

Group_6_Scraping_Images

April 16, 2021

1 Group 6 - Image Scraping (Google Arts & Culture)

```
[1]: import requests
```

```
[3]: res = requests.get("https://artsandculture.google.com/partner/albertina-vienna")
```

```
[4]: res.raise_for_status()
```

```
[5]: content = res.content
```

```
[6]: content[:50]
```

```
[6]: b'<!DOCTYPE html><html lang="en"><head><meta charset'
```

```
[7]: !pip install beautifulsoup4
```

```
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.9/site-packages (4.9.3)
```

```
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.9/site-packages (from beautifulsoup4) (2.1)
```

```
[8]: !pip freeze
```

```
beautifulsoup4==4.9.3
blessed==1.17.6
certifi==2020.12.5
cffi==1.14.4
chardet==4.0.0
click==7.1.2
cryptography==3.3.1
flake8==3.8.4
Flask==1.1.2
Flask-Cors==3.0.10
idna==2.10
inquirer==2.7.0
itsdangerous==1.1.0
Jinja2==2.11.3
MarkupSafe==1.1.1
```

```
mccabe==0.6.1
numpy==1.20.1
pandas==1.2.3
psycpg2-binary==2.8.6
pycodestyle==2.6.0
pycparser==2.20
pyflakes==2.2.0
pyOpenSSL==20.0.1
python-dateutil==2.8.1
python-editor==1.0.4
pythonping==1.0.16
pytz==2021.1
readchar==2.0.1
requests==2.25.1
requests-toolbelt==0.9.1
simple-term-menu==1.0.0
six==1.15.0
soupsieve==2.1
urllib3==1.26.2
wcwidth==0.2.5
Werkzeug==1.0.1
```

```
[15]: from bs4 import BeautifulSoup
```

```

↳ -----
ModuleNotFoundError                                Traceback (most recent call↳
↳ last)

<ipython-input-15-f33cf0cb8fd3> in <module>
----> 1 from bs4 import BeautifulSoup

ModuleNotFoundError: No module named 'bs4'
```

```
[17]: !pip3.8 install beautifulsoup4
```

```
Collecting beautifulsoup4
  Using cached beautifulsoup4-4.9.3-py3-none-any.whl (115 kB)
Collecting soupsieve>1.2
  Downloading soupsieve-2.2.1-py3-none-any.whl (33 kB)
Installing collected packages: soupsieve, beautifulsoup4
Successfully installed beautifulsoup4-4.9.3 soupsieve-2.2.1
```

```

[18]: from bs4 import BeautifulSoup

[19]: soup = BeautifulSoup(content)

[22]: image_hrefs = soup.find(id="exp_tab_popular").find_all("a")

[33]: image_hrefs[0].get("href")

[33]: '/asset/hase-albrecht-dürer/1QHEnzUGYMDG_w'

[34]: images_links = [anchor.get("href") for anchor in image_hrefs]

[35]: images_links

[35]: ['/asset/hase-albrecht-dürer/1QHEnzUGYMDG_w',
      '/asset/big-fish-eat-little-fish-1556/XQE024hSX0rhWQ',
      '/asset/the-large-piece-of-turf-1503/NgELdACk3I8Jkg',
      '/asset/praying-hands-1508/gwGj6BUX8D5Kug',
      '/asset/seated-couple-1915/6gGg3g6t0zmIOQ',
      '/asset/female-lovers-1915/OgEx9nCb-BQmUA',
      '/asset/standing-nude-girl/9wGwcjnXy_POyw',
      '/asset/the-tree-man-c-1505/zwHAFav7ln6uEA',
      '/asset/venice-the-pink-cloud-1909/rAGxIWLe1DFcxQ',
      '/asset/loïe-fuller-in-the-dance-the-archangel-1902/zQHoKKwoqqGSNw',
      '/asset/half-length-figure-of-an-apostle-1493-1495/bAGjZ7zpYVEKyQ',
      '/asset/the-painter-and-the-buyer-1565/eQHTzz_SGMMZJQ',
      '/asset/seated-young-male-nude-and-two-arm-studies-recto-c-1510-11/JgGU41RjZpT-Q',
      '/asset/mary-among-a-multitude-of-animals-c-1503/AAETMvKqvXDNXg',
      '/asset/old-houses-in-krumau-1914/1AFT3iyW1Tp2eQ']

[36]: URL_PREFIX = "https://artsandculture.google.com"

[47]: for image_link in images_links:
      res = requests.get(f"{URL_PREFIX}{image_link}")
      res.raise_for_status()

      image_soup = BeautifulSoup(res.content)
      image_id = image_link.split("/")[-1]
      metadata = image_soup.find(id=f"metadata-{image_id}").find("ul").
      ↪find_all("li")

      image_data = {}

      for item in metadata:
          image_data_record = item.get_text()
          key, val = image_data_record.split(": ")

```

```

        image_data[key] = val
    print(image_data)

    print()

```

```

{'Title': 'Hase', 'Creator': 'Albrecht Dürer', 'Date': '1502', 'Physical
Dimensions': '25 x 22,5 cm', 'Type': 'Drawing', 'Rights': 'Albertina', 'External
Link': ' Albertina Collections Online', 'Medium': 'Watercolor and body color,
heightened with white body color'}

```

```

{'Title': 'Big Fish Eat Little Fish, 1556', 'Date Created': '1556', 'Painter':
'Pieter Bruegel the Elder', 'Type': 'Brush and pen and gray and black ink,
transfer lines', 'Rights': 'Albertina, Vienna'}

```

```

{'Title': 'The Large Piece of Turf, 1503', 'Date Created': '1503', 'Painter':
'Albrecht Dürer', 'Type': 'Watercolor and body color, heightened with white body
color', 'Rights': 'Albertina, Vienna'}

```

```

{'Title': 'Praying Hands, 1508', 'Date Created': '1508', 'Painter': 'Albrecht
Dürer', 'Type': 'Brush, gray and white ink, gray wash, on blue prepared paper',
'Rights': 'Albertina, Vienna'}

```

```

{'Title': 'Seated Couple, 1915', 'Date Created': '1915', 'Painter': 'Egon
Schiele', 'Type': 'Brush, watercolor, pencil, die stamp, mounted on cardboard',
'Rights': 'Albertina, Vienna'}

```

```

{'Title': 'Female Lovers, 1915', 'Date Created': '1915', 'Painter': 'Egon
Schiele', 'Type': 'Body color, pencil', 'Rights': 'Albertina, Vienna'}

```

```

{'Title': 'Standing Nude Girl', 'Date Created': 'Undated', 'Painter': 'Francois
Boucher', 'Type': 'Black, white, and red chalk, green and violet pastel pencil',
'Rights': 'Albertina, Vienna'}

```

```

{'Title': 'The Tree-Man, c. 1505', 'Date Created': 'c. 1505', 'Painter':
'Hieronymus Bosch', 'Type': 'Pen and brown ink', 'Rights': 'Albertina, Vienna'}

```

```

{'Title': 'Venice, The Pink Cloud, 1909', 'Date Created': '1909', 'Painter':
'Paul Signac', 'Type': 'Oil on canvas', 'Rights': 'Albertina, Vienna - The
Batliner Collection'}

```

```

{'Title': 'Loïe Fuller in the Dance "The Archangel", 1902', 'Date Created':
'1902', 'Painter': 'Koloman Moser', 'Type': 'Indian ink and watercolor on
squared paper', 'Rights': 'Albertina, Vienna'}

```

```

{'Title': 'Half-Length Figure of an Apostle, 1493-1495', 'Date Created':
'1493-1595', 'Painter': 'Leonardo da Vinci', 'Type': 'Silverpoint, pen and brown
ink on blue prepared paper', 'Rights': 'Albertina, Vienna'}

```

```
{'Title': 'The Painter and the Buyer, 1565', 'Date Created': '1567', 'Painter':  
'Pieter Bruegel the Elder', 'Type': 'Pen and brown ink', 'Rights': 'Albertina,  
Vienna'}
```

```
{'Title': 'Seated Young Male Nude and Two Arm Studies (recto), c. 1510-11',  
'Date Created': 'c. 1510-11', 'Painter': 'Michelangelo Buonarroti', 'Type': 'Red  
chalk, heightened with white, over traces of a slate-stylus preliminary  
drawing', 'Rights': 'Albertina, Vienna'}
```

```
{'Title': 'Mary among a Multitude of Animals, c. 1503', 'Date Created': 'c.  
1503', 'Painter': 'Albrecht Dürer', 'Type': "Pen and blackish-brown ink,  
watercolor, squared in fine black chalk from the Madonna's head downwards",  
'Rights': 'Albertina, Vienna'}
```

```
{'Title': 'Old Houses in Krumau, 1914', 'Date Created': '1914', 'Painter': 'Egon  
Schiele', 'Type': 'Brush, body color, pencil on Japan paper', 'Rights':  
'Albertina, Vienna'}
```

[]: