

國立屏東科技大學 資訊管理系

資料科學導論

Introduction to Data Science

Week 5

講師：余榮裕



余榮裕

jungyuyu@gmail.com

屏東科技大學 資訊管理碩士

經歷：

屏東科技大學資技中心 程式設計師

馨藝資訊有限公司 技術經理

金帝盛股份有限公司 研發總監

中華全球食物股份有限公司 資深主任

屏東科技大學電算中心 程式設計師

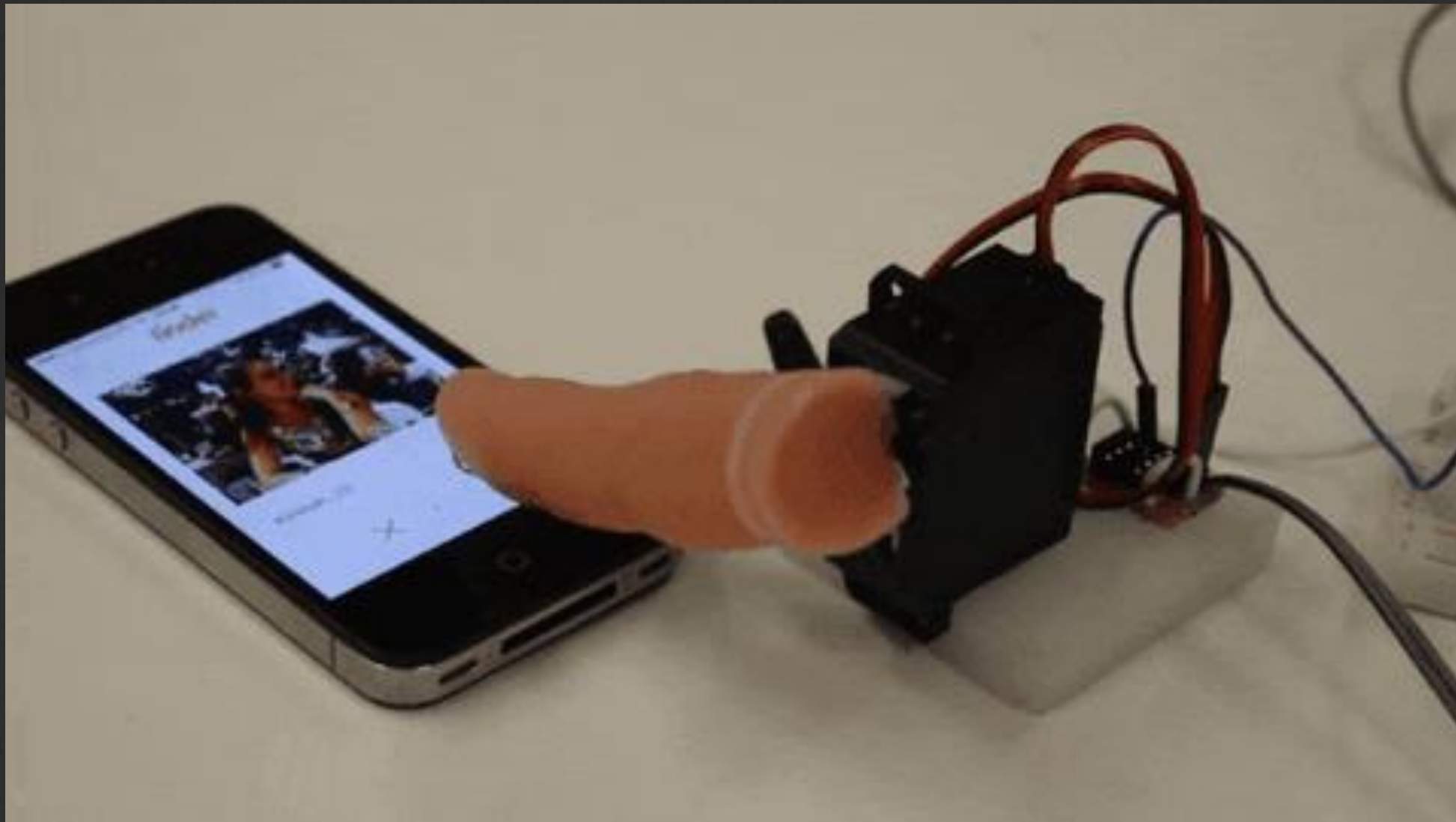
xt:Commerce GmbH 漢語站長

Python + Selenium + MySQL

Simple

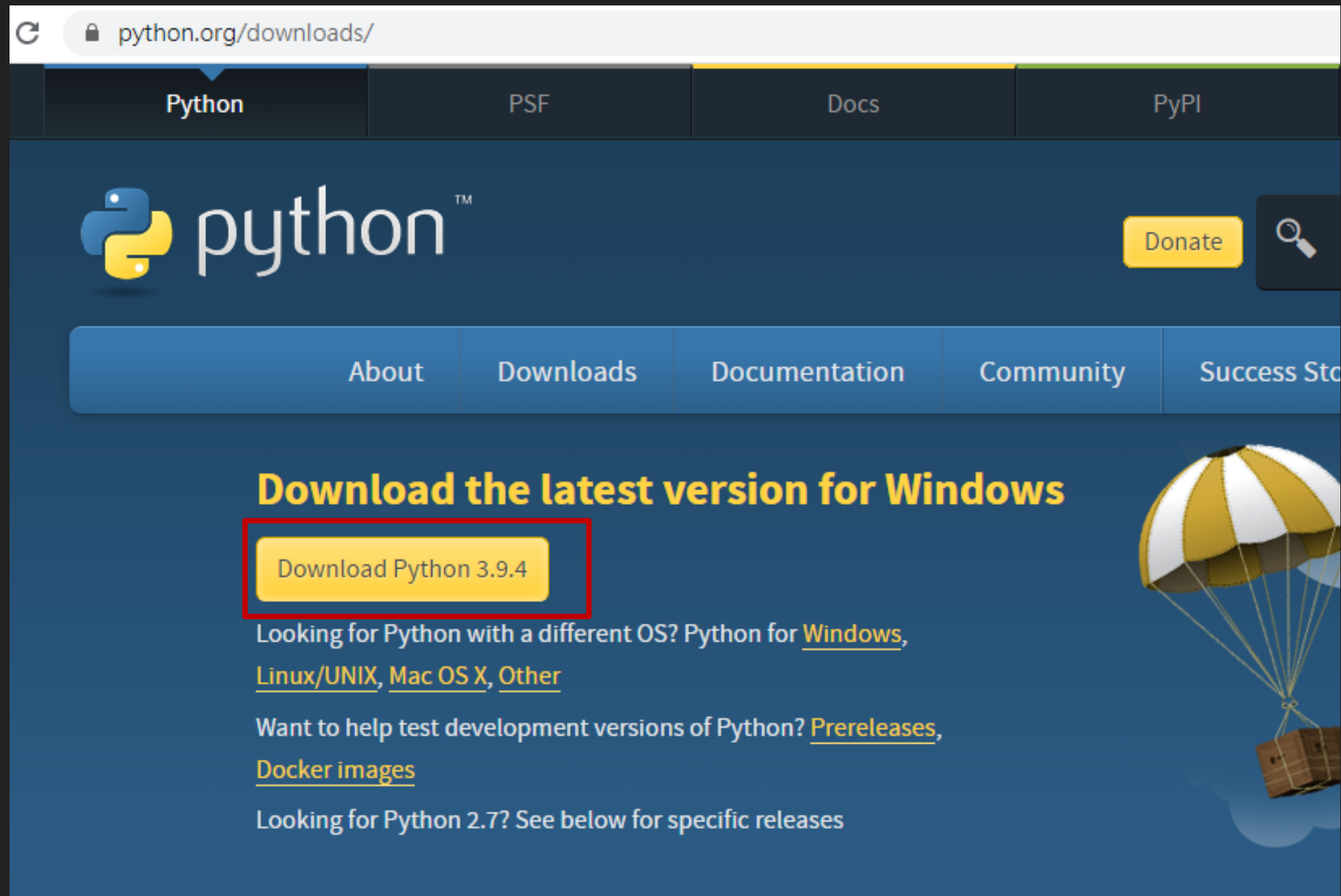
1. Python 基本安裝及路徑
2. MySQL 基本安裝
3. 安裝並使用 **Selenium** 模擬真實人類操作網頁動作
4. 介接雲端**驗證碼圖片辨識**平台，破解市面上 90% 以上圖形驗證程序
5. 練習4：整合 Selenium 由爬蟲機器人進行**自動註冊及登入**網站(排程購票)
6. 將爬取獲得資料寫入資料庫

不用偽裝，改真實的瀏覽器操作



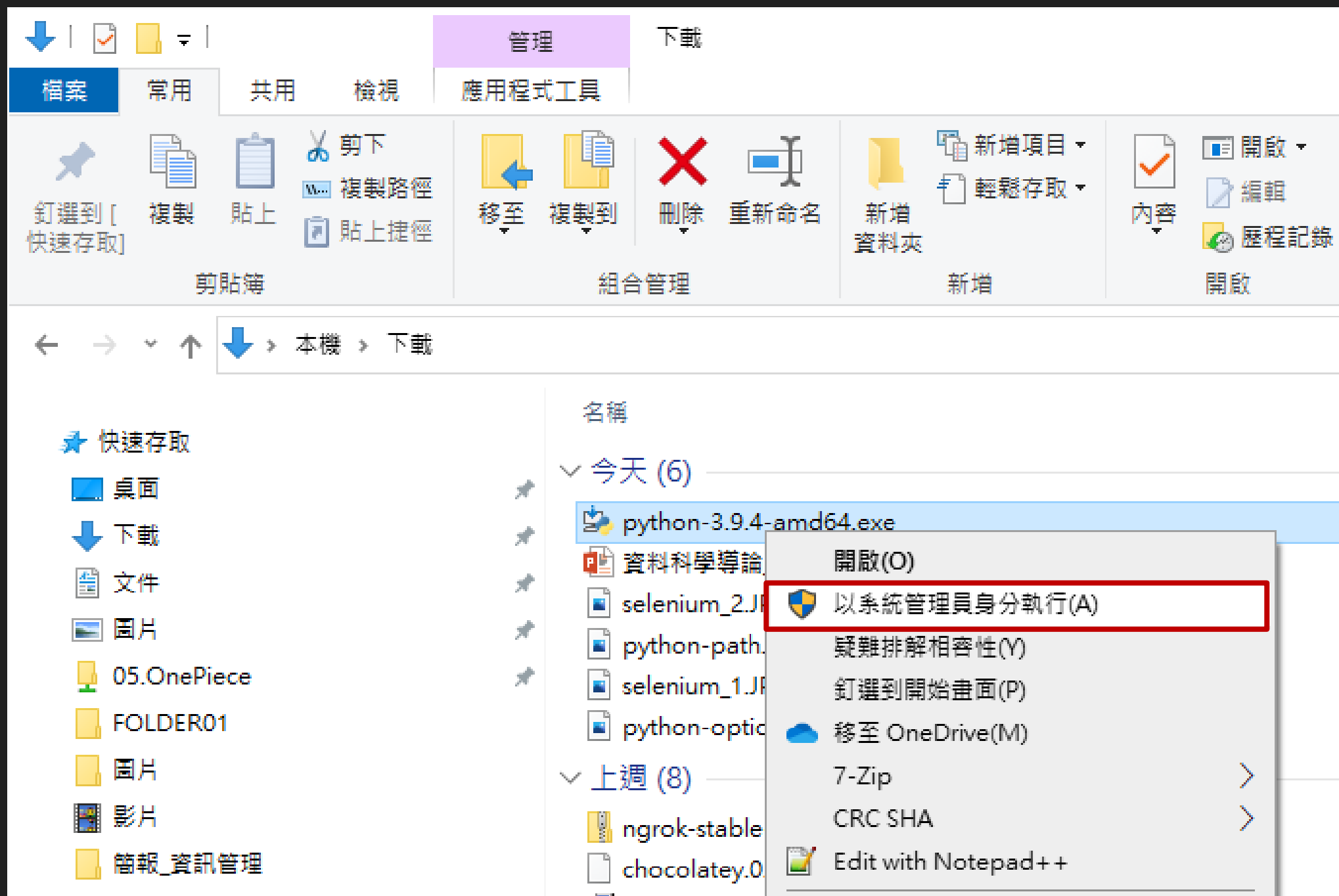
使用目前 Python 環境或全新下載

▶ 檢查安裝路徑 `python -c "import sys; print(sys.executable)"`



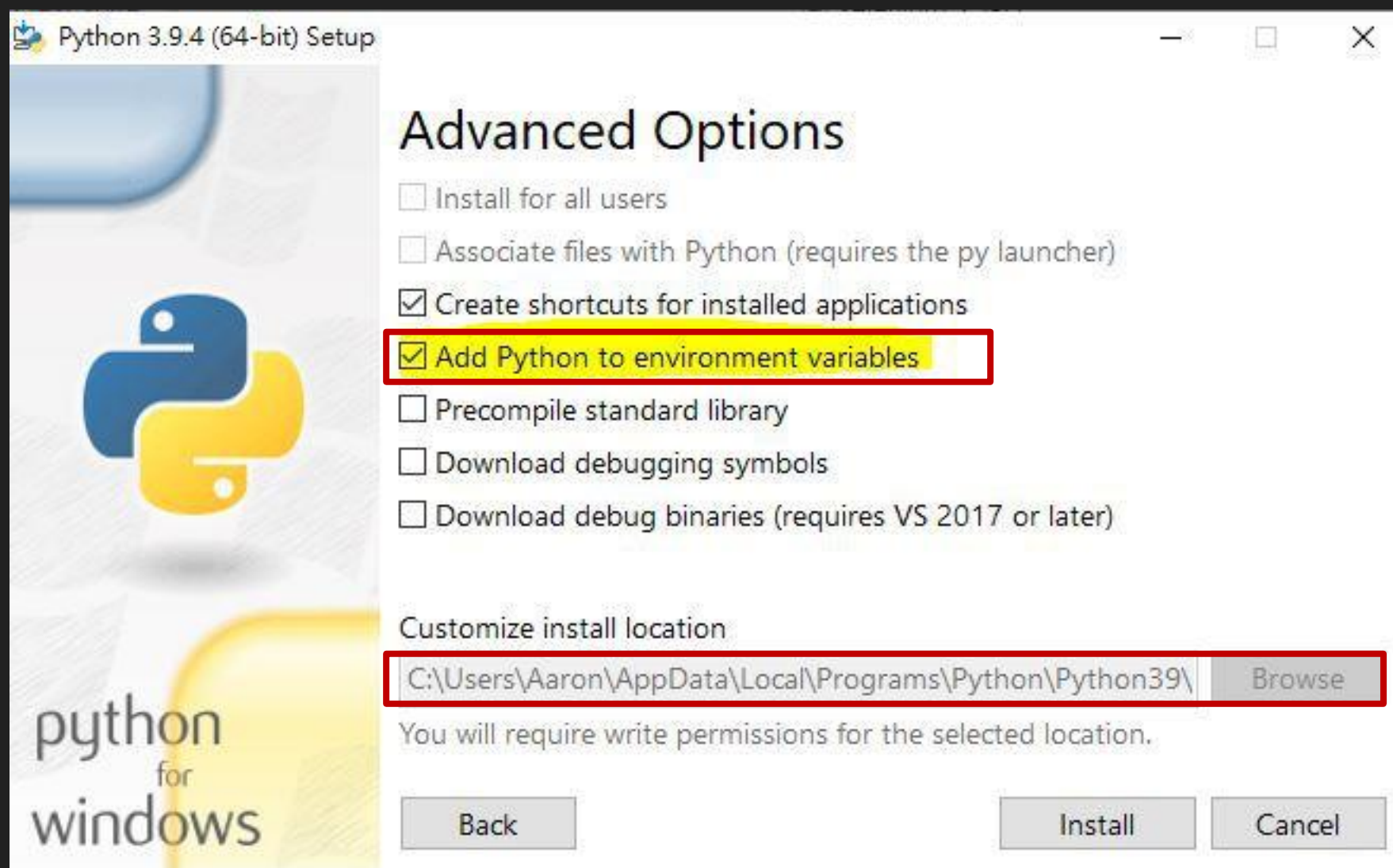
Python 基本安裝

▶ 以系統管理員身分執行



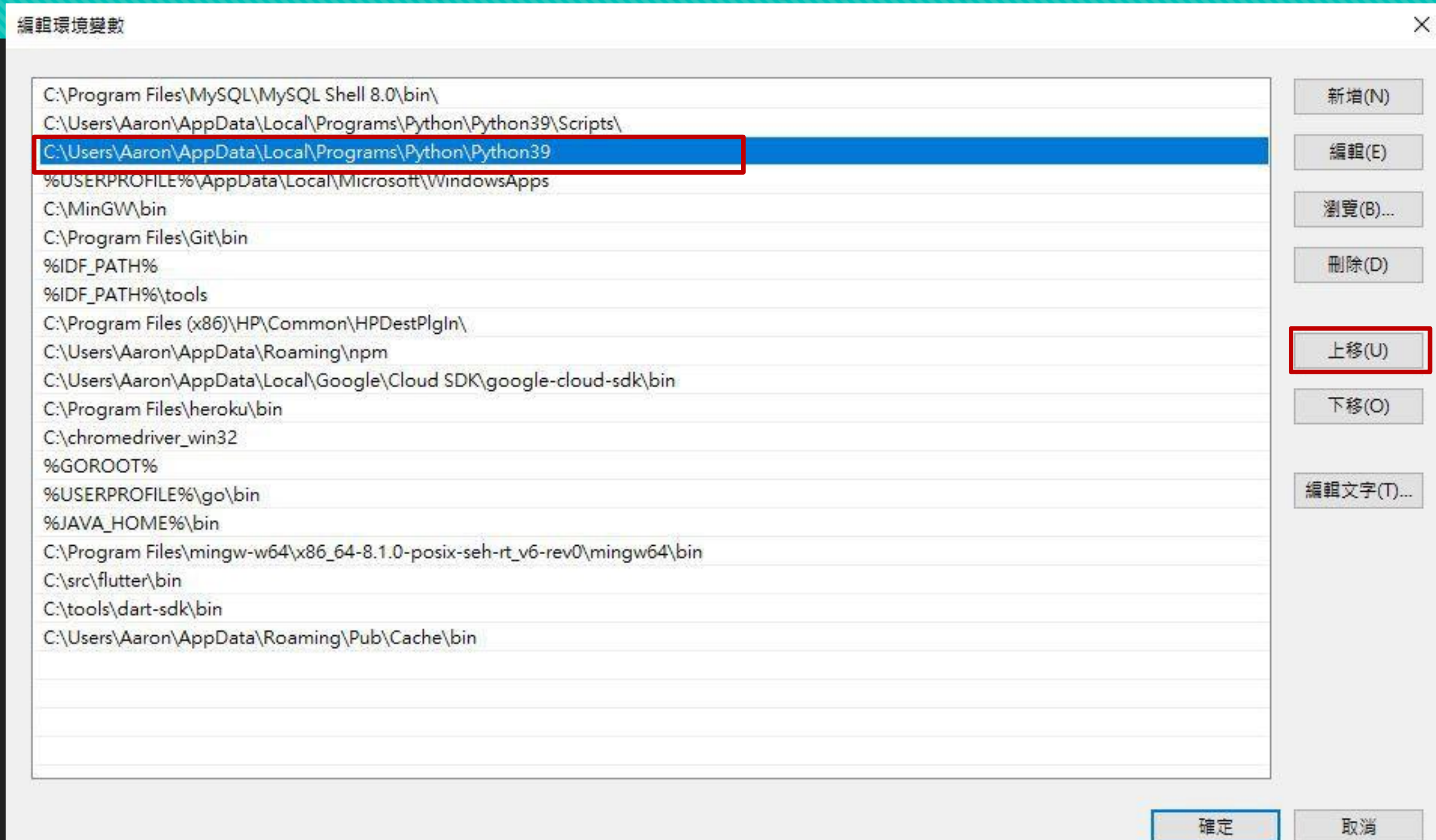
Python 基本安裝

▶ 建議路徑 C:\Python39\



Python 基本安裝

▶ 檢查安裝路徑 `python -c "import sys; print(sys.executable)"`



WebDriver：讓程式可以控制瀏覽器

▶ <https://sites.google.com/a/chromium.org/chromedriver/>

← → ↻ Chrome | chrome://settings/help ☆ 人

設定 🔍 搜尋設定

你與 Google

自動填入

安全檢查

隱私權和安全性

外觀

搜尋引擎

預設瀏覽器


起始畫面

進階 ▾

擴充功能 📄

關於 Chrome

關於 Chrome

 Google Chrome

✔ Google Chrome 目前是最新版本
版本 89.0.4389.114 (正式版本) (64 位元)

前往 Chrome 說明頁面 📄

回報問題 📄

Google Chrome

Copyright 2021 Google LLC. 保留所有權利。

Google Chrome 的開發仰賴 [Chromium](#) 開放原始碼計劃與其他開放原始碼軟體的協助才得以完成。

[服務條款](#)

WebDriver：讓程式可以控制瀏覽器

▶ <https://sites.google.com/a/chromium.org/chromedriver/>

The screenshot shows the ChromeDriver website with the following content:

ChromeDriver - WebDriver for Chrome

Search this site

Downloads

Current Releases

- If you are using Chrome version 90, please download [ChromeDriver 90.0.4430.24](#)
- If you are using Chrome version 89, please download [ChromeDriver 89.0.4389.23](#)
- If you are using Chrome version 88, please download [ChromeDriver 88.0.4324.96](#)
- If you are using Chrome version 87, please download [ChromeDriver 87.0.4280.88](#)
- For older version of Chrome, please see below for the version of ChromeDriver that supports it.

If you are using Chrome from Dev or Canary channel, please following instructions on the [ChromeDriver Canary](#) page.

For more information on selecting the right version of

Left Sidebar:

- CHROMEDRIVER
- CAPABILITIES & CHROMEOptions
- CHROME EXTENSIONS
- CHROMEDRIVER CANARY
- CONTRIBUTING
- ▼ DOWNLOADS
 - VERSION SELECTION
- ▼ GETTING STARTED
 - ANDROID
 - CHROMEOS
- ▼ LOGGING
 - PERFORMANCE LOG
- MOBILE EMULATION
- ▼ NEED HELP?
 - CHROME DOESN'T START OR CRASHES IMMEDIATELY
 - CHROMEDRIVER CRASHES

Selenium 自動化測試

▶ 除了爬蟲應用外，也可以作為網站開發之除錯、壓力工具

```
from selenium import webdriver
from selenium.webdriver.support.ui import Select
from selenium.webdriver.chrome.options import Options
from selenium.common.exceptions import NoSuchElementException
from selenium.webdriver.common.keys import Keys
```

```
options = Options()
options.binary_location = "C:\\Program Files (x86)\\Google\\Chrome\\Application\\chrome.exe"
webdriver_path = 'C:\\chromedriver_win32\\chromedriver.exe'
```

```
url = 'https://網站網址'
driver = webdriver.Chrome(executable_path=webdriver_path, options=options)
driver.get(url)
```

```
login_id = driver.find_element_by_name('login_id')
captcha_img = driver.find_element_by_id('captcha_img')
```

```
checknum = driver.find_element_by_xpath('//span[@id="checknum_pan"]/img')
checknum.screenshot('captcha_image.png')
```

```
code = driver.find_element_by_xpath('//input[@name="rand_code"]')
code.clear()
code.send_keys(captcha_code)
confirm = driver.find_element_by_xpath('//img[@name="Image2"]').click()
```



2Captcha 驗證碼辨識服務

▶ <https://2captcha.com/>



Captcha recognition service

Work for us

API

Software

Blog

 Sign in

Registration

✓ Solve Captcha



50¢

Starting from 0.5 USD for 1000 solved CAPTCHAs



API available for most popular programming languages



Av. response time: less than 12 sec

Learn more

Quick START

\$ Work on Captcha Recognition



Home data entry work



Instant payments



Easy to start

Learn more

Quick START

2Captcha 驗證碼辨識服務

▶ <https://2captcha.com/>

```
from captcha_solver import CaptchaSolver
```

```
solver = CaptchaSolver('2captcha', api_key= 'API 金鑰' )
```

```
checknum = driver.find_element_by_xpath('//span[@id="checknum_pan"]/img')
```

```
checknum.screenshot('captcha_image.png')
```

```
raw_data = open('captcha_image.png', 'rb').read()
```

```
captcha_code = solver.solve_captcha(raw_data)
```

```
code = driver.find_element_by_xpath('//input[@name="rand_code"]')
```

```
code.clear()
```

```
code.send_keys(captcha_code)
```

博客來書局：圖形驗證

▶ <https://www.books.com.tw/>



博客來

請輸入關鍵字

全館



2020
會員分級辦法
&回饋方案

立即了解



會員登入

帳號ID或email

密碼 Password

請輸入驗證碼 (不分大小寫)



[更新驗證碼](#)

登入

[忘記帳號密碼](#) | 還不是會員? [加入會員](#)

台鐵訂票 / 登入 reCaptcha

▶ <https://www.railway.gov.tw/tra-tip-web/tip/tip008/tip811/memberLogin>

會員登入

身分別

個人

企業

身分證號/會員編號

請輸入身分證字號

密碼

8-12字元，至少1個英文與1個數字



[忘記密碼？](#)



我不是機器人



reCAPTCHA
隱私權 - 條款

[切換成語音驗證](#)

登入

內政部不動產交易實價查詢

▶ <https://lvr.land.moi.gov.tw/login.action>

內政部
不動產交易實價查詢服務網



不動產買賣



不動產租賃



預售屋買賣

使用說明

首頁

聯絡資訊

區域查詢

地標查詢

*縣市區域 屏東縣 內埔鄉

*種類 ☒ 房地交易價格 ☐ 土地交易價格 *交易期間 108 年 8 月~

交易資料

第 1-200 筆

頁次 1 2 3 4 .. 20

排序：總價

區段位置或區段門牌	交易年月	總價(萬)	單價(萬/坪)	總面積(坪)	型態	屋齡	樓別樓高
1 - 桑原巷1~30號	109/8	750		60.62	透	1	全/3
2 - 福原路111巷1弄1~30號	109/8	1,580		131.49	透	13	全/4
3 - 四維路15巷1~30號	109/8	230		41.33	透	26	全/4
4 - 勝利路257巷9弄1~30號	109/8	330		39.04	透	29	全/3
5 - 壽比路241~270號	109/8	40	6.5	6.16	華	17	4/5
6 - 南華路211~240號	109/8	480		41.33	透	25	全/4
7 - 勝利路121~150號	109/8	520		51.52	透	39	全/3
8 - 桑原巷1~30號	109/8	768		60.62	透	1	全/3
9 - 昭勝路223巷61~90號	109/8	300		23.22	透	34	全/2
10 - 福泰路211~240號	109/8	350		42.85	透	25	全/3