

Python程式設計#9補充

林奇賦 daky1983@gmail.com

Outline

- POST
- GET
- 如何送表單、改變header

POST

- POST方法是將要傳送的資訊放在message-body中
- 使用POST方法就不用擔心資料大小的限制，可以防止使用者操作瀏覽器網址，表單的資料被隱藏在message-body中，因此，在大多數的情況下，使用POST方法將表單資料傳到Web Server端是更有效的方法。

GET

- GET就是指在網址上指定變數的名稱及變數的值給網頁伺服器，使用GET是有一個上限的所以較不適合用來傳送大量的資料或訊號。
- 一個簡單的GET就是在網址尾部加上一個問號？之後進行宣告傳送的變數。而傳送的格式是：“變數名稱＝變數的值”。
- 若有多個變數需要傳遞則用 & 符號隔開。

如何送表單、改變header

- urllib.parse
- Reference
 - <https://docs.python.org/3/library/urllib.parse.html#urllib.parse.urlencode>

範例

```
# -*- coding: utf-8 -*-
import urllib.request
import urllib.parse
from html.parser import HTMLParser

data = urllib.parse.urlencode({'sselect': '039', 'action':
'query', 'sselect': '051', 'submit': '確定'})
data = data.encode('utf-8')
request =
urllib.request.Request("http://web.trtc.com.tw/c/2stainfo_new.asp")
request.add_header("Content-Type", "application/x-www-form-
urlencoded; charset=utf-8")
response = urllib.request.urlopen(request, data)
html = response.read().decode('utf_8')
```

```
class myparser(HTMLParser):
    def __init__(self):
        HTMLParser.__init__(self)
        self.isNumber = 0
        self.numbers = []
    def handle_data(self, data):
        if self.isNumber == 1:
            print(data)
            self.isNumber = 0
    def handle_starttag(self, tag, attrs):
        if tag == 'font' and attrs == [('size', '-1')]:
            self.isNumber = 1
```

```
Parser = myparser()
Parser.feed(html)
```

課堂練習

- 抓取高鐵網頁的時刻表
 - <http://www.thsrc.com.tw/tw/TimeTable/SearchResult>
 - 出發站:台北
 - 到達站:左營
 - 日期時間任意
 - 不需要解析內容，查詢的結果網頁內容抓取下來即可