Resume Parsing and Information Extraction Using Gemini AI

Imports and Setup

```python
import os, re
import google.generativeai as genai
import pandas as pd
from pdfminer.high_level import extract_text
from docx import Document
```

Explanation:
- Loads libraries for file handling, text extraction from PDFs/DOCX, AI interaction, and data management.
- Google Generative AI: Configures the AI model for processing resumes.

Text Preprocessing

```python
def preprocess_text(text):
    return re.sub(r'[^\x00-\x7F]+', ' ', re.sub(r'\s+', ' ', text))
```

Explanation:
- Cleans up the extracted text by removing unnecessary spaces making it easier for the AI model to work with.

Text Extraction Functions

```python
def extract_text_from_pdf(file_path):
    return extract_text(file_path)
def extract_text_from_docx(file_path):
    doc = Document(file_path)
    return '\n'.join([para.text for para in doc.paragraphs])
```

Explanation:
- These functions extract raw text from PDF and DOCX files using respective libraries.

Main Processing Loop

```python
directory = "Resumes/Sample"
all_extracted_details = []
genai.configure(api_key="YOUR_API_KEY")
model = genai.GenerativeModel(model_name='gemini-1.5-flash')
```

```python
for filename in os.listdir(directory):
    file_path = os.path.join(directory, filename)
    if filename.endswith('.pdf'):
        text = extract_text_from_pdf(file_path)
    elif filename.endswith('.docx'):
        text = extract_text_from_docx(file_path)
    else:
        continue
    processed_text = preprocess_text(text)
```

Explanation:
- Iterates over all files in the resume folder.
- Extracts text based on whether the file is PDF or DOCX, ignoring unsupported file types.

AI Prompt Creation and Response

```python
    prompt = """
    Extract Name, Email, Phone, DOB, College, City, Stream of Work,
Top Skills, Experience. If unavailable, use 'NA'.
    """
    response = model.generate_content([processed_text,
prompt]).text.strip()
    extracted_details = {
        "Name": "NA", "Email": "NA", "Phone": "NA", "DOB": "NA",
        "College": "NA", "City": "NA", "Stream of Work": "NA",
        "Top Skills": "NA", "Experience": "NA"
    }
    for line in response.splitlines():
        for key in extracted_details.keys():
            if key in line:
                extracted_details[key] =
line.split(":")[-1].strip().strip('"')
    all_extracted_details.append(extracted_details)
```

Explanation:
- Sends the cleaned resume text to the AI along with a prompt asking for specific details (name, email, etc.).
- Captures the AI's response and extracts relevant fields into a structured format (`extracted_details`).

Save Results to CSV

```python
df = pd.DataFrame(all_extracted_details)
df.to_csv("Resumes/Sample/extracted_resume_details.csv", index=False)
```

Explanation:
Converts the extracted resume data into a table format and saves it as a CSV for easy access.

Result:

```python
# Optionally, save the DataFrame to a CSV file
df.to_csv("Resumes/Sample/extracted_resume_details.csv", index=False)
```

```
            Name                           Email            Phone  \
0   Chitturi Prasad",         chprasad@kluniversity.in",     9700674515",
1    Jatin Varlyani",        varlyanijatin88@gmail.com",  +91 8806502484",
2       Junhao Dong",          junhao.dong96@gmail.com",   (678) 343-1817",
3      Kumar Saurav",  kumarsauravsmart2010@gmail.com",  +91-8420538839",
4   Rachelle Beaudry",        hello@reallygreatsite.com",   123-456-7890",
5    SAGAR BHANDARI",       sbhandari1@stcloudstate.edu",            NA",

      DOB                                        College          City  \
0    NA",   Working as Assistant Professor in KL Universit...            NA",
1    NA",          Veermata Jijabai Technological Institute",       Mumbai",
2    NA",                          Northeastern University",   Boston, MA",
3    NA",   Indian Institute of Engineering Science and Te...       Kolkata",
4    NA",   University of Finance and Management, City Col...      Any City",
5    NA",                        St. Cloud State University",  St Cloud,MN",

              Stream of Work  \
0          Assistant Professor",
1   Full Stack Web Developer",
2          Software Engineer",
3      Full Stack Developer",
4                Accounting",
5         Information Systems",

                                        Top Skills  \
0   Java, JSP, HTML, CSS, JavaScript, C, C++, SQL,...
1   PHP, HTML, CSS, Bootstrap, jQuery, SQL, JavaSc...
2   Python, Java, JavaScript, Ruby, C, Hack, Schem...
3   Core Java, Spring MVC, Oracle Database, LINUX,...
4   Financial Analysis, Budgeting and Forecasting,...
5   SQL, Java, HTML, CSS, Excel, Access, SAP HANA,...

                                        Experience
0                                               NA
1   PHP Developer Intern at Mehaar Technologies (M...
2   Facebook SOFTWARE ENGINEER  Menlo Park, CA Jul...
3   Associate Software Engineer at NRI FinTech Ind...
4   Accounting Executive, Borcelle Corporation\nJa...
5   Quality Analyst Intern at Century Bank Limited...
```

jupyter extracted_resume_details.csv Last Checkpoint: 44 minutes ago

File   Edit   View   Settings   Help

Delimiter: [, ▾]

	Name	Email	Phone	DOB	College	City	Stream of Work	Top Ski
1	Chitturi Prasad",	:hprasad@kluniversity.in",	9700674515",	NA",	m July 2017 to July 2018.	NA",	Assistant Professor",	L, Python, Programming
2	Jatin Varlyani",	lyanijatin88@gmail.com",	+91 8806502484",	NA",	ai Technological Institute",	Mumbai",	ıll Stack Web Developer",	Flask, React, Blockchair
3	Junhao Dong",	hao.dong96@gmail.com",	(678) 343-1817",	NA",	Northeastern University",	Boston, MA",	Software Engineer",	Docker, GraphQL, Flask
4	Kumar Saurav",	vsmart2010@gmail.com",	+91-8420538839",	NA",	and Technology, Shibpur",	Kolkata",	Full Stack Developer",	abase, LINUX, Windows
5	Rachelle Beaudry",	ello@reallygreatsite.com",	123-456-7890",	NA",	anagement, City College",	Any City",	Accounting",	, Financial Managemen
6	SAGAR BHANDARI",	ıdari1@stcloudstate.edu",	NA",	NA",	t. Cloud State University",	St Cloud,MN",	Information Systems",	ableau, COBOL, Pythor