Webscraping Project

November 21, 2021

[1]:
```
from splinter import Browser
from bs4 import BeautifulSoup
import re
import time
```

[2]:
```
!which chromedriver
```

```
/usr/local/bin/chromedriver
```

[3]:
```
executable_path = {'executable_path': '/usr/local/bin/chromedriver'}
browser = Browser('chrome', **executable_path)
```

0.1 Visit the NASA mars news site

[4]:
```
url = 'https://mars.nasa.gov/news/'
browser.visit(url)

# Delay for loading the page
browser.is_element_present_by_css("ul.item_list li.slide", wait_time=1)
```

[4]:
```
True
```

[5]:
```
html = browser.html
news_soup = BeautifulSoup(html, 'html.parser')

slide_elem = news_soup.select_one('ul.item_list li.slide')
```

[6]:
```
slide_elem.find("div", class_='content_title')
```

[6]:
```
<div class="content_title"><a href="/news/8613/a-year-of-surprising-science-
from-nasas-insight-mars-mission/" target="_self">A Year of Surprising Science
From NASA's InSight Mars Mission</a></div>
```

[7]:
```
news_title = slide_elem.find("div", class_='content_title').get_text()
news_title
```

[7]:
```
"A Year of Surprising Science From NASA's InSight Mars Mission"
```

[8]:
```
news_p = slide_elem.find('div', class_="article_teaser_body").get_text()
news_p
```

[8]:
```
"A batch of new papers summarizes the lander's findings above and below the
surface of the Red Planet."
```

0.2 JPL Space Images Featured Image

[9]:
```
url = 'https://www.jpl.nasa.gov/spaceimages/?search=&category=Mars'
browser.visit(url)
```

[10]:
```
full_image_elem = browser.find_by_id('full_image')
full_image_elem.click()
```

[11]:
```
browser.is_element_present_by_text('more info', wait_time=1)
more_info_elem = browser.find_link_by_partial_text('more info')
more_info_elem.click()
```

[12]:
```
html = browser.html
img_soup = BeautifulSoup(html, 'html.parser')
```

[13]:
```
img_url_rel = img_soup.select_one('figure.lede a img').get("src")
img_url_rel
```

[13]:
```
'/spaceimages/images/largesize/PIA01384_hires.jpg'
```

[14]:
```
img_url = f'https://www.jpl.nasa.gov{img_url_rel}'
img_url
```

[14]:
```
'https://www.jpl.nasa.gov/spaceimages/images/largesize/PIA01384_hires.jpg'
```

0.3 Mars Weather

[15]:
```
url = 'https://twitter.com/marswxreport?lang=en'
browser.visit(url)

time.sleep(5)
```

[16]:
```
html = browser.html
weather_soup = BeautifulSoup(html, 'html.parser')
```

[17]:
```
mars_weather_tweet = weather_soup.find('div', attrs={"class": "tweet",␣
↪"data-name": "Mars Weather"})
```

[18]:
```
try:
    mars_weather = mars_weather_tweet.find("p", "tweet-text").get_text()
    mars_weather
```

```
except AttributeError:

    pattern = re.compile(r'sol')
    mars_weather = weather_soup.find('span', text=pattern).text
    mars_weather

mars_weather
```

[18]:
```
'InSight sol 444 (2020-02-25) low -93.8ºC (-136.8ºF) high -12.0ºC
(10.5ºF)\nwinds from the SSW at 6.2 m/s (13.9 mph) gusting to 21.2 m/s (47.4
mph)\npressure at 6.30 hPa'
```

0.4 Mars Hemispheres

[19]:
```
url = 'https://astrogeology.usgs.gov/search/results?
 ↪q=hemisphere+enhanced&k1=target&v1=Mars'
browser.visit(url)
```

[20]:
```
hemisphere_image_urls = []

links = browser.find_by_css("a.product-item h3")

for i in range(len(links)):
    hemisphere = {}

    browser.find_by_css("a.product-item h3")[i].click()

    sample_elem = browser.find_link_by_text('Sample').first
    hemisphere['img_url'] = sample_elem['href']

    hemisphere['title'] = browser.find_by_css("h2.title").text

    hemisphere_image_urls.append(hemisphere)

    browser.back()
```

[21]:
```
hemisphere_image_urls
```

[21]:
```
[{'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/cerbe
rus_enhanced.tif/full.jpg',
  'title': 'Cerberus Hemisphere Enhanced'},
 {'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/schia
parelli_enhanced.tif/full.jpg',
  'title': 'Schiaparelli Hemisphere Enhanced'},
 {'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/syrti
s_major_enhanced.tif/full.jpg',
  'title': 'Syrtis Major Hemisphere Enhanced'},
```

```
 {'img_url': 'http://astropedia.astrogeology.usgs.gov/download/Mars/Viking/valle
s_marineris_enhanced.tif/full.jpg',
  'title': 'Valles Marineris Hemisphere Enhanced'}]
```

0.5 Mars Facts

[22]:
```
import pandas as pd
df = pd.read_html('https://space-facts.com/mars/')[0]
df.columns=['description', 'value']
df.set_index('description', inplace=True)
df
```

[22]:
```
                                          value
description
Equatorial Diameter:                      6,792 km
Polar Diameter:                           6,752 km
Mass:                  6.39 × 10^23 kg (0.11 Earths)
Moons:                            2 (Phobos & Deimos)
Orbit Distance:               227,943,824 km (1.38 AU)
Orbit Period:                     687 days (1.9 years)
Surface Temperature:                    -87 to -5 °C
First Record:                      2nd millennium BC
Recorded By:                     Egyptian astronomers
```

[23]:
```
df.to_html()
```

[23]:
```
'<table border="1" class="dataframe">\n  <thead>\n    <tr style="text-align:
right;">\n      <th></th>\n      <th>value</th>\n    </tr>\n    <tr>\n
<th>description</th>\n      <th></th>\n    </tr>\n  </thead>\n  <tbody>\n
<tr>\n      <th>Equatorial Diameter:</th>\n      <td>6,792 km</td>\n    </tr>\n
<tr>\n      <th>Polar Diameter:</th>\n      <td>6,752 km</td>\n    </tr>\n
<tr>\n      <th>Mass:</th>\n      <td>6.39 × 10^23 kg (0.11 Earths)</td>\n
</tr>\n    <tr>\n      <th>Moons:</th>\n      <td>2 (Phobos &amp;amp; Deimos)</td>\n
</tr>\n    <tr>\n      <th>Orbit Distance:</th>\n      <td>227,943,824 km (1.38
AU)</td>\n    </tr>\n    <tr>\n      <th>Orbit Period:</th>\n      <td>687 days
(1.9 years)</td>\n    </tr>\n    <tr>\n      <th>Surface Temperature:</th>\n
<td>-87 to -5 °C</td>\n    </tr>\n    <tr>\n      <th>First Record:</th>\n
<td>2nd millennium BC</td>\n    </tr>\n    <tr>\n      <th>Recorded By:</th>\n
<td>Egyptian astronomers</td>\n    </tr>\n  </tbody>\n</table>'
```

[24]:
```
browser.quit()
```