

```

# -*- coding: utf-8 -*-
"""Projet3_web_scraping.ipynb

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/1suvSUA9lRbBzQ3ax6KwIuV6dxDGjBgka

### ***Importer les librairies nécessaires***
"""

from urllib.request import Request, urlopen
import pandas as pd
from bs4 import BeautifulSoup

"""### ***Définir les liens à scraper***
Le 1er lien est pour les films et le second est pour les acteurs et animateurs
"""

url = 'https://www.themoviedb.org/movie'
url_actors = 'https://www.themoviedb.org/person'

"""##Ici, le code d'etat est de 403: Cela nous empeche de scraper la page...
###Pour parler ca, nous avons notre Agent-Utilisateur de notre navigateur.
##En suite, nous utiliser la fonction ***BeautifulSoup***
"""

UserAgent = Request(url, headers={'User-agent': 'Chrome/95.0.4638.69'})
html = urlopen(UserAgent)
data = BeautifulSoup(html, 'html.parser')

UserAgent = Request(url_actors, headers={'User-agent': 'Chrome/95.0.4638.69'})
html1 = urlopen(UserAgent)
data1 = BeautifulSoup(html1, 'html.parser')

"""## Extraction du ***titre*** de la page web"""

title = data.title.text
print(title)

"""###Après analyse du code HTML, accedons a la balise conteneur tous les
films***"""

movie_containers = data.find_all('div', class_ = 'card style_1')
print(type(movie_containers))
print(len(movie_containers))

"""##Nous pouvons accéder au premier conteneur, qui contient des informations
sur un seul film"""

first_movie = movie_containers[0]
first_movie

"""###Quelques informations sur le premier flim

```

```

"""

#Information du film
first_movie.div

#l'hypertext du film
first_movie.a

#Le nom du film
first_name = first_movie.h2.a.text
first_name

"""###Après analyse du code HTML, accedons a la balise conteneur tous les
acteurs###"""

actors_containers = data1.find_all('div', class_ = 'fifty_square')
print(type(actors_containers))
print(len(actors_containers))

"""###Nous allons définir une fonction qui va envoyer la listes des films, la
liste des acteurs et animateurs.###

##Fonction pour afficher les films
"""

#Definir une fonction get_all_movies qui va contenir l'ensemble des films

def get_all_movies():
    print("\n-----")
    print("\nMovies : ")
    print("-----\n")

    containers = data.find_all("div", class_="card style_1")
    for movies in containers:
        titre_1 = movies.find("div", class_="content")
        titre_2 = titre_1.find("h2")
        titre = titre_2.find("a")
        lien = titre_2.find("a", href=True)
        date_sortie = titre_1.find("p")
        print("\n-----")
        print("Titre : ", titre.text)
        print("Date : ", date_sortie.text)
        print("-----\n")
        #print(movies)

"""###Fonction pour afficher les acteurs ou animateurs."""

def get_liste_artors_or_animators():
    print("\n-----")
    print("\tArtistes : ")
    print("-----\n")

    listes_acteurs = data1.find_all("div", class_="fifty_square")

```

```

for acteurs_data in listes_acteurs:
    parent_data = acteurs_data.find("div", class_="meta")
    nom_artiste = parent_data.find("p", class_="name")
    film_jouer = parent_data.find("p", class_="sub")
    print("\n-----")
    print("Nom : ", nom_artiste.text)
    print("Film : ", film_jouer.text)
    print("-----\n")

"""#Appelle des deux fonctions pré-définies"""

if __name__ == '__main__':
    print("-----\n")
    print("\tMENU\t\n")
    print("1- Liste des films\n")
    print("2- Liste des artistes\n")
    print("Tapez une autre touche pour quitter\n")
    print("-----\n")

    x = input("Tapez quelque chose ici : ")
    if x == "1":
        get_all_movies()
    elif x == "2":
        get_liste_artors_or_animators()
    else:
        print("AU REVOIR")

```