

Application 1

Kossi

11/01/2020

Extraction des principaux taux de change sur MSN

Dans cette partie, nous allons mettre en pratique tout ce que nous avons déjà appris dans ce cours. En mobilisant les différents points déjà abordés, nous irons récupérer des données financières en temps réel sur MSN.

MSN est un portail de services web détenu par Microsoft.

Ouvrons la page des taux de change MSN avec la librairie **requests** :

```
import requests
from bs4 import BeautifulSoup
msn = requests.get("https://www.msn.com/en-us/money/markets/currencies")
print( "code d'état :", msn.status_code)
```

```
## code d'état : 200
```

```
soup = BeautifulSoup(msn.text)
```

Q1 Quelle est la plus petite classe qui englobe notre table d'intérêt ? Appelons cette classe `currency_class`. Modifiez le code ci-dessous pour ne sélectionner que la classe `currency_class`.

```
# N'oubliez pas de définir currency_class
cur = soup.find( class_ = currency_class) # Retenez cette syntaxe, elle permet d'extraire une classe q
if cur and len(cur) :
    print("Bravo: il semblerait que j'ai trouvé la classe!")
else:
    print("Je n'ai toujours pas trouvé la classe mère")
```

```
## Bravo: il semblerait que j'ai trouvé la classe!
```

Pour trouver la classe `'currency_class'`, vous pouvez vous servir de votre navigateur Web en faisant 'C

Q2 Quelle est la classe qui représente l'en-tête de la table ? Notons-la `thead_class`. Alors l'en-tête s'extrait de la façon suivante :

```
# N'oubliez pas de définir thead_class
header = cur.find(class_ = thead_class)
headers = header.find_all(class_ = "mctblheading") # Avez-vous pu retrouver la classe 'mctblheading' d
header_values = []
for header in headers :
    header_values.append(header.find("p").attrs["title"])
print(header_values)
```

```
## ['Major Currencies', 'Price', 'Change', 'Change%', '52 Week High', '52 Week Low']
```

Q3 Quelle est la classe qui englobe toutes les lignes de la table ? Elle sera notée `trows_class`. Pouvez-vous identifier la classe qui caractérise une ligne dans la table ? Notons cette classe `'trow_class'`.

```
rows_container = cur.find(class_ = trows_class)
rows = rows_container.find_all(class_ = "mcrow")
row = rows[0]
print(row)
```

```
## <div class="mcrow">
## <span aria-hidden="True" class="flagcol"><span class="cntryflag" id="EUR"> </span></span>
## <div class="cntrycnvrsn">
## <span class="cntrycol"><p class="truncated-string" title="Euro">Euro</p></span>
## <span class="cnvrsncol"><p class="truncated-string" title="USD/EUR">USD/EUR</p></span>
## </div>
## <span class="pricecol"><p aria-label="Price 0.8991" class="truncated-string" title="0.8991">0.8991</p>
## <div class="chcpclub">
## <span class="chnghcol chng-value decrease"><p aria-label="Change -0.0003" class="truncated-string" title="Change -0.0003%">-0.0003%</p>
## <span class="chpcol chp-value decrease"><p aria-label="Change% -0.0334%" class="truncated-string" title="Change% -0.0334%">-0.0334%</p>
## </div>
## <span class="rng52wkcol"><p aria-label="52 Week High 0.9191" class="truncated-string" title="0.9191">0.9191</p>
## <span class="rng52wkcol"><p aria-label="52 Week Low 0.8642" class="truncated-string" title="0.8642">0.8642</p>
## </div>
```

Q4 En vous servant de la variable **row** définie dans la question précédente, trouvez les classes qui représentent :

- La monnaie de base
- La paire de monnaies de change
- Le cours/prix de change
- La variation absolue du cours
- La variation en pourcentage du cours

Extraire chacun des champs sus-mentionnés.

Quelle est la classe commune au maximum et minimum annuel des cours ? Pouvez-vous dire l'**attribut HTML** qui différencie le cours minimum du cours maximum ? Extraire les 2 cours. Extrayez les 2 cours.

Q5 En vous servant de la question **Q4**, écrire une fonction qui prend en entrée une ligne de la table et extrait ses champs. Utilisez cette fonction pour extraire toutes les lignes de la table dans une liste de dictionnaire.

Q6 En utilisant la librairie **pandas**, convertissez la liste de la question **5** en un **dataframe** (n'hésitez pas à vous inspirer de l'en-tête de la table pour définir le nom de vos colonnes).

Vous devriez obtenir une table de la forme :

```
##          currency    pairs  ... rng52wkcol_high rng52wkcol_low
## 0             Euro  USD/EUR  ...          0.9191          0.8642
## 1    Japanese Yen  USD/JPY  ...        112.3900        104.4400
## 2    British Pound  GBP/USD  ...          1.3514          1.1957
## 3   Canadian Dollar  USD/CAD  ...          1.3564          1.2947
## 4   Australian Dollar  USD/AUD  ...          1.4988          1.3702
##
## [5 rows x 7 columns]
```

Q7 Au vu de nos données, quel est le taux de conversion le plus élevé ? Pouvez-vous visualiser ces taux à l'aide d'un diagramme en bâtons (barplot) ?

Q8 Reconstruire la table pandas de la question **7** en n'utilisant cette fois que des **sélecteurs CSS**.

BRAVO, vous avez réussi à scraper votre premier page Web !.