

自訂請求表頭

Responsive Web Design(rwd): 自適應不同裝置大小

network(重新整理) -> header -> user-agent(代理人)(瀏覽器)

沒發送表頭，Google會發送手機版的版面，
所以抓到的資料跟瀏覽器的不一樣

class_=""，以空格分割的字串，可選其中一段來篩選
.find_all('h3',class_="LC20lb MBeuO DKV0Md")
.select("h3.LC20lb.MBeuO.DKV0Md") # 找多個
.select_one("h3.LC20lb.MBeuO.DKV0Md") # 找一個
div -> a ... href
class搜索器: .select()
links = soup.select("div.yuRUbf")
for i in link: print(i.a.get("href"))
links = soup.select("div.yuRUbf a") 或是空格換成>
for i in link: print(i.text))
>: 抵達下一層，空格: 找到為止
links = soup.select("div.yuRUbf > div > span > a")
可混合使用，soup.select("div.yuRUbf > div a")
select無法用attrs，如果是id而非class用#

正規表示法select會有問題

向前、向上、向後搜尋，sibling兄弟
with open("index.html") as f:
soup=BeautifulSoup(f)

表頭會送到到伺服器(後端)

GET請求: 把搜尋字放到網址送過去，網址可被看到，
且網址有限定長度，理論可傳50bytes的檔案

POST請求: 寫表單送出去(登入、註冊)，也可送檔案；
後端接收後，Headers旁邊Payload頁會有送出的資料

files = {'my_filename': open('','rb')
request.post('',files=files)

搜尋引擎右鍵檢查，input參數上面有form參數，
會顯示由GET還是POST送出

例子: PTT八卦18歲認證，Network拉到最上面，
over18 -> Headers會看到POST請求

Cookie: 資料暫存在客戶端(瀏覽器)

Session: 資料暫存在伺服器，如帳號資訊

F12 -> Application -> Cookies -> 可以Clear
Cookies會伴隨Session，

存入Session會產生Session ID存在Cookies，
如果沒看到Session ID就代表不是用Session存
抓取時間可用來做停止條件判斷

兩層迴圈要結束，需要變成一個def

索引片段子: [1:] = [1::]，沒結尾可省略；
[::-1]由尾巴到頭部，頭尾都被省略剩下":"連結
{':'} = dict(x = '')

int 4個byte，一次加32個bit；
1 個字節(byte)等於 8 個位元 (bits)

1 KB = 1,024 字節 (Bytes)：
計算機內存和存儲，如硬碟容
量、文件大小

1 Kb = 1,000 位元 (bits)：
網絡傳輸速率，如網絡速度，
1Kbps(Kilobits per second)