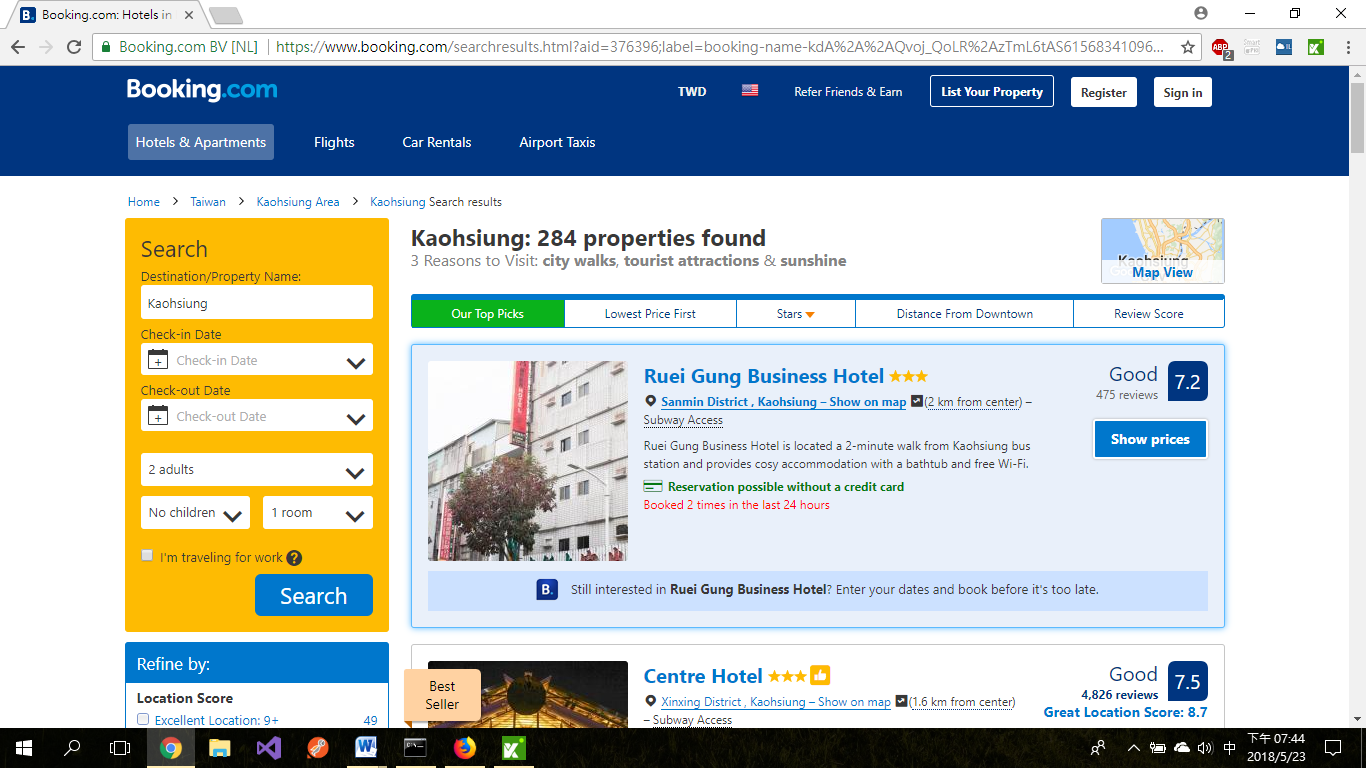
**如何抓booking每頁的旅館評價**

**安裝Katalon Selenium IDE(for Chrome)、chromedriver.exe**

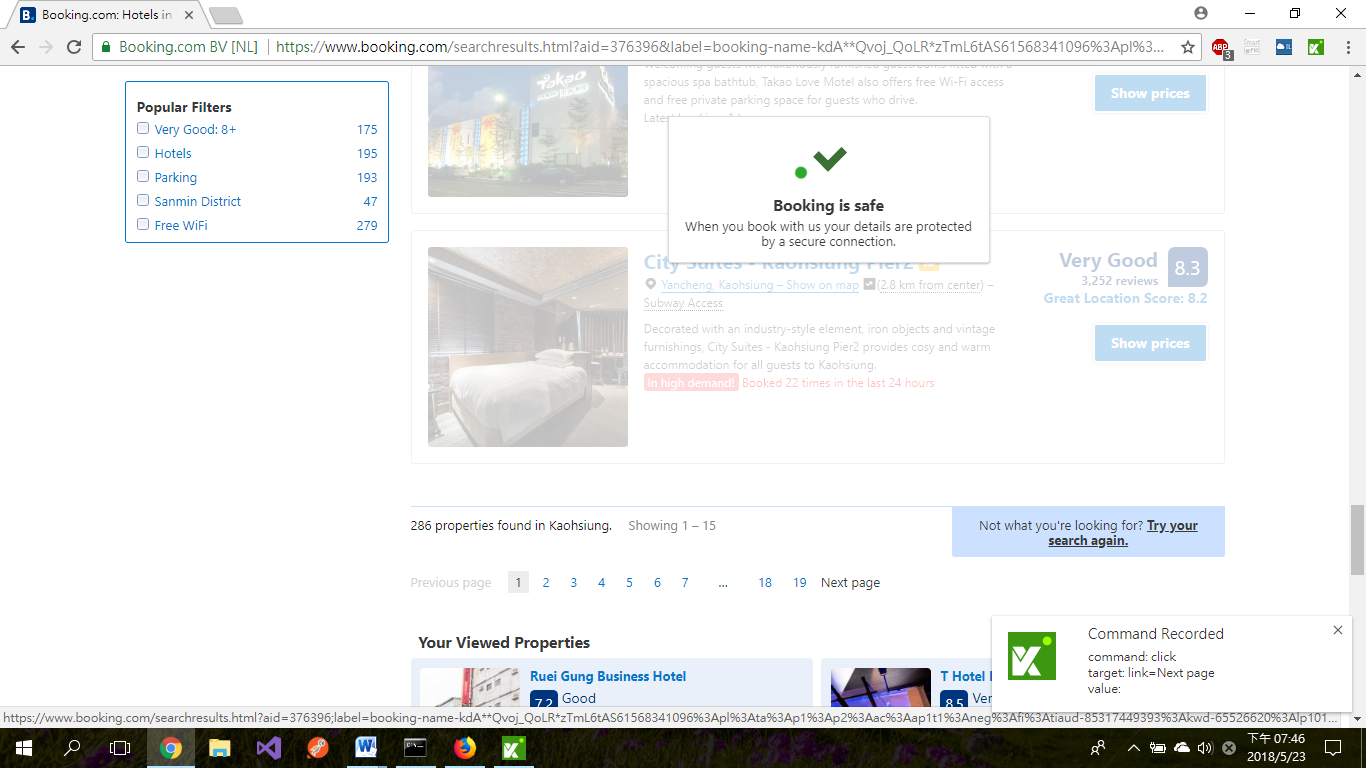


按這啟動

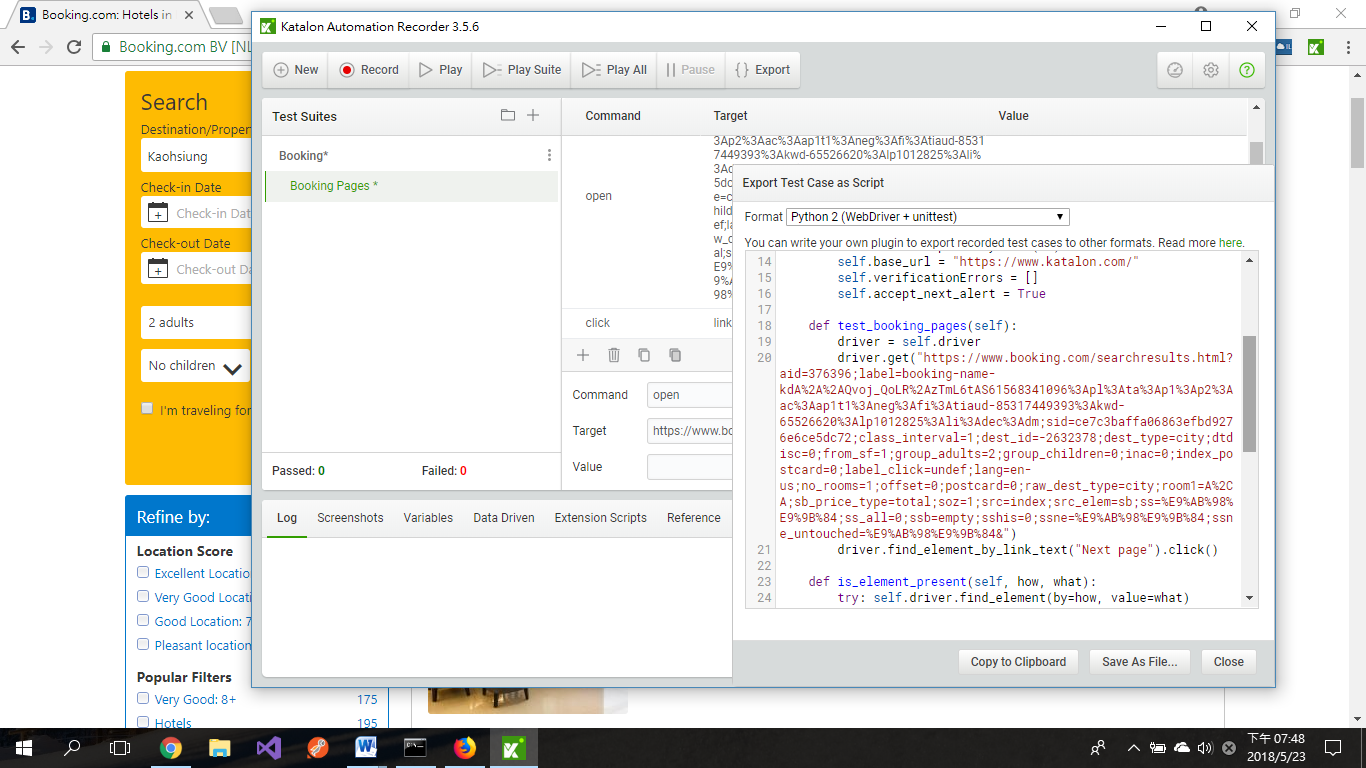
**第一部分:翻頁和抓取旅館英文名字**

**用英文版是因為後面需抓出英文旅館名放進各連結裡**

1. **按New新建case**
2. **按Record開始錄製**
3. **按下一頁**



1. **按Stop後，再按Export**



複製 ->

**部分程式**

chrome\_path = "D:/cdriver/chromedriver.exe"

#chromedriver.exe執行檔所存在的路徑

driver = webdriver.Chrome(chrome\_path) #預設使用Chrome開啟

#英文版booking，使用 GET 方式下載普通網頁

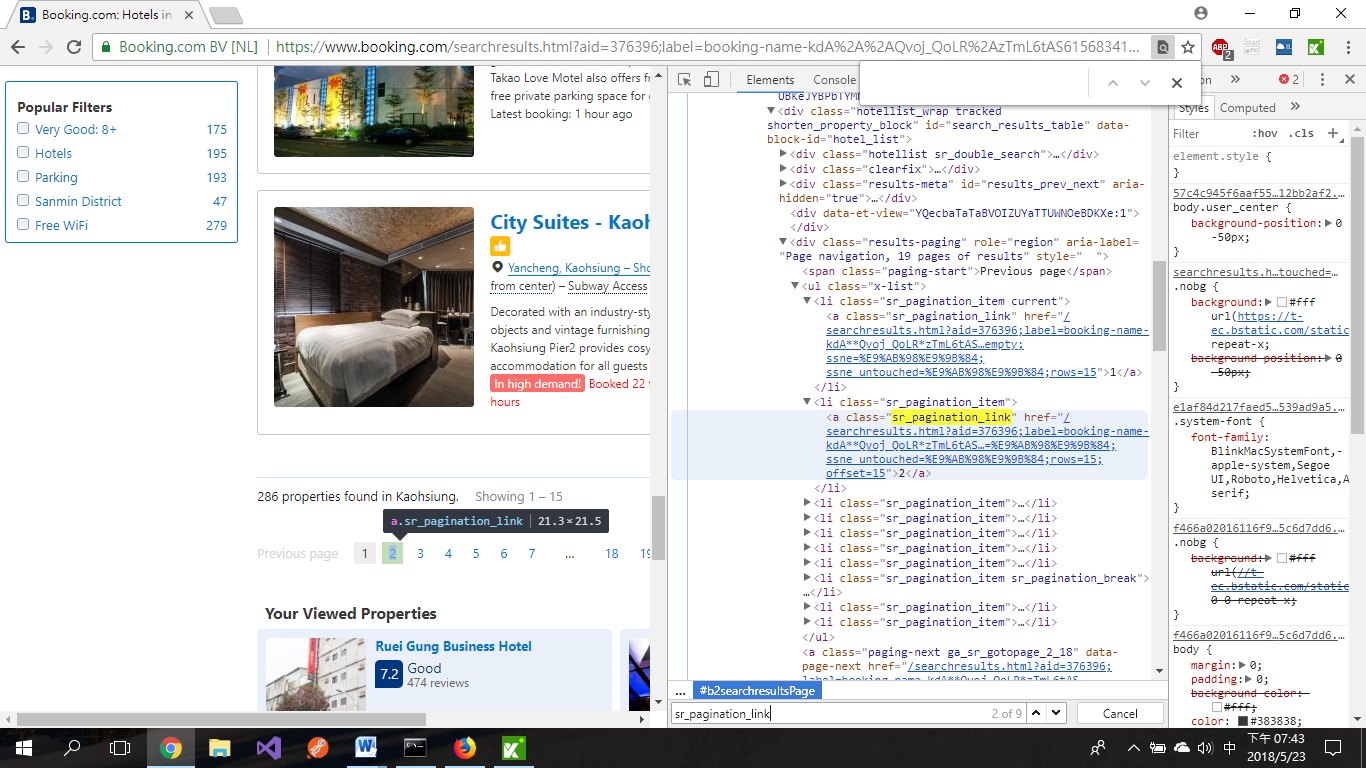
driver.get("https://www.booking.com/searchresults.en-gb.html?aid=376396&label=booking-name-kdA......")

soup=BeautifulSoup(driver.page\_source,'lxml')

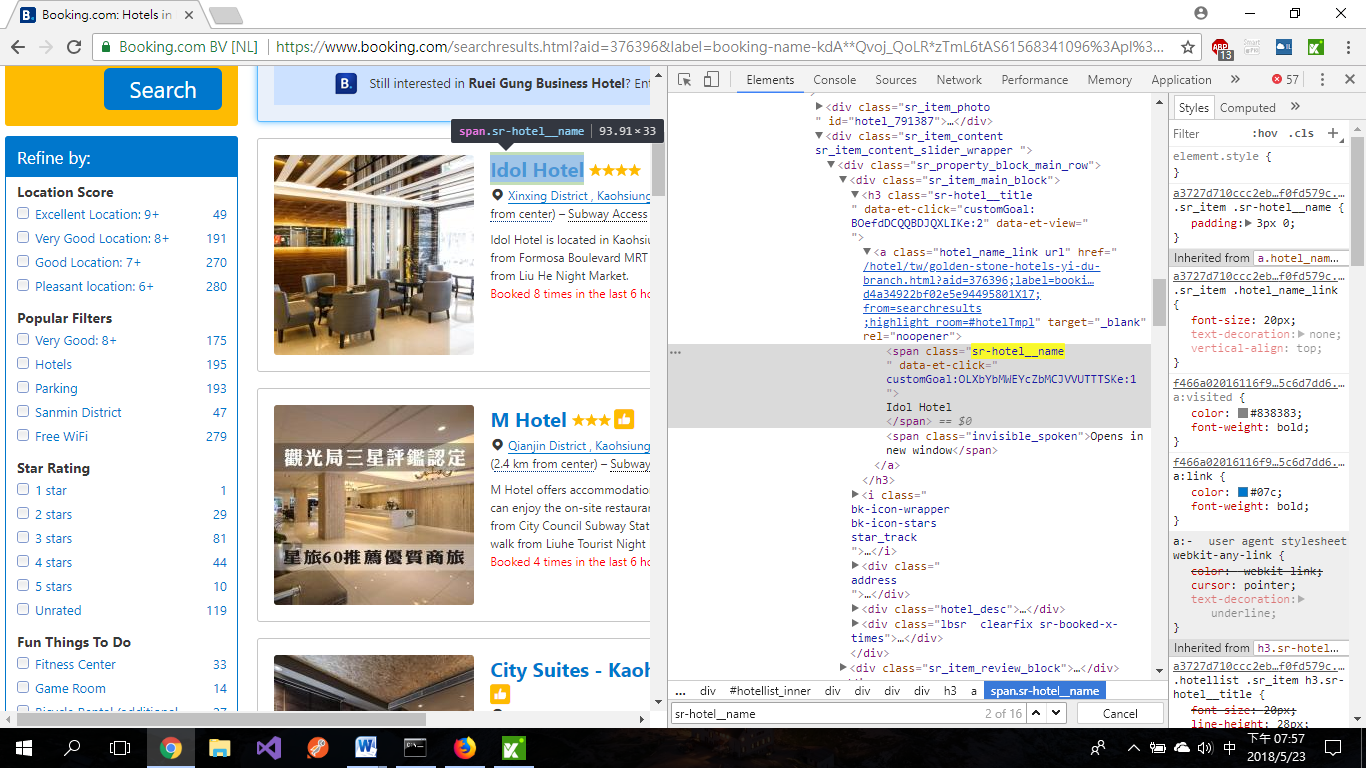
#得到他的原始碼，並用BeautifulSoup剖析內容，lxml乃一好用的解析器

**檢視原始碼**

while len(soup.select('.sr\_pagination\_link')) > 0: #(找(各頁))這個長度是否 > 0

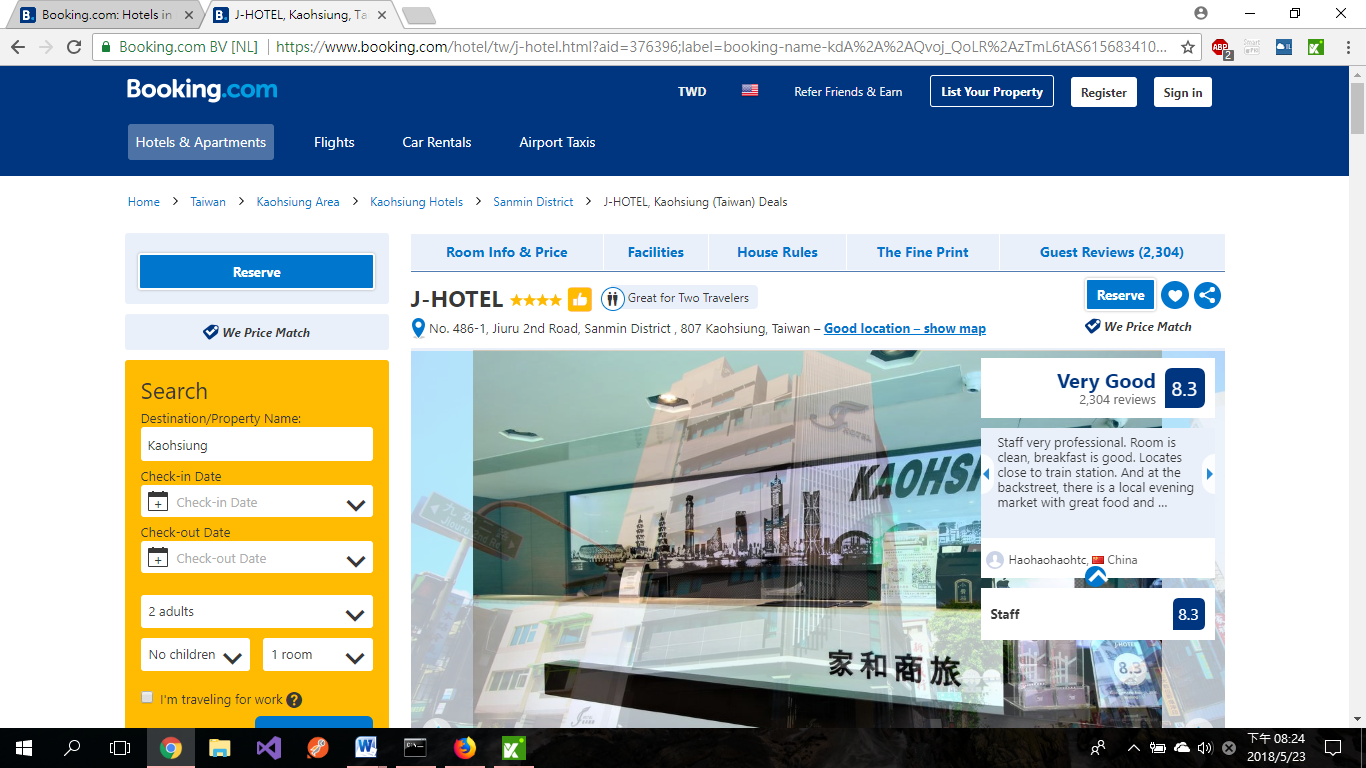


for hotelname in soup.select('.sr-hotel\_\_name'): #把這頁裡面的連結全部找出來(旅館的英文名字)



**第二部分:轉換成中文網站**

**觀察各旅館的網址，大部分都是以這形式表示**





**部分程式**

turnname=hotelname.text #.text 可以取得不含 html 標籤的純文字內容。

lowname=turnname.lower() #把大寫轉為小寫

lowname = "-".join(lowname.split()) #按字符串的每個空格分割，換成-連接

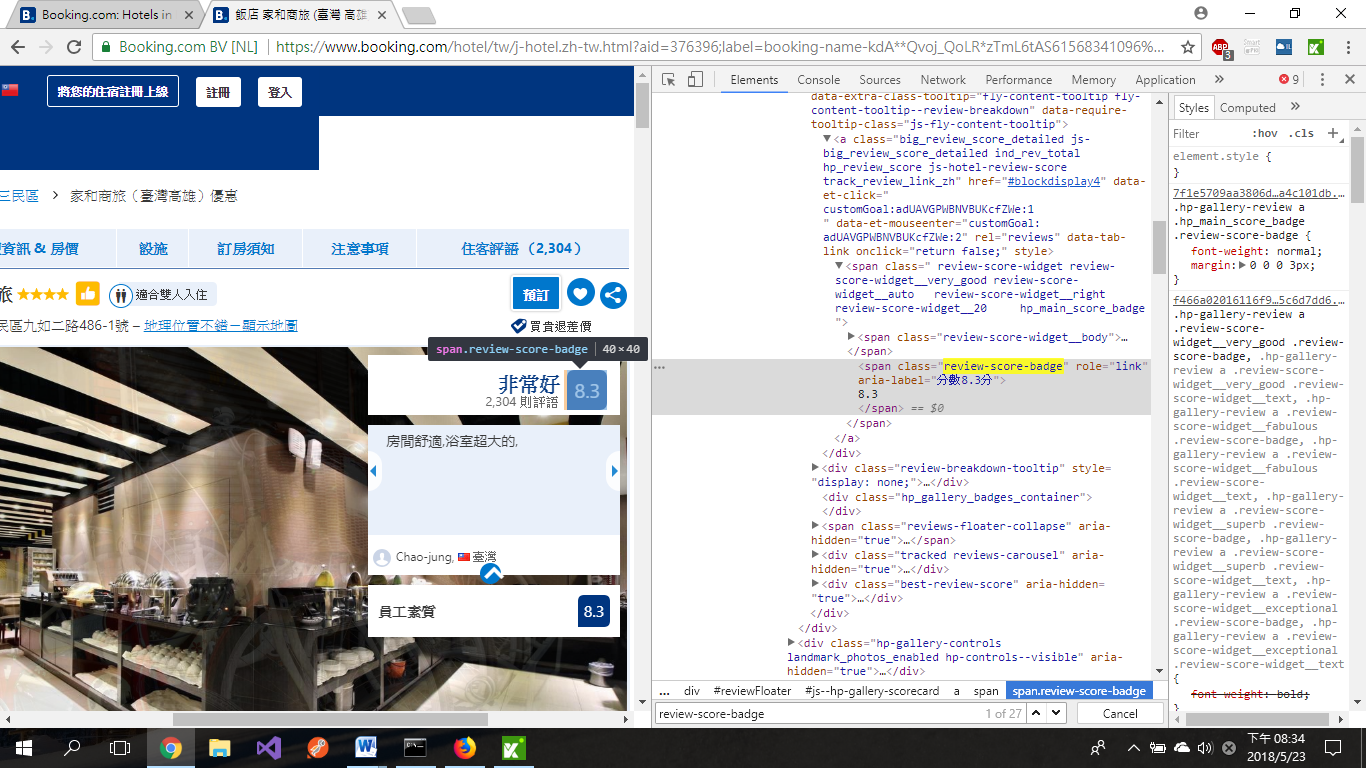
#用format，如把 t-hotel-kaohsiung 放進連結{}裡 commentsURL='https://www.booking.com/hotel/tw/{}.zh-tw.html?aid=304142; ...... '.format(lowname)

**第三部分:抓取每個旅館的中文名和評價**

title=soupc.select('#hp\_hotel\_name') #找出中文版booking的旅館名



score=soupc.select('.review-score-badge') #整體評分



for review in soupc.select('.review\_item\_review'): #找該旅館的評論

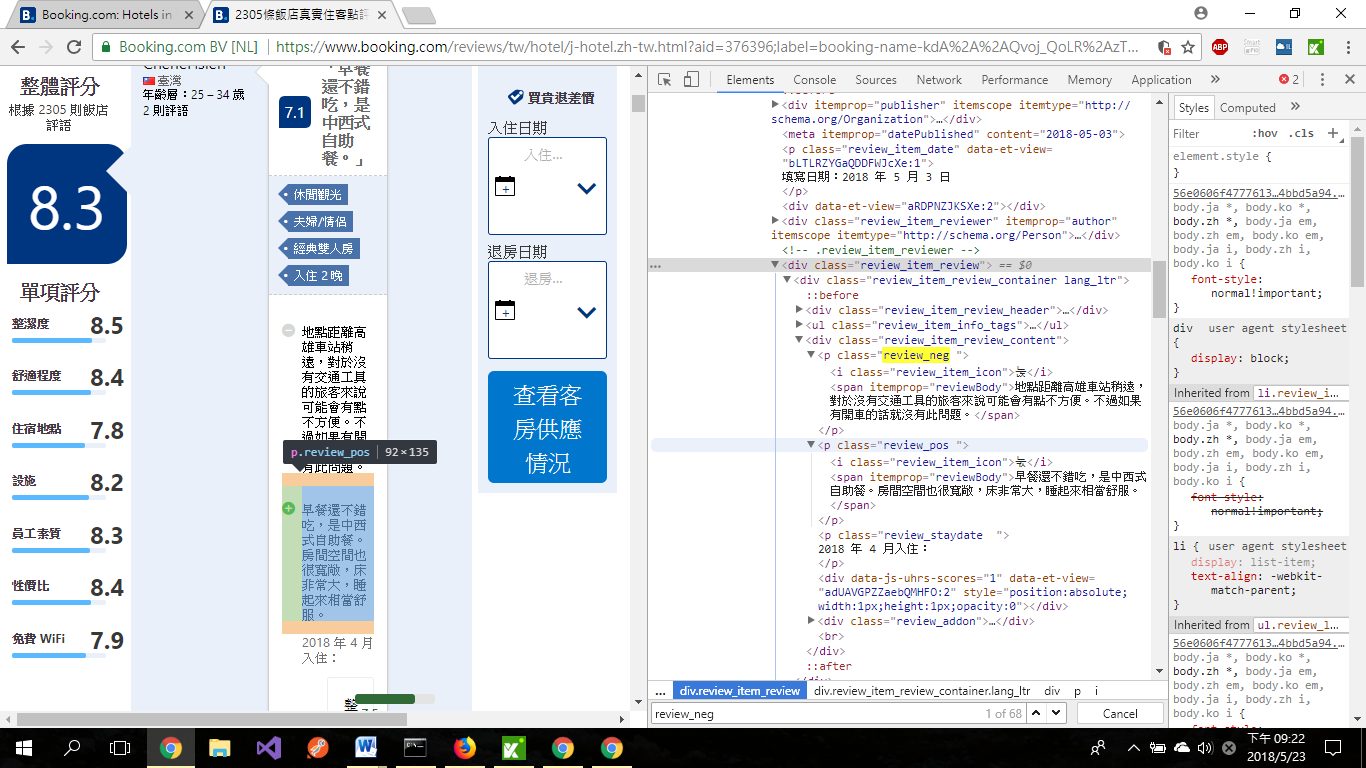
if(review.select('.review\_neg')): #負評

print("-:",review.select('.review\_neg')[0].i.next\_sibling)

# .i.next\_sibling，往旁邊爬留言，回傳同一階層的標籤。<i class="review\_item\_icon">눉</i>。即不要눉

if (review.select('.review\_pos')): #正評

print("+:",review.select('.review\_pos')[0].i.next\_sibling)



**完整程式和結果**

