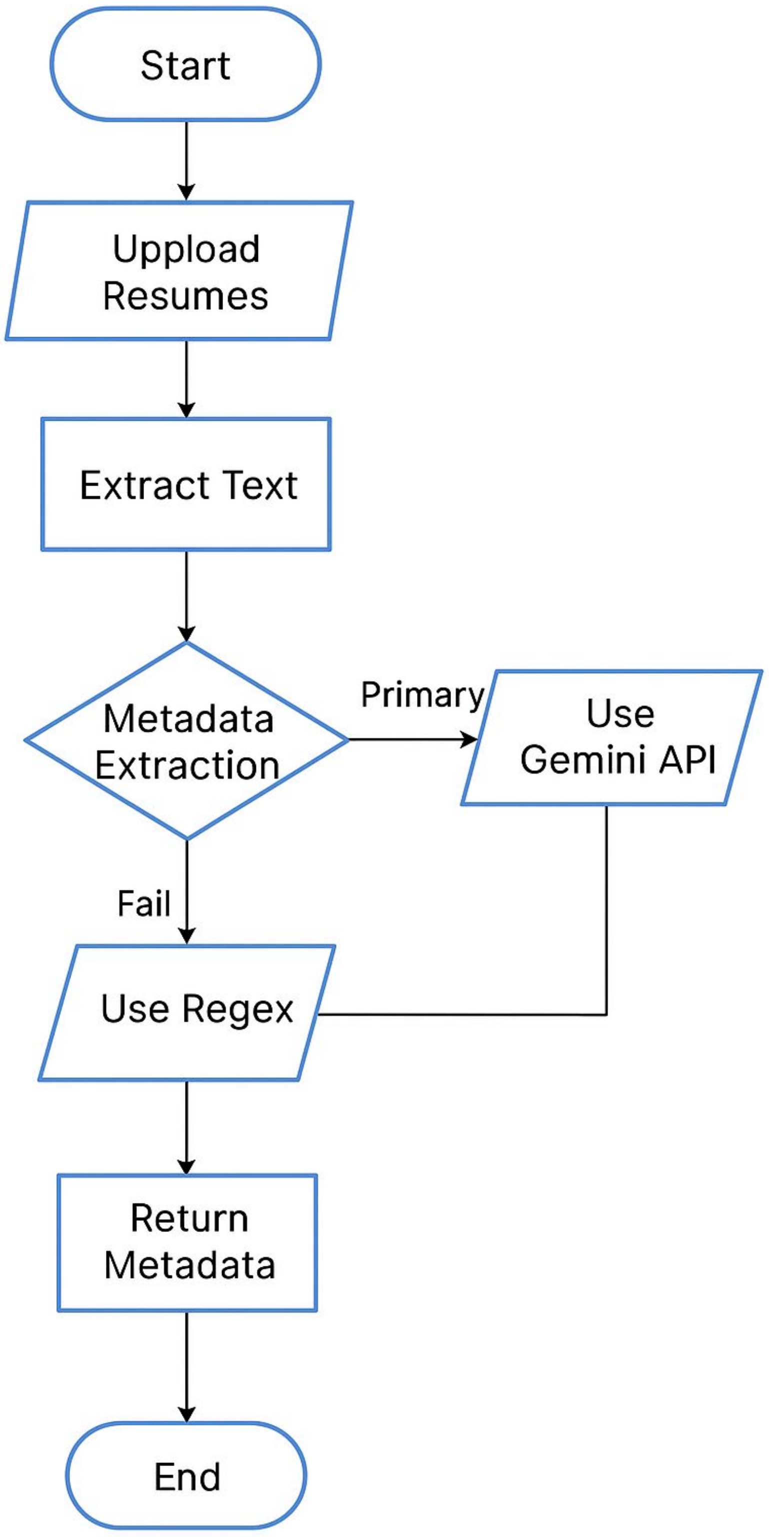
Disadvantages

- **Dependency on Gemini API**: Requires internet access and an API key.
- **Cost**: Using Gemini API at scale may incur charges.
- **Accuracy Limitations**: Regex methods might miss nuanced information or be inaccurate.
- **Privacy Concerns**: Resume data handling needs strict privacy/security practices.
- **Format Dependency**: Text extraction from poorly formatted PDFs may fail.

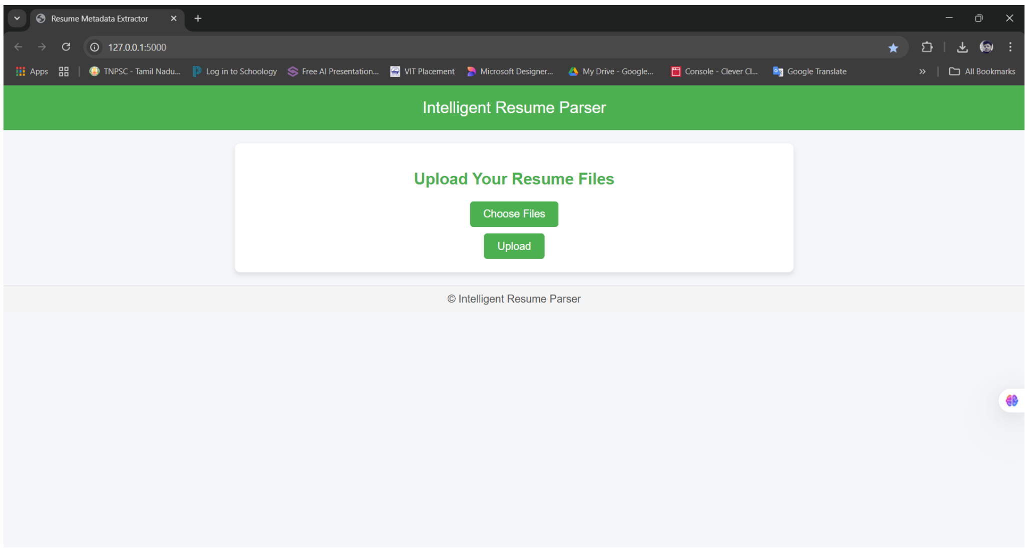
Approach

1. **Setup**:
   - Configure Flask backend and define upload directory.
   - Define allowed file types (.pdf, .docx).
2. **Upload Handling**:
   - Accept multiple file uploads from user.
   - Save securely using secure_filename().
3. **Text Extraction**:
   - For .pdf: Use PyPDF2 to extract text.
   - For .docx: Use python-docx.
4. **Metadata Extraction**:
   - **Primary**: Call Gemini API to extract metadata via prompt engineering.
   - **Fallback**: Use regex and keyword-based searches if Gemini fails.
5. **Chunking (optional)**:
   - Break large text into manageable pieces for better Gemini performance.
6. **Output**:
   - Return structured metadata as JSON (e.g., skills, contact info, education).
7. **Render Result**:
   - Load result on result_ai.html or return via API.

Flowchart:


Start
Uppload Resumes
Extract Text
Metadata Extraction
Primary
Use Gemini API
Fail
Use Regex
Return Metadata
End

Home Page:

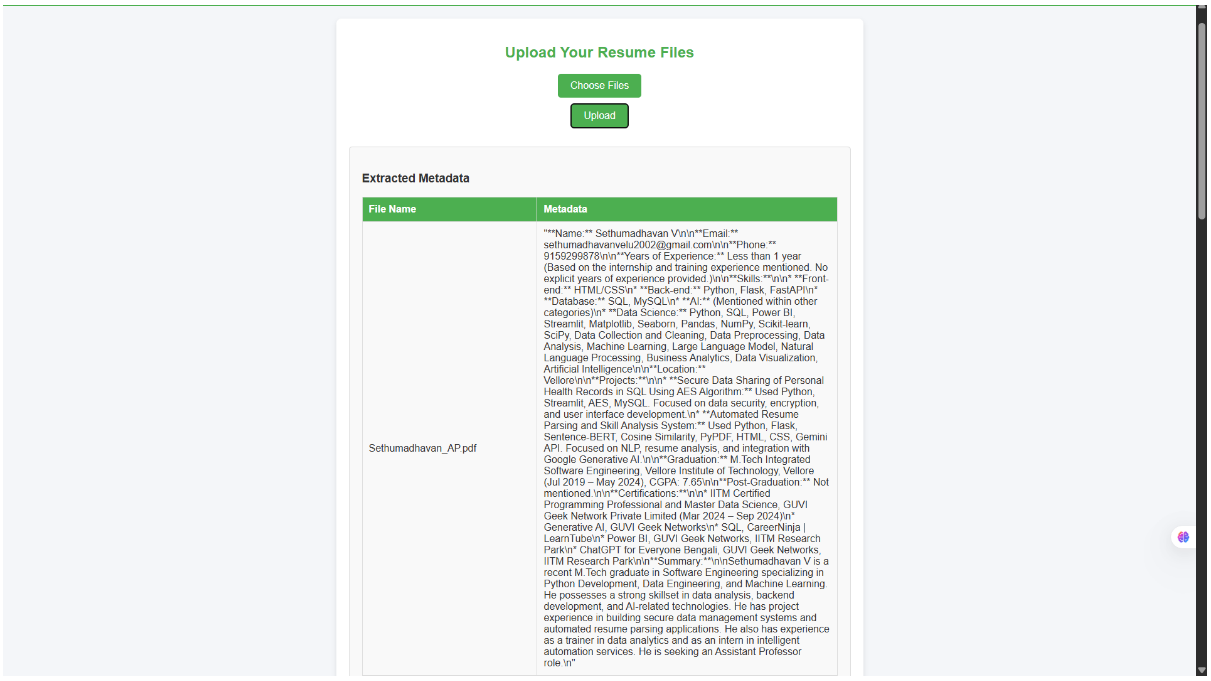


Intelligent Resume Parser

Upload Your Resume Files

© Intelligent Resume Parser

Result:



Upload Your Resume Files

Extracted Metadata

File Name	Metadata
Sethumadhavan_AP.pdf	"**Name:** Sethumadhavan V\n\n**Email:** sethumadhavanvelu2002@gmail.com\n\n**Phone:** 9159299878\n\n**Years of Experience:** Less than 1 year (Based on the internship and training experience mentioned. No explicit years of experience provided.)\n\n**Skills:**\n\n* **Front-end:** HTML/CSS\n* **Back-end:** Python, Flask, FastAPI\n* **Database:** SQL, MySQL\n* **AI:** (Mentioned within other categories)\n* **Data Science:** Python, SQL, Power BI, Streamlit, Matplotlib, Seaborn, Pandas, NumPy, Scikit-learn, SciPy, Data Collection and Cleaning, Data Preprocessing, Data Analysis, Machine Learning, Large Language Model, Natural Language Processing, Business Analytics, Data Visualization, Artificial Intelligence\n\n**Location:** Vellore\n\n**Projects:**\n\n* **Secure Data Sharing of Personal Health Records in SQL Using AES Algorithm:** Used Python, Streamlit, AES, MySQL. Focused on data security, encryption, and user interface development.\n* **Automated Resume Parsing and Skill Analysis System:** Used Python, Flask, Sentence-BERT, Cosine Similarity, PyPDF, HTML, CSS, Gemini API. Focused on NLP, resume analysis, and integration with Google Generative AI.\n\n**Graduation:** M.Tech Integrated Software Engineering, Vellore Institute of Technology, Vellore (Jul 2019 – May 2024), CGPA: 7.65\n\n**Post-Graduation:** Not mentioned.\n\n**Certifications:**\n\n* IITM Certified Programming Professional and Master Data Science, GUVI Geek Network Private Limited (Mar 2024 – Sep 2024)\n* Generative AI, GUVI Geek Networks\n* SQL, CareerNinja | LearnTube\n* Power BI, GUVI Geek Networks, IITM Research Park\n* ChatGPT for Everyone Bengali, GUVI Geek Networks, IITM Research Park\n\n**Summary:**\n\nSethumadhavan V is a recent M.Tech graduate in Software Engineering specializing in Python Development, Data Engineering, and Machine Learning. He possesses a strong skillset in data analysis, backend development, and AI-related technologies. He has project experience in building secure data management systems and automated resume parsing applications. He also has experience as a trainer in data analytics and as an intern in intelligent automation services. He is seeking an Assistant Professor role.\n"