Resume parsing Task

Introduction:

The task was to extract the most important information from the resumes and check if the resume qualified or not

As requested The problem is that we need to parse resumes to know if they are qualified or not, so we assumed that the input data are resumes in the format of pdf files or docs, mainly in an unstructured format

Data:

Resumes in format of pdf file or docs

Tools:

* Spacy → to extract entities pairs and relations
* Regex → to clean the data
* Nltk → To extract stopwords
* Doc2text → to parse docs file
* Pdfminer → to parse PDF files

Process:

First, we parse the pdf/docs files to get our raw data,

Then, we use our functions to extract the Name, Email, phone number, education, and Skills.

These functions mainly depend on spacy and Nltk to extract this information

I used a CSV (skills.csv) file that contains various skills to extract it from the resume.

Then we form a dictionary that contains the person's data.

I set the criteria for passing the screen as to have a bachelor/master's degree and to get have at least 70% of the skills required.

Test Cases:

The testing cases are two resumes containing different information and skills to get and passed them to our screen criteria that one of them passed and the other failed

Challenges:

The first challenge was utilizing spacy to extract entities and important information

Learned lessons:

Using spacy to extract various data from the text.

References:

<https://github.com/OmkarPathak/pyresparser>

<https://www.youtube.com/watch?v=HJy11kOlgvk&ab_channel=KGPTalkie>