

07_scrape_01: Web scraping

Objective: Basic Scraping a Web Page HTML by BeautifulSoup

Problems

Q1 Function:

Write function Q1(file_path). that counts the number of Buddhist holy days (วันพระ) in each year provided. The user will first enter a file path HTML for which it wants to count Buddhist holy day. The function will count the Buddhist holy days in each specified year and return a list that contains the number of Buddhist holy days for each year, from the first Monday (index 0) until the following Sunday.

Input: file_path HTML (e.g. "2566.html")

Output: A list that contains the number of Buddhist holy days for each year, counted from the first Monday until the following Sunday (e.g., [8, 6, 8, 6, 6, 7, 9])

Q2 Function:

Write function Q2(file_path) that finds Visakha Bucha Day (วันวิสาขบูชา) for each year provided in the specified HTML file. The user will input a file path to an HTML document, as in Q1. The function will return a string representing the Visakha Bucha Day in the format (dayWeek, monthNumber, month, year), for example, "วันเสาร์ที่ 3 มิถุนายน 2566".

Input: file_path HTML (e.g. "2566.html")

Output: A string containing the Visakha Bucha Day for each year (e.g., "วันเสาร์ที่ 3 มิถุนายน 2566").

Expected Results

Input	Output
print(Q1("2566.html"))	[8, 6, 8, 6, 6, 7, 9]
print(Q2("2566.html"))	วันเสาร์ที่ 3 มิถุนายน 2566

Template codes

```
import ...

def Q1(file_path): # DO NOT modify this line
    # remove pass and replace with you code
    Pass

def Q2(file_path): # DO NOT modify this line
    # remove pass and replace with you code
    Pass

exec(input().strip()) # do not delete this line
```

Example HTML file attached

```
<!DOCTYPE html>
<!-- saved from
url=(0162)https://www.myhora.com/%E0%B8%9B%E0%B8%8F%E0%B8%B4%E0%B8%97%E0%B8%B4%E0%B8%99
/%E0%B8%A7%E0%B8%B1%E0%B8%99%E0%B8%9E%E0%B8%A3%E0%B8%B0-%E0%B8%9E.%E0%B8%A8.2566.aspx -
->
<html class="fontawesome-i2svg-active fontawesome-i2svg-complete"><head><meta
http equiv="Content-Type" content="text/html; charset=UTF-8">

...

```