***Name:*** Nada Belaidi.

***Email:*** [nadabelaidi98@gmail.com](mailto:nadabelaidi98@gmail.com)

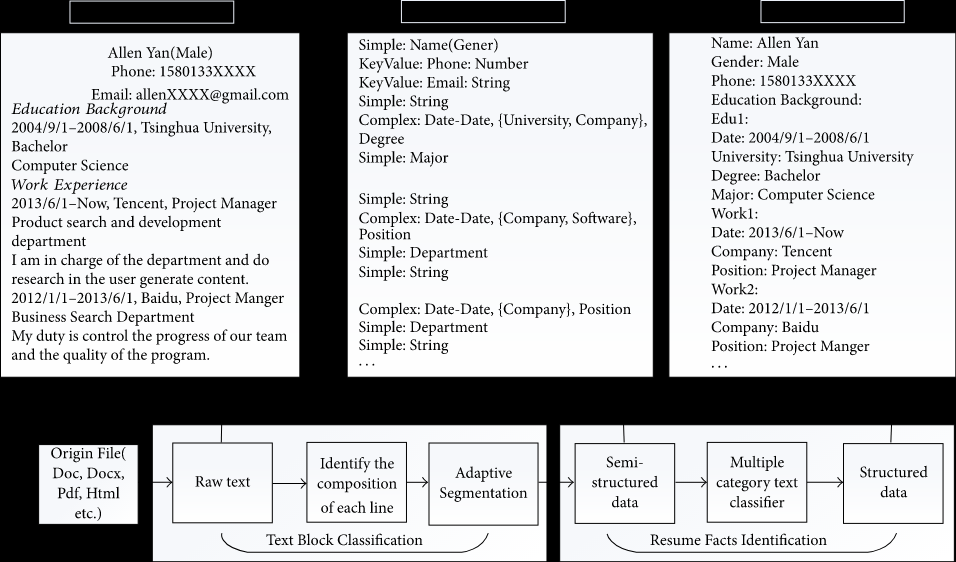
***Country***: Tunisia

***College:*** ESPRIT (The Private Higher School of Engineering and Technology)

***Specialization:*** NLP

***Problem description:***

Resumes contain surfeit information that is not relevant for the HR/authority, and they have to manually process the resumes to shortlist the promising candidates for them. And, thus making the shortlisting task a herculean task for HR. By making use of the NER (Named Entity Recognition) model of NLP this problem can be solved by finding and classifying the entities that are present in each resume into predefined classes such as person name, college name, academics information, relevant experiences, skill set, etc.



***Data understanding:***

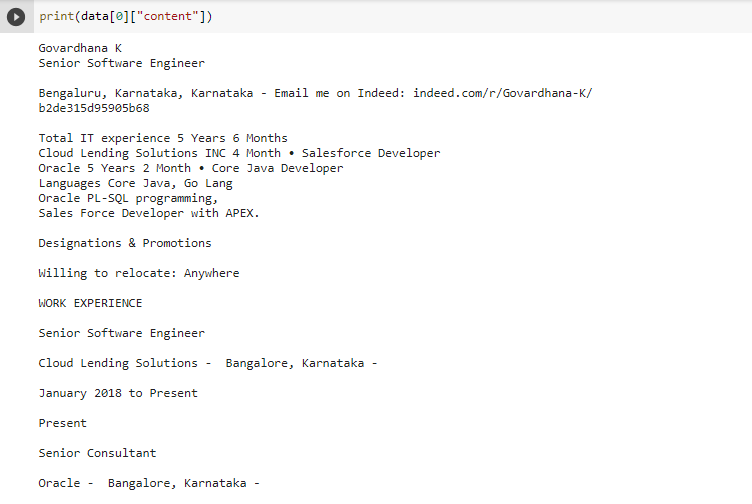
The data that has been provided is a json file.

This is what the data looks like:

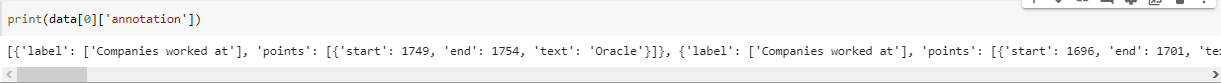


Its s composed of 200 resumes. Each line represents a 2 keys dictionary:

“content”: is the key to the plain resume text which looks like this:



“annotation” is the key to the labeled resume which looks like this:



Each resume feature is represented with a dictionary:

dict\_keys(['label', 'points'])

‘points’ is the key to a dictionary that looks like this:

[{'start': 1749, 'end': 1754, 'text': 'Oracle'}]

The data provided doesn’t seem to have a few problems:

-The text isn’t all in low case.

-Characters like “\n” and “\r” exists in the text.

- The dates seem odd too.

I intend to fix those problems using simple python commands, and Spacy.

(Because it has Built in visualizers for syntax and NER)

***GitHub Repo link:*** https://github.com/NadaBelaidi/NLP-Resume-Extraction