



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

[Python爬蟲教學]Python網頁爬蟲動態翻頁的實作技巧

6月 14, 2020



分享

標籤

Python 爬蟲教學

Photo by True

Agency on Unsplash



【免費課程】三個你不知道的Python網頁爬蟲秘密

下，除了會使用
滾動捲軸來動態
載入更多的資料
外，另一種最常
見的方式就是利
用分頁，來分別
顯示特定筆數的
資料。

如果Python網頁
爬蟲遇到需滾動
捲軸，才可爬取
更多資料的實作
方式，可以參考
[\[Python 爬蟲教
學\] 整合 Python](#)



【免費課程】三個你不知道的Python網頁爬蟲秘密

頁的網站時，該如何翻頁爬取內容，就是本文要來分享的主題。其中的重點包含：

- HTTP GET
vs POST方法
- 頁碼使用
GET的方式
- 頁碼使用
POST的方式

一、HTTP GET vs POST方法



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

GET 及 POST 方法。

HTTP 的 GET 方法就像明信片，在上面需要撰寫目的地的地址及內容，其中的地址就像是請求的網址，而內容就像是接在網址後面的參數 (querystring)，只要拿到這個明信片(網址)的人，都知道要寄送到



【免費課程】三個你不知道的Python網頁爬蟲秘密

件，在信封上需
撰寫目的地的地
址，而信件內容
會在信封中，所
以拿到這封信的
人只知道會寄送
到哪裡，但不會
知道內容是什
麼，就像是知道
請求的網址，但
是其中送到伺服
器端的資料，無
法在網址知道。

瞭解了這兩個
HTTP 方法的基



【免費課程】三個你不知道的Python網頁爬蟲秘密

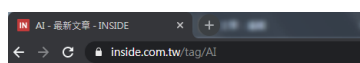
立即索取

的網站，Python
網頁爬蟲該如何
翻頁爬取資料。

二、頁碼使用

GET的方式

大部分的網站分
頁，都是使用
GET 方法來處
理，這邊以
**Inside 硬塞的網
路趨勢觀察**網站
的「AI」類別為
例，可以看到網
址如下：





【免費課程】三個你不知道的Python網頁爬蟲秘密

式，從網址就可以知道目的地網址及傳至伺服器端的頁碼參數(page)，當變換page參數的值時，就可以前往對應的頁碼網頁。

接下來，在網頁上按F12鍵，進入開發者模式後，先來看一下要爬取的文章標



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

BeautifulSoup 套件爬取第一頁的內容，如下範例：

```
1.  from bs4 import BeautifulSoup
2.  import requests
3.
4.
5.  # 連結網站
6.  response = requests.get("https://www.learnpython.org/en/Hello_Python_World")
7.
8.
9.  # HTML原始碼解碼
10. soup = BeautifulSoup(response.text, 'html.parser')
11.
12. # 取得所有class="list-item"的標題
13. titles = soup.find_all(class_='list-item')
14.
15. for title in titles:
16.     print(title.get_text())
```

執行結果



【免費課程】三個你不知道的Python網頁爬蟲秘密

圈，透過變換網址
頁碼參數
(page)的方式，
翻頁爬取網頁內
容了，如下範
例：

```
1. from bs4 import BeautifulSoup
2. import requests
3.
4. for page in range(1, 10):
5.
6.     # 連結網站
7.     response = requests.get(
8.         "https://www.learncodewithmike.com/2020/06/how-to-scrape-different-pages-using-python-scraper.html",
9.         headers={'User-Agent': 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7; rv:68.0) Gecko/20100101 Firefox/68.0'})
10.
11.     # HTML 原始碼
12.     soup = BeautifulSoup(response.text, 'html.parser')
13.
14.     # 取得所有 title
15.     titles = soup.find_all('h1')
```



【免費課程】三個你不知道的Python網頁爬蟲秘密

如果要更多頁的
話可以自行調整
第4行的 `range()`
參數。另外，要
特別注意的是，
Python迴圈讀取
`range()` 函式時，
`page` 的型別為數
值 (Integer)，所
以為了要加在網
址後面，需透過
`str()` 函式轉型成
字串，如範例中
第8行。詳細的
Python 轉型教學
可以參考

[Python 教



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

POST的方式

另一個網站頁碼
切換的方式會採
用 POST 方法，
就像信件一樣，
知道寄送的地址
(網址)，卻不會
知道信封中的內
容(送往伺服器的
頁碼參數)。

這邊以 **591 房屋
交易網** 的「租
屋」類別為例，
可以看到網址如
下：



【免費課程】三個你不知道的Python網頁爬蟲秘密

頁時，會發現網址並不會像 **Inside 硬塞的網路趨勢觀察** 網站一樣，會出現頁碼的參數，而是和第一頁的網址一模一樣，都沒有動靜。

這時候，就需要結合 Selenium 套件，透過模擬使用者點擊下一頁按鈕，來進行換頁的動作。同樣



【免費課程】三個你不知道的Python網頁爬蟲秘密

接著，就可以利用 Selenium 套件開啟瀏覽器並且連線至 [591 房屋交易網](#) 的「租屋」類別，再使用 BeautifulSoup 套件爬取第一頁的租屋標題。首先，利用以下指令安裝開啟瀏覽器的 WebDriver 套件：

```
1. $ pip install
```



【免費課程】三個你不知道的Python網頁爬蟲秘密

```
2.     from selenium
3.     from webdriver
4.     from bs4 import
5.
6.
7.     options = Opti
8.     options.add_ar
9.
10.    browser = webd
11.        ChromeDriv
12.
13.    browser.get("
14.
15.    area = browser
16.    area.click()
17.
18.    soup = Beautif
19.
20.    # 取得所有cLas
21.    elements = sou
22.
23.    for element in
24.        # 取得<Li>
25.        title = el
26.        print(titl
```



【免費課程】三個你不知道的Python網頁爬蟲秘密

怎麼稱呼

電子郵件

立即索取

全新裝潢，近基隆長庚醫院國聯國社區
無仲介費，房東親自帶看屋，價格優惠
近海大二信，環境安全優美，交通便利
市區透天雅房，精心設計，鬧中取靜超舒服！
經國、全民一家設備齊全獨立套房
海景獨立房有電梯近廟口機能通風佳
基隆全新裝潢小套房出租
全新裝潢出租套房
巴塞隆納社區一間女雅房出租(近海洋大學)
43巷29號三樓(崇佑技術學院、崇佑企專
近省立醫院近信義區公所崇佑技術學院近肯德
瑞陵坑，鄰近七堵及大埔交流道

由於要點擊下一
頁按鈕來進行換
頁，所以，來觀
察一下頁碼區域
的 HTML 原始
碼，如下圖：

```
<!--分页-->
<div class="page-limit">
  <div class="pageBar">
    <a href="#" class="pageFirst" data-first="0" data-total="523" class="pageFirst">1</a>
    <a href="#" class="pagePrev" data-first="0" data-total="523" class="pagePrev"></a>
    <a href="#" class="pageNext" data-first="523" data-total="523" class="pageNext">524</a>
    <a href="#" class="pageLast" data-first="523" data-total="523" class="pageLast">524</a>
  </div>
  <div class="pageInfo">
    <span>共 523 頁</span>
  </div>
</div>
```

得知下一頁按鈕
的樣式類別
(class) 為
「pageNext」
後，就可以利用



【免費課程】三個你不知道的Python網頁爬蟲秘密

```
2.     from selenium
3.     from webdriver
4.     from bs4 import
5.
6.
7.     options = Opti
8.     options.add_ar
9.
10.    browser = webd
11.        ChromeDriv
12.
13.    browser.get("
14.
15.    area = browser
16.    area.click()
17.
18.    soup = Beautif
19.
20.    # 取得所有cLas
21.    elements = sou
22.
23.    for element in
24.        # 取得<Li>
25.        title = el
26.        print(titl
```




【免費課程】三個你不知道的Python網頁爬蟲秘密

執行以上的動作，來達到動態翻頁爬取網頁內容的目的，如下範例：

```
1. from selenium
2. from selenium
3. from webdriver
4. from bs4 import
5. import time
6.
7.
8. options = Options
9. options.add_argument
10.
11. browser = webdriver
12. ChromeDriver
13.
14. browser.get("
15.
16. area = browser
17. area.click()
```



【免費課程】三個你不知道的Python網頁爬蟲秘密

```
26.         print(f"==> {element.get('title')}")
27.     for element in soup.find_all('div', class_='list-item'):
28.         # 取得
29.         title
30.         print(
31.
32.         page_next
33.         page_next
34.
35.         time.sleep(1)
```

四、小結

以上就是Python
網頁爬蟲在遇到
有分頁的網站
時，能夠動態換
頁爬取網頁內容
的實作方式，詳



【免費課程】三個你不知道的Python網頁爬蟲秘密

[立即索取](#)

如果您喜歡我的
文章，請幫我按
五下Like(使用
Google或
Facebook帳號免
費註冊)，支持我
創作教學文章，
回饋由LikeCoin
基金會出資，完
全不會花到錢，
感謝大家。



27 個讚

註冊 LikeCoin ID 免費投讚



Mike Ku

[按月贊助](#)



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

有想要看的教學

內容嗎？歡迎利

用以下的Google

表單讓我知道，

將有機會成為教

學文章，分享給

大家😊

[https://forms.gle/](https://forms.gle/UW8u9XddoY17HjaSA)

[UW8u9XddoY17](https://forms.gle/UW8u9XddoY17HjaSA)

[HjaSA](https://forms.gle/UW8u9XddoY17HjaSA)

Python學習資源

- Python學習

資源整理



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

戰

- Python 網頁

爬蟲入門實

戰

Python網頁爬蟲

- BeautifulSoup

教學

- [Python爬蟲

教學]7個

Python使用

BeautifulSou

p開發網頁爬

蟲的實用技

巧



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

Selenium 及

BeautifulSou

p實現動態網

頁爬蟲

- [Python 爬蟲教學]3 個建構 Python 動態網頁爬蟲重要的等待機制
- [Python 爬蟲教學]學會使用 Selenium 及 BeautifulSou



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

載入網頁的 爬取秘訣

Python非同步網 頁爬蟲

- [Python爬蟲
教學]整合
asyncio與
aiohttp打造
Python非同
步網頁爬蟲
- [Python爬蟲
教學]非同步
網頁爬蟲使
用GRequests
套件提升爬



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

- [\[Python爬蟲教學\]Python網頁爬蟲結合LINE Notify打造自動化訊息通知服務](#)
- [\[Python爬蟲教學\]有效利用Python網頁爬蟲幫你自動化下載圖片](#)
- [\[Python+LINE Bot教學\]建構具網頁爬蟲功能的](#)



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

有網頁爬蟲

功能的關鍵

字查詢網頁

Python網頁爬蟲

部署

- [Python爬蟲

教學]教你如何

何部署

Python網頁

爬蟲至

Heroku雲端

平台

Python網頁爬蟲

資料儲存



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

庫的整合方式

- [Python爬蟲教學]活用openpyxl套件將爬取的資料寫入Excel檔案
- [Python爬蟲教學]解析如何串接Google Sheet試算表寫入爬取的資料

Python網頁爬蟲技巧



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

用方法

- [\[Python爬蟲教學\]開發Python網頁爬蟲前需要知道的五個基本觀念](#)
- [解析Python網頁爬蟲如何有效整合Pandas套件提升資料處理效率](#)



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

網頁爬蟲線上課程 問卷

現在網路發達的時代，
如何在有限的時間及專
注力下，有效掌握網路
上的資料並且加以應
用，將會是大數據時代
下非常重要的能力。

而**網頁爬蟲**就是最常使
用的技術，透過定時自

0%



Learn Code With Mike

歡迎訂閱以收到免費的
Python教學內容及活動訊
息

對什麼主題有興趣呢



【免費課程】三個你不知道的Python網頁爬蟲秘密

[立即索取](#)

BUILT WITH  ConvertKit

標籤：

[PYTHON爬蟲教學](#)

[分享](#)

留言



genius

· 2021年6月30日

晚上11:57

你好，謝謝你詳細的分享～～想另外請教一下，目前的教學比較著重在去爬「已知網站」的網頁資料；那如果是想要爬「未知網站」的資料，因為是未知的網頁結構，會有比較建議用什麼方式或套件嗎～應用場景大概是：使



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

...

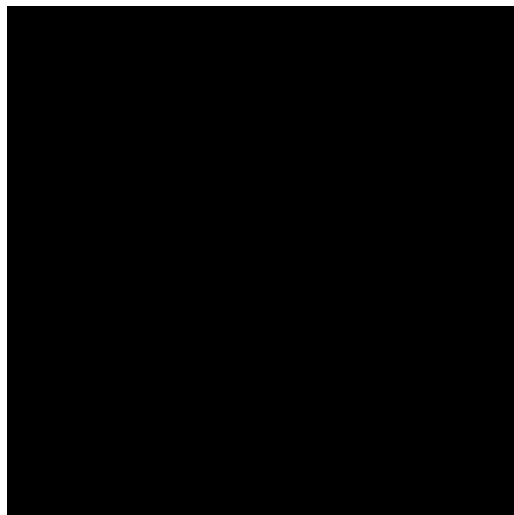


這個網誌中的熱門文章

**[Python
爬蟲教
學]7個
Python
使用
Beautifu
lSoup開
發網頁爬
蟲的實用
技巧**

2月 28, 2020

Photo by
Stanley Dai
on
Unsplash ...



【免費課程】三個你不知道的Python網頁爬蟲秘密

怎麼稱呼

電子郵件

立即索取

向]淺談
Python
類別
(Class)

1月 02, 2020

Photo by
Bram Naus
on
Unsplash ...

分享
27 則留言
閱讀完整內容

[Pandas
教學]資
料分析必
懂的
Pandas
DataFra

×

【免費課程】三個你不知道的Python網頁爬蟲秘密

怎麼稱呼

電子郵件

立即索取

on
Unsplash ...

分享
1 則留言
閱讀完整內容

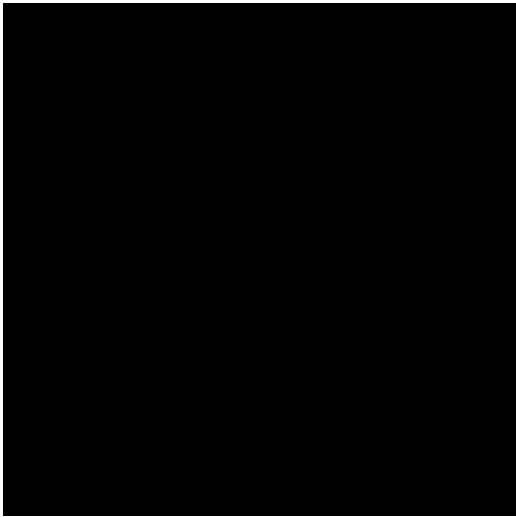
[Python
教學]搞
懂5個
Python
迴圈常見
用法

12月 04, 2019

Photo by
Scott Webb

on
Unsplash ...

分享
4 則留言





【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

快速上手
LINE
Bot機器人

6月 28, 2020

Photo by
Yura Fresh
on
Unsplash ...

分享

47 則留言

閱讀完整內容

[Python
教學]5個
必知的
Python
Functio

【免費課程】三個你不知道的Python網頁爬蟲秘密

怎麼稱呼

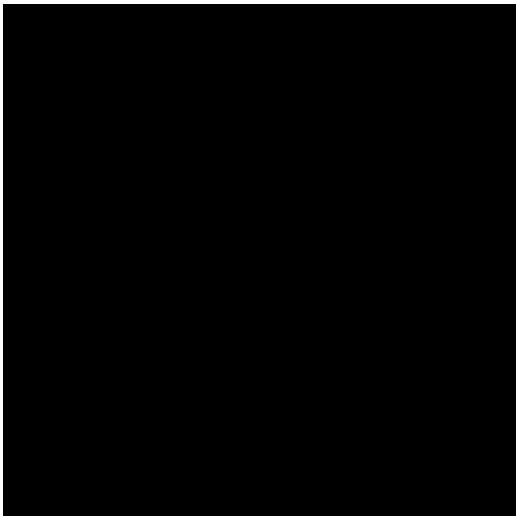
電子郵件

立即索取

Unsplash ...

- 分享
- 張貼留言
- 閱讀完整內容

[Pandas
教學]5個
實用的
Pandas
讀取
Excel檔
案資料技
巧



12月 06, 2020

Photo by
LinkedIn
Sales
Navigator ...

分享



【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

n Lambda Function n應用技 巧分享

12月 29, 2019

Photo by

Fatos

Bytyqi on

Unsplash ...

分享

6 則留言

閱讀完整內容

[Python 爬蟲教 學]整合 Python Seleniu

【免費課程】三個你不知道的Python網頁爬蟲秘密

怎麼稱呼

電子郵件

立即索取

Photo by
LAUREN
GRAY on
Unsplash ...

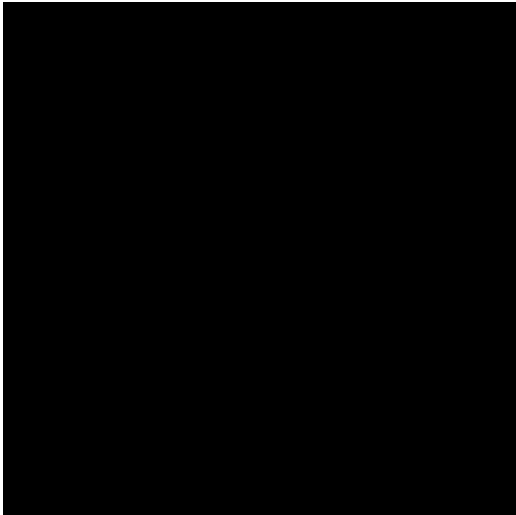
分享

18 則留言

閱讀完整內容

[Python
爬蟲教
學]有效
利用
Python
網頁爬蟲
爬取免費
的Proxy
IP清單

10月 24, 2021





【免費課程】三個你不知道的Python網頁爬蟲秘密

[立即索取](#)[閱讀完整內容](#)

關於創辦人



Mike Ku

是一位喜歡旅遊的網頁工程師，由於對Python程式語言非常有興趣，所以創辦了「Learn Code With Mike」網站，提供線上的Python「入門教學、爬蟲應用、資料分析與網頁開發」等主題的教學，透過小專案實作的方式來幫助初學者們學習Python程式語言，並且有能力開發屬於自己的應用程式。

[瀏覽簡介](#)

取得最新發佈的免費Python教學



【免費課程】三個你不知道的Python網頁爬蟲秘密

[立即索取](#)

[Pandas資料清理](#) (8)

[Python初學入門](#) (13)

[Python爬蟲教學](#) (29)

[Python物件導向](#) (6)

[Python資料結構](#) (3)

[Python實戰應用](#) (5)

[顯示更多](#)



MikeKu1116



Learn Code With Mike



learncodewithmike@gmail.com

Python推薦課程

[The Complete Python Course](#)





【免費課程】三個你不知道的Python網頁爬蟲秘密

立即索取

線上課程平台推薦



資料分析學習平台推薦



技術提供 : Blogger

COPYRIGHT © 2019-2021 BY MIKE 古耕全 @ Learn Code With Mike版權所有