

【課題 19 : school_bus.py】

BeautifulSoup を用いて、配布した bus.html から「八王子駅南口行」の時刻情報を抽出し、図 1 のように出力せよ。BeautifulSoup に加えて正規表現も使用しても問題ない。読み込みファイルのパスは、コマンドライン引数で指定できるようにすること。

```
PS C:\Users\admin\Desktop\ProA2> python .\lec09\kadai19\school_bus.py .\lec09\data\bus.html
1 ['7:15', '7:30', '7:48']
2 ['7:28', '7:40', '7:58']
3 ['7:33', '7:45', '8:03']
4 ['7:38', '7:50', '8:08']
5 ['7:43', '7:55', '8:13']

80 ['20:50', '21:01', '21:19']
81 ['21:00', '21:11', '21:29']
82 ['21:10', '21:21', '21:39']
83 ['21:20', '21:31', '21:45']
```

図 1 課題 19 の実行結果：「八王子駅南口行」のバスは全部で 83 往復する。

リスト 1 school_bus.py

```
1 from bs4 import BeautifulSoup
2 import re
3 import sys
4
5 if __name__ == "__main__":
6     file_path = sys.argv[1]
7     pattern = re.compile("[0-9:]{1,}")
8     soup = BeautifulSoup(open(file_path, encoding="utf-8"), "html.parser")
9     table_tag = soup.find_all("table")[1]
10    time_table = []
11    for tr_tag in table_tag.find_all("tr"):
12        td_tags = tr_tag.find_all("td")
13        if td_tags:
14            if pattern.search(td_tags[0].string):
15                time_table.append([td_tags[j].string for j in range(3)])
16    for i, tt in enumerate(time_table, 1):
17        print(i, tt)
```

【課題 20 : school_bus.py】

Selenium を用いて以下のページを順に辿り、

- (1) 大学トップ : <https://www.teu.ac.jp/>
- (2) 「交通案内」 : <https://www.teu.ac.jp/campus/access/006644.html>
- (3) 「2022 年度スクールバス運行時刻表《月～金曜日運行》(2022 年 4 月 7 日 (木) より運行)2022 年 4 月 14 日更新」 : https://www.teu.ac.jp/campus/access/2022_0401bus.html

図 2 のように「八王子駅南口行」のバス時刻表のスクリーンショットを保存するコードを実装せよ。ただし、トップページを除いて、URL をハードコードせず Selenium により当該ページをクリック、または、URL を取得すること。出力画像のパスは、コマンドライン引数で指定できるようにすること。

キャンパス発	駅 着 発	キャンパス着	備 考
7:15	7:30	7:48	
7:28	7:40	7:58	
7:33	7:45	8:03	
7:38	7:50	8:08	
7:43	7:55	8:13	
7:48	8:00	8:18	
～	～	～	シャトル運行 (約3～5分間隔)
9:08	9:20	9:38	
9:13	9:25	9:43	
9:18	9:30	9:48	
9:23	9:35	9:53	
9:28	9:40	9:58	
9:33	9:45	10:03	
9:38	9:50	10:08	
9:43	9:55	10:13	
9:48	10:00	10:18	
9:53	10:05	10:23	
9:58	10:10	10:28	
10:03	10:15	10:33	

図 2 課題 20 の実行結果

リスト 2 school_bus.py

```

1 from selenium import webdriver
2 from selenium.webdriver.common.by import By
3 from time import sleep
4 import sys
5
6 if __name__ == "__main__":
7     driver = webdriver.Chrome("chromedriver")
8     driver.maximize_window() #ウインドウ最大化
9     driver.get("https://www.teu.ac.jp/")
10    driver.find_element(By.LINK_TEXT, "交通案内").click()
11    driver.find_element(By.LINK_TEXT, "2022年度スクールバス運行時刻表《月～金曜日運行》(2022年 4月 7
        日(木)より運行)2022年 4月 14日更新").click()
12    driver.find_element(By.LINK_TEXT, "八王子駅南口行").click()
13    driver.save_screenshot(sys.argv[1])

```

【課題 21 : highschool_code.py】

Selenium を用いて以下のページを順に辿り、

- (1) 大学トップ : <https://www.teu.ac.jp/>
- (2) 「入試・入学案内」 : <https://www.teu.ac.jp/entrance/index.html>
- (3) 「高校コード検索」 : <https://ientry.teu.ac.jp/tokyokoka/schoolsearch/index.aspx>

図 3 のように、学校名「東京工業高専」の学校コードを検索した結果のスクリーンショットを保存するコードを実装せよ。ただし、トップページを除いて、URL をハードコードせず Selenium により当該ページをクリック、または、URL を取得すること。出力画像のパスは、コマンドライン引数で指定できるようにすること。



図 3 課題 21 の実行結果

リスト 3 highschool_code.py

```

1 from selenium import webdriver
2 from selenium.webdriver.common.by import By
3 from time import sleep
4 import sys
5
6 if __name__ == "__main__":
7     driver = webdriver.Chrome("chromedriver")
8     driver.maximize_window() #ウインドウ最大化
9     driver.get("https://www.teu.ac.jp/")
10    driver.find_element(By.LINK_TEXT, "入試・入学案内").click()
11    a_tag = driver.find_element(By.LINK_TEXT, "高校コード検索")
12    url = a_tag.get_attribute("href")
13    driver.get(url)
14    search_bar = driver.find_element(By.NAME, "txtKokoNm")
15    search_bar.send_keys("東京工業高専")
16    driver.find_element(By.LINK_TEXT, "検索").click()
17    driver.save_screenshot(sys.argv[1])

```