

D37：自動化更新機制(排程)



簡報閱讀



範例與作業



問題討論

加速：非同步爬蟲



加速：非同步爬蟲

本日知識點目標



為什麼要自動化更新



自動化更新的做法



解題時間



Day 37 提升爬蟲程式執行效率

加速：非同步爬蟲



出題教練：張維元



1

本日知識點目標



本日知識點目標

- 如何定期執行爬蟲程式
- 更新時需要注意的事情是什麼？

真實世界中的資料是瞬息萬變的，也代表資料會有更新的需求。但爬蟲爬的資料只是一個片刻，所以必須要思考如何與資料源上的資料做同步或是更新，確保拿到的資料部會是錯誤或是假的。

自動化更新的做法

在迴圈中加上
Sleep

利用 threading 的
Timer

第三方套件
schedule

在迴圈中加上 Sleep

```
3 """
    每n秒執行一次
    """
    while True:
        print time.strftime('%Y-%m-%d %X',time.localtime())
        yourTask() # 此處為要執行的任務
        time.sleep(n)
```

利用 threading 的 Timer

```
1 def printHello():
2     print "Hello World"
3     t = Timer(2, printHello)
    t.start()
    if __name__ == "__main__":
        printHello()
```

第三方套件 schedule



```
3 def job():  
    print("I'm working...")  
  
    schedule.every(10).minutes.do(job)  
    schedule.every().hour.do(job)  
    schedule.every().day.at("10:30").do(job)  
    schedule.every(5).to(10).minutes.do(job)  
    schedule.every().monday.do(job)  
    schedule.every().wednesday.at("13:15").do(job)  
    schedule.every().minute.at(":17").do(job)  
  
    while True:  
        schedule.run_pending()  
        time.sleep(1)
```

解題時間



解題時間 LET'S CRACK IT

Sample Code & 作業

開始解題



下一步：閱讀範例與完成作業