Name: Alexis Collier

Email: info@alexiscollier.com

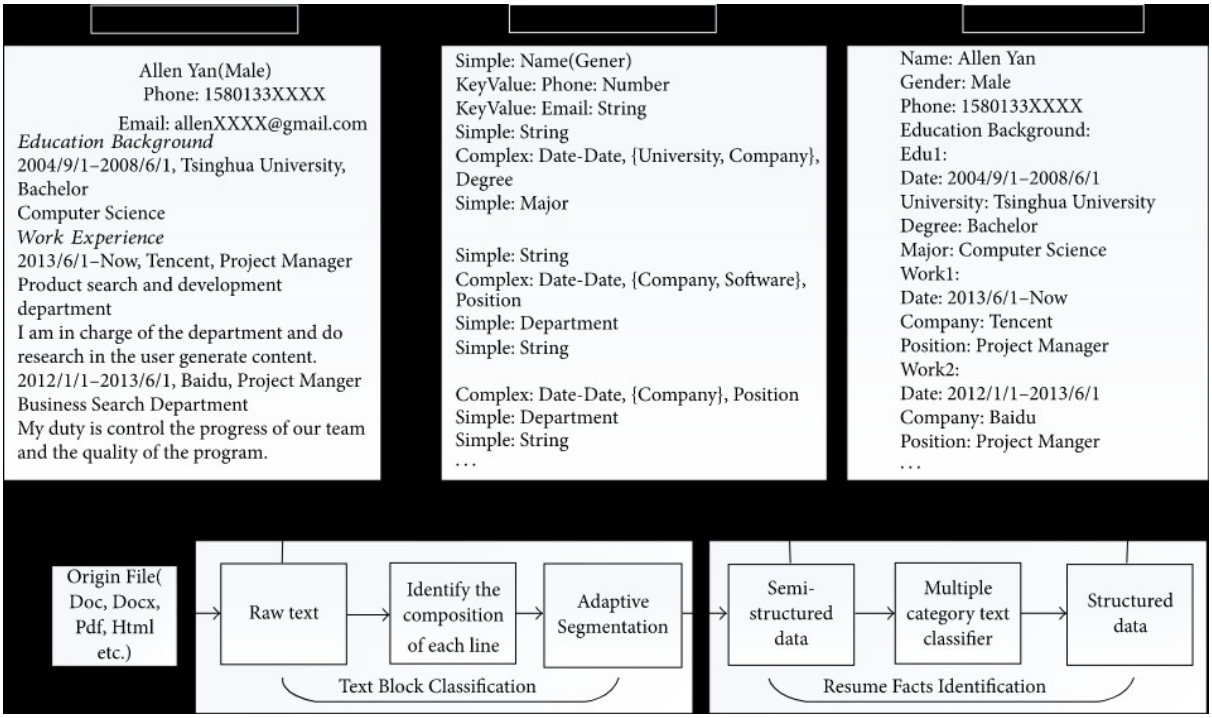
Country: United States

College: Fullstack Academy

Specialization: NLP

Problem description:

Resumes contain a surplus of information irrelevant to the HR/authority, and they have to manually process the resumes to shortlist the promising candidates, thus making the shortlisting task a herculean task for HR. Using the NER (Named Entity Recognition) model of NLP, this problem can be solved by finding and classifying the entities present in each resume into predefined classes such as person name, college name, academic 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:

```
{'content': 'Govardhana K\nSenior Software Engineer\n\nBengaluru, Karnataka, Karnataka - Email me on Indeed: indeed.com/r/Govardhana-K/\nb2
{'content': "Harini Komaravelli\nTest Analyst at Oracle, Hyderabad\n\nHyderabad, Telangana - Email me on Indeed: indeed.com/r/Harini-\nKoma
{'content': 'Hartej Kathuria\nData Analyst Intern - Oracle Retail\n\nBengaluru, Karnataka - Email me on Indeed: indeed.com/r/Hartej-Kathuri
{'content': "Ijas Nizamuddin\nAssociate Consultant - State Street\n\nIrinchayam B.O, Kerala - Email me on Indeed: indeed.com/r/Ijas-\nNizam
{'content': 'Imgeeyaul Ansari\njava developer\n\nPune, Maharashtra - Email me on Indeed: indeed.com/r/Imgeeyaul-Ansari/a7be1cc43a434ac4\n\n
{'content': 'Jay Madhavi\nNavi Mumbai, Maharashtra - Email me on Indeed: indeed.com/r/Jay-\nMadhavi/1e7d0305af766bf6\n\nI look forward to b
{'content': "Jitendra Babu\nFI/CO Consultant in Tech Mahindra - SAP FICO\n\nChennai, Tamil Nadu - Email me on Indeed: indeed.com/r/Jitendra
{'content': 'Jyotirbindu Patnaik\nAssociate consultant@SAP labs , Bangalore Karnataka\n\nBengaluru, Karnataka - Email me on Indeed: indeed.
{'content': 'Karthihayini C\nSystems Engineer - Infosys Limited\n\nRajapalaiyam, Tamil Nadu - Email me on Indeed: indeed.com/r/Karthihayini
{'content': 'Karthik GV\nArchitect - Microsoft India\n\nHyderabad, Telangana - Email me on Indeed: indeed.com/r/Karthik-GV/1961c4eff806e6f4
{'content': "Kartik Sharma\nSystems Engineer - Infosys Ltd\n\nDelhi, Delhi - Email me on Indeed: indeed.com/r/Kartik-Sharma/cc7951fd7809f35
{'content': "Kasturika Borah\nTeam Member - Cisco\n\nBengaluru, Karnataka - Email me on Indeed: indeed.com/r/Kasturika-\nBorah/9e71468914b3
{'content': "Kavitha K\nSenior System Engineer - Infosys Limited\n\nSalem, Tamil Nadu - Email me on Indeed: indeed.com/r/Kavitha-K/8977ce8c
{'content': 'Kavya U.\nNetwork Ops Associate - Accenture\n\nBengaluru, Karnataka - Email me on Indeed: indeed.com/r/Kavya-U/049577580b3814e
{'content': 'Khushboo Choudhary\nDeveloper\n\nNoida, Uttar Pradesh - Email me on Indeed: indeed.com/r/Khushboo-Choudhary/\nb10649068fcdfa42
{'content': 'kimaya sonawane\nThane, Maharashtra - Email me on Indeed: indeed.com/r/kimaya-\nsonawane/1f27a18d2e4b1948\n\nQuality education
{'content': "Koushik Katta\nDevops\n\nHyderabad, Telangana - Email me on Indeed: indeed.com/r/Koushik-Katta/a6b19244854199ec\n\nDevOps Admi
{'content': 'Kowsick Somasundaram\nCertified Network Associate Training Program\n\nErode, Tamil Nadu - Email me on Indeed: indeed.com/r/Kow
{'content': 'Lakshika Neelakshi\nSenior Systems Engineer - Infosys Limited\n\nBengaluru, Karnataka - Email me on Indeed: indeed.com/r/Laksh
{'content': "Madas Peddaiah\nAnantapur, Andhra Pradesh - Email me on Indeed: indeed.com/r/Madas-\nPeddaiah/557069069de72b14\n\n• Having 3 m
{'content': "Madhuri Sripathi\nBanglore, Karnataka, Karnataka - Email me on Indeed: indeed.com/r/Madhuri-\nSripathi/04a52a262175111c\n\nAro
{'content': "Mahesh Vijay\nBengaluru, Karnataka - Email me on Indeed: indeed.com/r/Mahesh-Vijay/a2584aabc9572c30\n\nOver 6.5 years of funct
{'content': "Manisha Bharti\nSoftware Automation Engineer\n\nPune, Maharashtra - Email me on Indeed: indeed.com/r/Manisha-Bharti/3573e36088
{'content': "Manjari Singh\nSenior Software Analyst - Accenture\n\nBengaluru, Karnataka - Email me on Indeed: indeed.com/r/Manjari-Singh/fd
{'content': 'Mohamed Ameen\nSystem engineer\n\nBengaluru, Karnataka - Email me on Indeed: indeed.com/r/Mohamed-Ameen/\nba052bfa70e4c0b7\n\n
```


It is composed of 200 resumes. Each line represents a two keys dictionary:

"content" is the key to the plain resume text, which looks like this:

```
print(data[0]["content"])
```

```
Govardhana K
Senior Software Engineer

Bengaluru, Karnataka, Karnataka - Email me on Indeed: indeed.com/r/Govardhana-K/
b2de315d95905b68

Total IT experience 5 Years 6 Months
Cloud Lending Solutions INC 4 Month • Salesforce Developer
Oracle 5 Years 2 Month • Core Java Developer
Languages Core Java, Go Lang
Oracle PL-SQL programming,
Sales Force Developer with APEX.

Designations & Promotions

Willing to relocate: Anywhere

WORK EXPERIENCE

Senior Software Engineer

Cloud Lending Solutions -  Bangalore, Karnataka -

January 2018 to Present

Present

Senior Consultant

Oracle -  Bangalore, Karnataka -
```

"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 does seem to have a few problems:

- There are unnecessary spaces or punctuation
- The text contains characters such as "\n" and "\r"

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/colla00/NLP-Resume-Extraction-Project