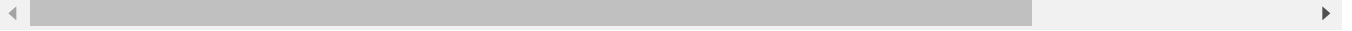```
pip install python-docx
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
Requirement already satisfied: python-docx in /usr/local/lib/python3.7/dist-packages (0
Requirement already satisfied: lxml>=2.3.2 in /usr/local/lib/python3.7/dist-packages (f
```

```
pip install docx
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
Requirement already satisfied: docx in /usr/local/lib/python3.7/dist-packages (0.2.4)
Requirement already satisfied: lxml in /usr/local/lib/python3.7/dist-packages (from doc
Requirement already satisfied: Pillow>=2.0 in /usr/local/lib/python3.7/dist-packages (f
```

```
pip install pyresparser
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
Requirement already satisfied: pyresparser in /usr/local/lib/python3.7/dist-packages (1
Requirement already satisfied: preshed>=2.0.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: urllib3>=1.25.3 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: spacy>=2.1.4 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: python-dateutil>=2.8.0 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: jsonschema>=3.0.1 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: six>=1.12.0 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: blis>=0.2.4 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: chardet>=3.0.4 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: attrs>=19.1.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: certifi>=2019.6.16 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: pytz>=2019.1 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: nltk>=3.4.3 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: sortedcontainers>=2.1.0 in /usr/local/lib/python3.7/dist
Requirement already satisfied: cymem>=2.0.2 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: srsly>=0.0.7 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: pycryptodome>=3.8.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: tqdm>=4.32.2 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: pandas>=0.24.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pyrsistent>=0.15.2 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: wasabi>=0.2.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: docx2txt>=0.7 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: thinc>=7.0.4 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: pdfminer.six>=20181108 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: numpy>=1.16.4 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: requests>=2.22.0 in /usr/local/lib/python3.7/dist-packag
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: importlib-resources>=1.4.0 in /usr/local/lib/python3.7/d
Requirement already satisfied: zipp>=3.1.0 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from nl
```

```
      Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.7/dist-package
      Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from n
      Requirement already satisfied: cryptography>=36.0.0 in /usr/local/lib/python3.7/dist-pa
      Requirement already satisfied: charset-normalizer>=2.0.0 in /usr/local/lib/python3.7/di
      Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.7/dist-packages (fr
      Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages (fro
      Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/di
      Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-pack
      Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (fr
      Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/dist
```

```
pip install spacy==2.3.7
```
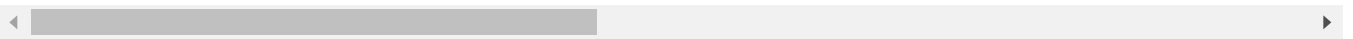
```
      Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
      Requirement already satisfied: spacy==2.3.7 in /usr/local/lib/python3.7/dist-packages (
      Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.7/dist
      Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-pack
      Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/di
      Requirement already satisfied: thinc<7.5.0,>=7.4.1 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-packages
      Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pa
      Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: blis<0.8.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pack
      Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (fr
      Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-p
      Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/dist
      Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (fro
      Requirement already satisfied: typing-extensions>=3.6.4 in /usr/local/lib/python3.7/dis
      Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packa
      Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (
      Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pack
      Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/li
```

```
!python -m spacy download en_core_web_sm
```

```
      Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
      Collecting en_core_web_sm==2.3.1
        Using cached en_core_web_sm-2.3.1-py3-none-any.whl
      Requirement already satisfied: spacy<2.4.0,>=2.3.0 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (fr
      Requirement already satisfied: thinc<7.5.0,>=7.4.1 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-pack
      Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/dist
      Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-p
      Requirement already satisfied: blis<0.8.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pack
      Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-pac
      Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/di
      Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.7/dist
```
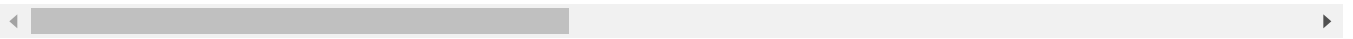
```
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: typing-extensions>=3.6.4 in /usr/local/lib/python3.7/dis
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/li
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (
✓ Download and installation successful
You can now load the model via spacy.load('en_core_web_sm')
```

```python
!python -m nltk.downloader words
```

```
/usr/lib/python3.7/runpy.py:125: RuntimeWarning: 'nltk.downloader' found in sys.modules
  warn(RuntimeWarning(msg))
[nltk_data] Downloading package words to /root/nltk_data...
[nltk_data]   Package words is already up-to-date!
```

```python
import nltk
nltk.download('stopwords')
from pyresparser import ResumeParser

import os
from docx import Document
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```python
##file format should be in .txt , .docx or .pdf only
filed=input()
```

```
/content/ResumeOfRafi1.docx
```

```python
filed
```

```
'/content/ResumeOfRafi1.docx'
```

```python
data = ResumeParser(filed).get_extracted_data()
```

```
/usr/local/lib/python3.7/dist-packages/spacy/util.py:275: UserWarning: [W031] Model 'en
  warnings.warn(warn_msg)
```

```python
try:
    doc = Document()

    with open(filed, 'r') as file:
        doc.add_paragraph(file.read())
    doc.save("text.docx")
    data = ResumeParser('text.docx').get_extracted_data()
    print("skilss",data['skills'])
except:
    data = ResumeParser(filed).get_extracted_data()
    print("Candidate skills",data['skills'])
    print("Email",data['email'])
    print("Name",data['name'])
    print("total_experience",data['total_experience'])
    print("company_names",data['company_names'])
    print("degree",data['degree'])
    print("designation",data['designation'])
    print("Candidate skills",type(data['skills']))
```

```
Candidate skills ['Mysql', 'Database', 'C++', 'Email', 'Engineering', 'Python', 'Progra
Email kazimusaddirafi@gmail.com
Name Kazi Musaddi Rafi
total_experience 0
company_names None
degree None
designation ['Software Engineer']
Candidate skills <class 'list'>
```

```python
import pandas as pd
```

```python
print(type (data))
```

```
<class 'dict'>
```
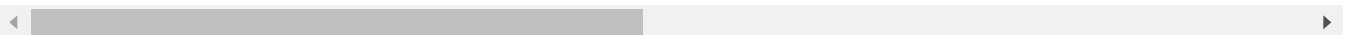
```python
df = pd.DataFrame(data['skills'])
print(type(df))
```

```
<class 'pandas.core.frame.DataFrame'>
```

```python
# Another approach
```

```
pip install resume-parser
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/pub
Requirement already satisfied: resume-parser in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: spacy>=2.3.2 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: docx2txt>=0.8 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: phonenumbers>=8.12.7 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: pdfminer.six>=20200517 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: tika>=1.24 in /usr/local/lib/python3.7/dist-packages (fr
Requirement already satisfied: pandas>=1.1.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: nltk>=3.5 in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: numpy>=1.19.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pdfplumber>=0.5.23 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: stemming>=1.0.1 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: click in /usr/local/lib/python3.7/dist-packages (from nl
Requirement already satisfied: tqdm in /usr/local/lib/python3.7/dist-packages (from nlt
Requirement already satisfied: joblib in /usr/local/lib/python3.7/dist-packages (from n
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: charset-normalizer>=2.0.0 in /usr/local/lib/python3.7/di
Requirement already satisfied: cryptography>=36.0.0 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.7/dist-packages (fr
Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: Pillow>=9.1 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: Wand>=0.6.10 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/dist
Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (fr
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.7/dist
Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: blis<0.8.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/di
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: thinc<7.5.0,>=7.4.1 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (fro
Requirement already satisfied: typing-extensions>=3.6.4 in /usr/local/lib/python3.7/dis
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packa
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/li
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pack
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (
```

```python
from resume_parser import resumeparse # Library 1
from pyresparser import ResumeParser #Library 2
import os
import spacy
spacy.load('en_core_web_sm')
```

```
<spacy.lang.en.English at 0x7fcc6deacc90>
```

```python
dir_list = os.listdir(r'/content/drive/MyDrive/CV Sorting Project/Resumes')
#print(dir_list)

res_list = []
paths = r'/content/drive/MyDrive/CV Sorting Project/Resumes'
c = 0
for i in dir_list:
    c = c + 1
    if c == 5000:
        break
    pfinal = os.path.join(paths, i)
    data = ResumeParser(pfinal).get_extracted_data()
    res_list.append(data)


print(len(res_list))
```

```
3
```

```python
import pandas as pd
df= pd.DataFrame(res_list)
```

```python
df
```

	name	email	mobile_number	skills	college_name	
0	Kazi Musaddi Rafi	kazimusaddirafi@gmail.com	8801626627	[Mysql, Database, C++, Email, Engineering, Pyt...	None	Inforr Com
1	Mostafa Mohiuddin	mostafamohiuddin.j@gmail.com	8801824493	[Algorithms, Operations, Robot, Video, Editing...	None	Inforr Com
				[Healthcare		

Colab paid products - Cancel contracts here

✓ 0s   completed at 9:50 PM   ● ✕