`Scraping Indeed`

In [992]:

```python
from bs4 import BeautifulSoup
import urllib.request
import csv
from lxml import html
import requests
from selenium.webdriver.support.ui import WebDriverWait
import pandas as pd
```

In [993]:

```python
df_temp=pd.read_csv('ideed_Eng.csv')

urlpage = 'https://www.indeed.com/viewjob?jk=c49c843df963e561&tk=1e9nvm46143h4801&from=serp&vjs=3'
page = urllib.request.urlopen(urlpage)
soup = BeautifulSoup(page, 'html.parser')
entreprise = soup.find_all('div', class_='icl-u-lg-mr--sm icl-u-xs-mr--xs')[0].get_text()
title=soup.find_all('h3', class_='icl-u-xs-mb--xs icl-u-xs-mt--none jobsearch-JobInfoHeader-title')[0].get_text()
description =soup.find_all('div', class_='jobsearch-jobDescriptionText')[0].get_text()
description = description.replace(u'\xa0', u' ')
description = description.replace(u'\n', u' ')

df = pd.DataFrame({
 'entreprise':[entreprise],
'titre':[title],
'description':[description]
})


df = pd.concat([df, df_temp])
#Delet index
df.reset_index(drop=True, inplace=True)
```

In [994]:

```
df
```

Out[994]:

	entreprise	titre	description
0	The Standard	Senior Data Scientist	At The Standard, you'll join a team focused on...
1	dunnhumby	Senior Applied Data Scientist	dunnhumby is the world's leading customer scie...
2	HEPCO, Inc.	Senior Data Scientist	SENIOR DATA SCIENTIST HEPCO, Inc., is seeking...
3	One Concern	Data Scientist, Machine Learning	One Concern is a startup which specializes in ...
4	Grammarly, Inc.	Data Scientist	The opportunity Grammarly empowers people to ...
...	...	...	...
466	GETRESOURCE	Jr./Entry Level Middleware consultants	Job Requirements This requirement is for a ca...
467	Clark, Richardson and Biskup Consulting Engine...	Entry Level Process Simulation Engineer	Position Summary The Entry-Level Process Simul...
468	Logos Infotech Inc	Entry level Devops Engineer	Company Description Our company is a leading ...
469	SDSN Association	SDSN USA Analyst Intern	Organization Background The SDSN USA endeavor...
470	DataLab USA	Data Scientist / Statistician	Key responsibilities/duties include: Build, im...

471 rows × 3 columns

In [995]:

```
df.to_csv('ideed_Eng.csv',index=False)
```

In [ ]: