`Scraping TanitJobs `

In [2]:

```
from selenium import webdriver
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.by import By
from selenium.common.exceptions import TimeoutException
from selenium.webdriver.common.keys import Keys
import time
from selenium.webdriver.common.action_chains import ActionChains
from selenium.webdriver.common.keys import Keys

# bfs4
from bs4 import BeautifulSoup
import urllib.request
import requests

import pandas as pd
```

Use selenium to retrieve job offer urls

In [ ]:

```
options =webdriver.ChromeOptions()
options.headless=False
prefs={"profile.default_content_setting_values.notifications" :2}
options.add_experimental_option("prefs",prefs)
driver = webdriver.Chrome('C:/chromedriver/chromedriver.exe')

driver.get("https://www.tanitjobs.com/")
driver.find_element_by_xpath('//*[@id="keywords"]').send_keys("informatique",Keys.ENTER
)
time.sleep(1)

num_links = len(driver.find_elements_by_class_name('pad_right_small'))
num_links

a=0
for i in range(num_links):
    driver.find_element_by_xpath("/html/body/div[1]/div/div[2]/div[3]/div[4]/button").c
lick()
    a+=1

list_url=[]
for i in range(len(driver.find_elements_by_class_name("link"))):
    button = driver.find_elements_by_class_name("link")[i]
    button.click()
    print(driver.current_url)
    list_url.append(driver.current_url)
    driver.execute_script("window.history.go(-1)")
```

In [ ]:

```
# On supprimer l url de la page intiale de recherche
list_url.remove(list_url[0])
```

Scraping avec beautifulSoup

In [ ]:

```
a=0
for i in range(len(list_url)):
    urlpage = list_url[i]
    page = urllib.request.urlopen(urlpage)
    soup = BeautifulSoup(page, 'html.parser')

    description =soup.find_all('div', class_='details-body__content content-text')[0].g
et_text()
    description = description.replace(u'\xa0', u' ')
    description = description.replace(u'\n', u' ')

    uls = soup.select("div.bootstrap-tagsinput")
    text = [a.text for ul in uls for a in  ul.select("a")]
#-------------------------------------------------------------------------------

    df = pd.DataFrame({'description':[description],'Tags':[text]})
    df = pd.concat([df, df_temp])
    #Delet index
    df.reset_index(drop=True, inplace=True)

    df.to_csv('Tanitjobs.csv',index=False)
    a+=1
```

In [4]:

```
df.head()
```

Out[4]:

	Unnamed: 0	text	Tags
0	0	Skills and Qualifications Bachelor's degree i...	['informatique ', "Technologie de l'informatio...
1	1	Profile: You have completed your technical stu...	['Développeur ', 'Informatique ', 'Web ', '...
2	2	Technical Skills: SQL Server / SQL, basic Sync...	['Responsable Applicatif ', 'Informatique ',...
3	3	In our R & D team, you will participate in the...	['Ingénieur ', 'Informatique ', 'Développeme...
4	4	Job requirements you serious, motivated, punct...	['Formateur ', 'Développement ', 'Informatiq...

In [ ]:

```

```