



#### config.json

```
{"chromedriver_path": "Absolute Path"}
```

#### Assumptions

To handle different page structure, i have implemented page specific scraping method. And wrapped them under a class to create one system.

However Finding Careers page same for every site

Output will be stored in a json file

#### Limitation

This system has great dependencies on selenium with chromedriver. So, we have to make sure that we are always using compatible chromedriver with respect to host chrome.

This is not an universal system. This has been designed for specifically handle given URL's in assignment

We have to update the scraping method if page structure gets changed in future

#### Required Python Module

```

async-generator==1.10
attrs==21.4.0
beautifulsoup4==4.11.1
certifi==2022.6.15
cffi==1.15.1
charset-normalizer==2.1.0
cryptography==37.0.4
h11==0.13.0
idna==3.3
outcome==1.2.0
pycparser==2.21
pyOpenSSL==22.0.0
PySocks==1.7.1
requests==2.28.1
selenium==4.3.0
sniffio==1.2.0
sortedcontainers==2.4.0
soupsieve==2.3.2.post1
trio==0.21.0
trio-websocket==0.9.2
urllib3==1.26.10
wsproto==1.1.0
  
```

requests, bs4, selenium needs to be installed in environment by using python package manager. Which will automatically download dependent module

#### Output.json

```
{
  "https://www.druva.com": ["JOB URLS LIST"],
  "https://www.talentica.com": ["JOB URLS LIST"],
  "https://www.nuance.com": ["JOB URLS LIST"]
}
```

```
(venv) PS C:\Users\pulak\PycharmProjects\Alpha-Sense> python --version
Python 3.9.7
```