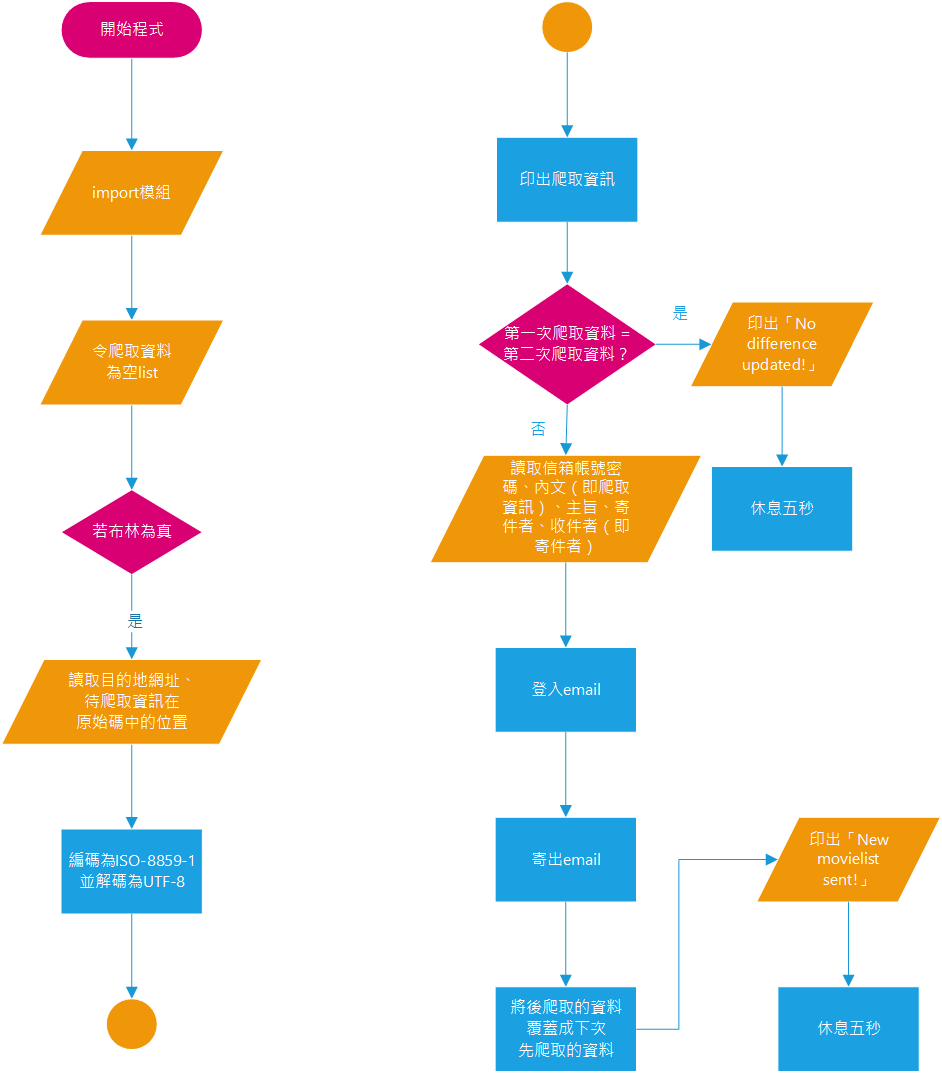
程式設計與實習(二) 期末專題報告

1072B0006 謝可安

一、程式目的摘要

電影院通常會以一週一次的頻率，在官網更新當週的上映時刻表。只要利用time模組，就可以讓程式每七天爬取一次時刻表，並且傳送email提醒自己。（註：程式中將時間間隔設定為5秒以方便觀察。）

二、流程圖

二、專題方法

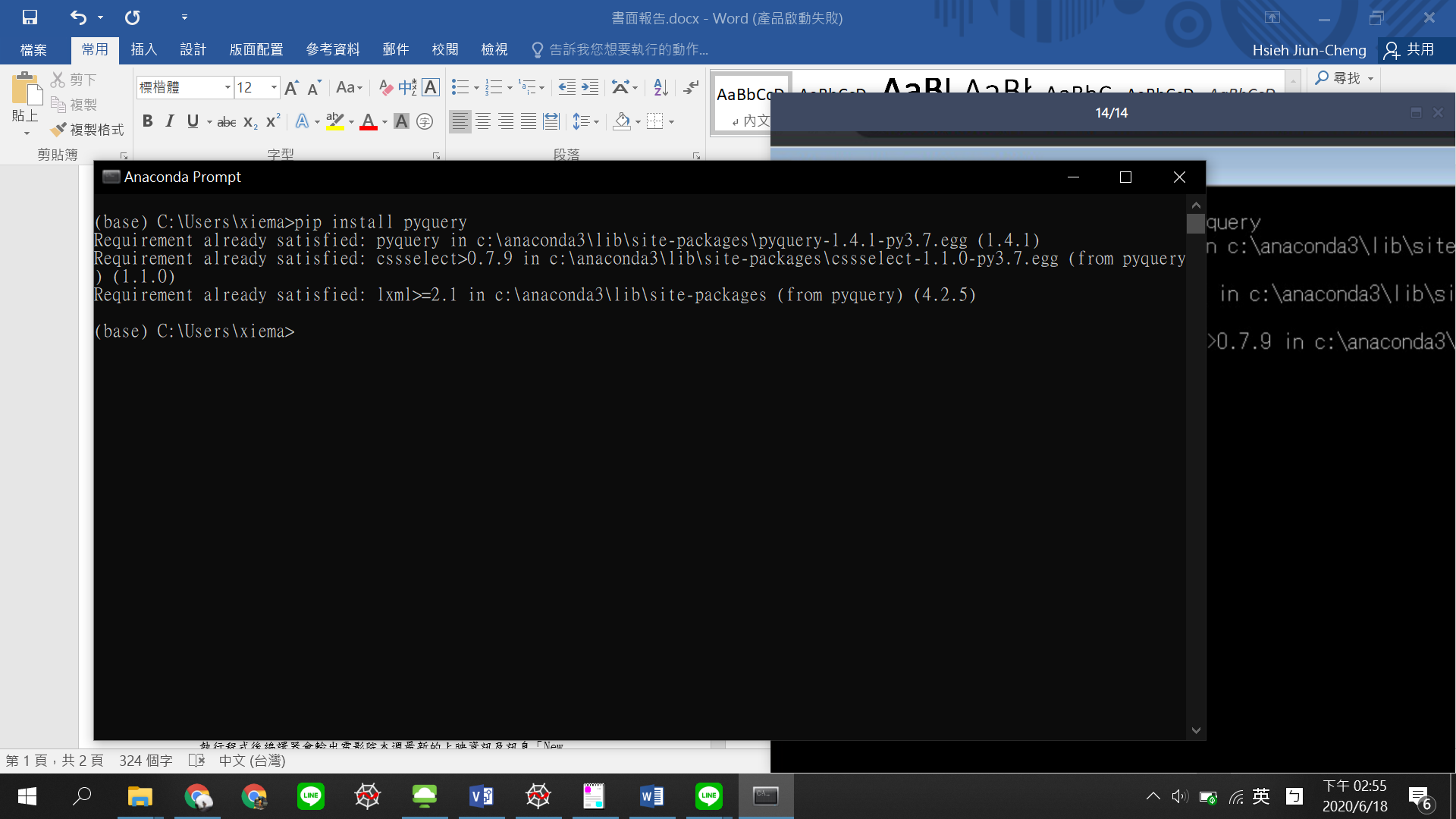
尋找範例程式→改變爬蟲目標網站→發現UNICODE顯示為亂碼→新增編碼、解碼區段→debug

使用模組：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 名稱 | pyquery | smtplib | MIMEText | time |
| 用途 | 爬蟲 | 寄email | 處理email用 | 計時 |

※需安裝PyQuery模組

安裝方式：在 Python 的命令提示字元下輸入指令pip install pyquery



*(顯示Requirement already satisfied表示已安裝過)*

※需在Google帳戶設定「**透過其他電子郵件平台查看 Gmail**」

（詳見<https://support.google.com/mail/answer/7126229?hl=zh-Hant>）



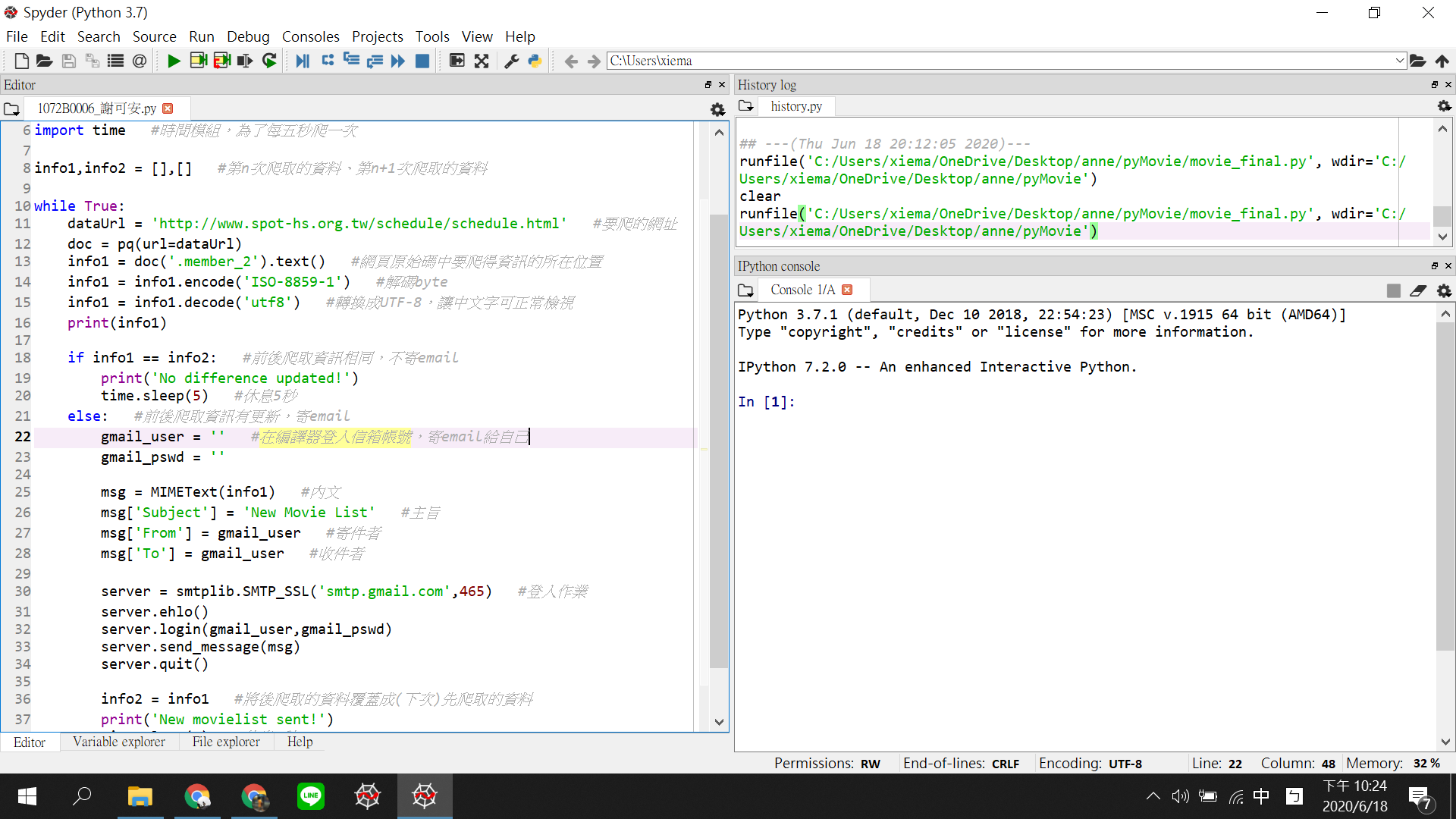
主要設定部分



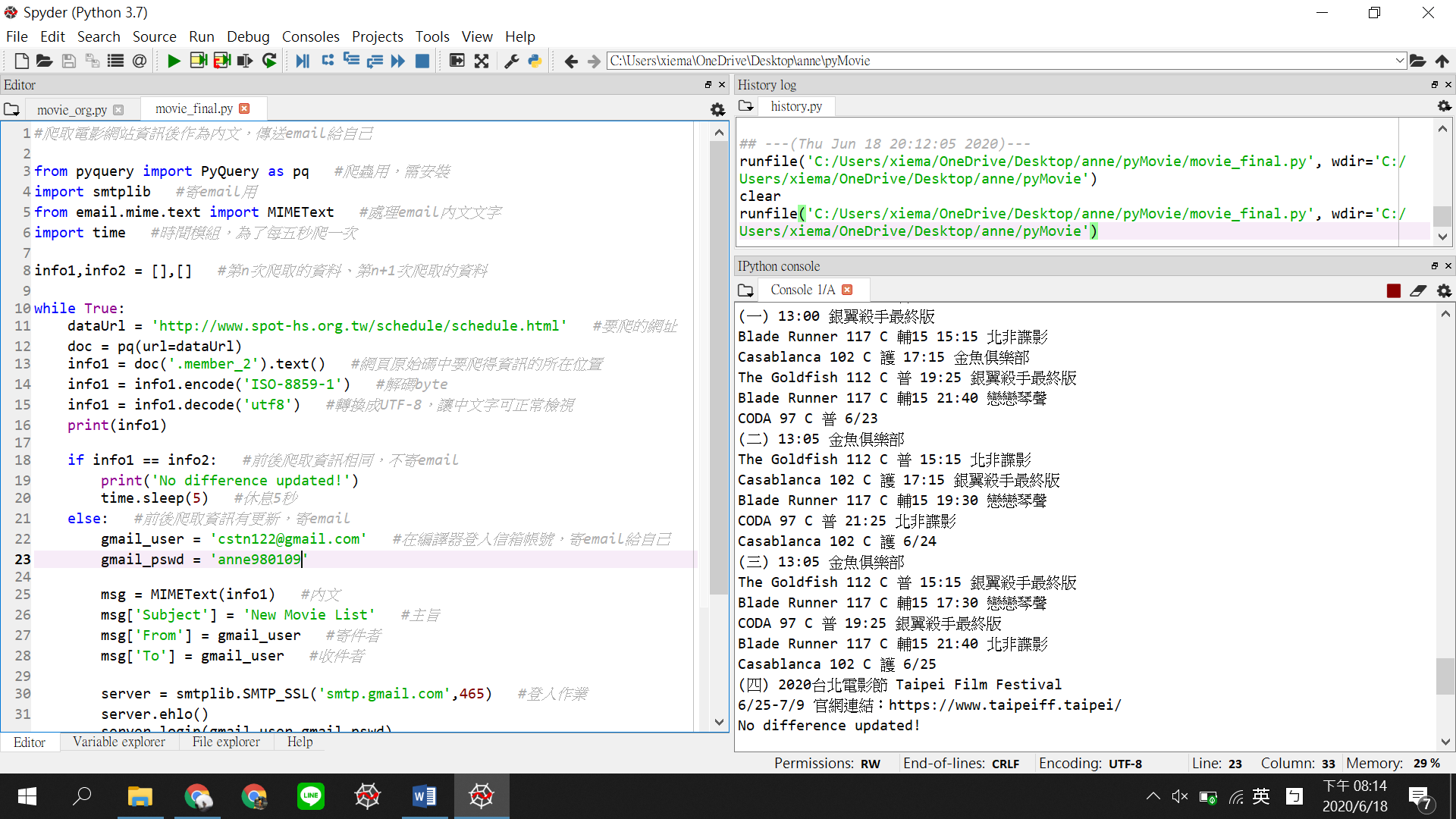
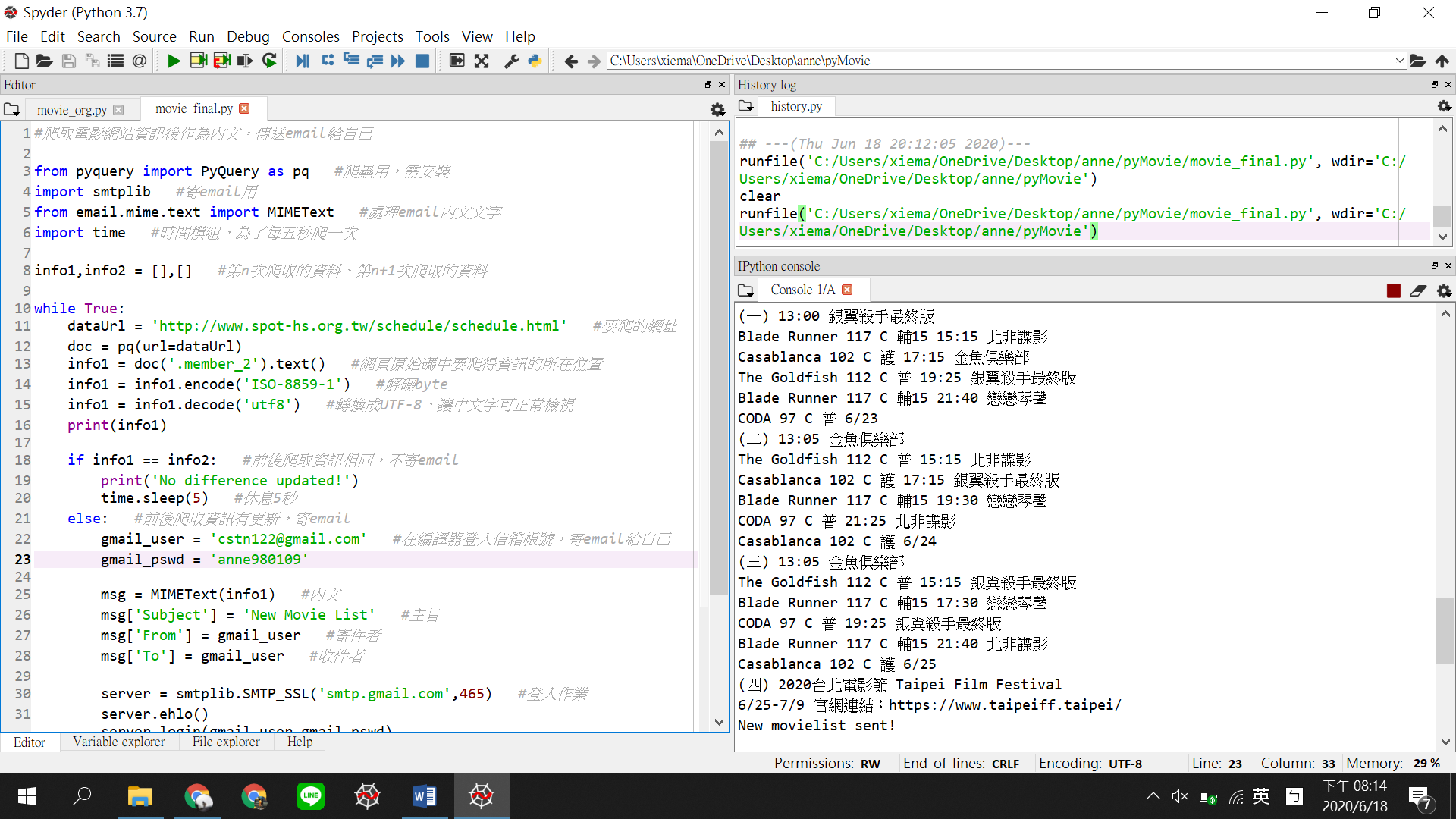
※需在程式檔中輸入自己的email帳號及密碼

gmail\_user = '信箱帳號'

gmail\_pswd = '信箱密碼'



三、結果



第一次

爬取資料

第二次

爬取資料

5

秒

後

第二次爬取，

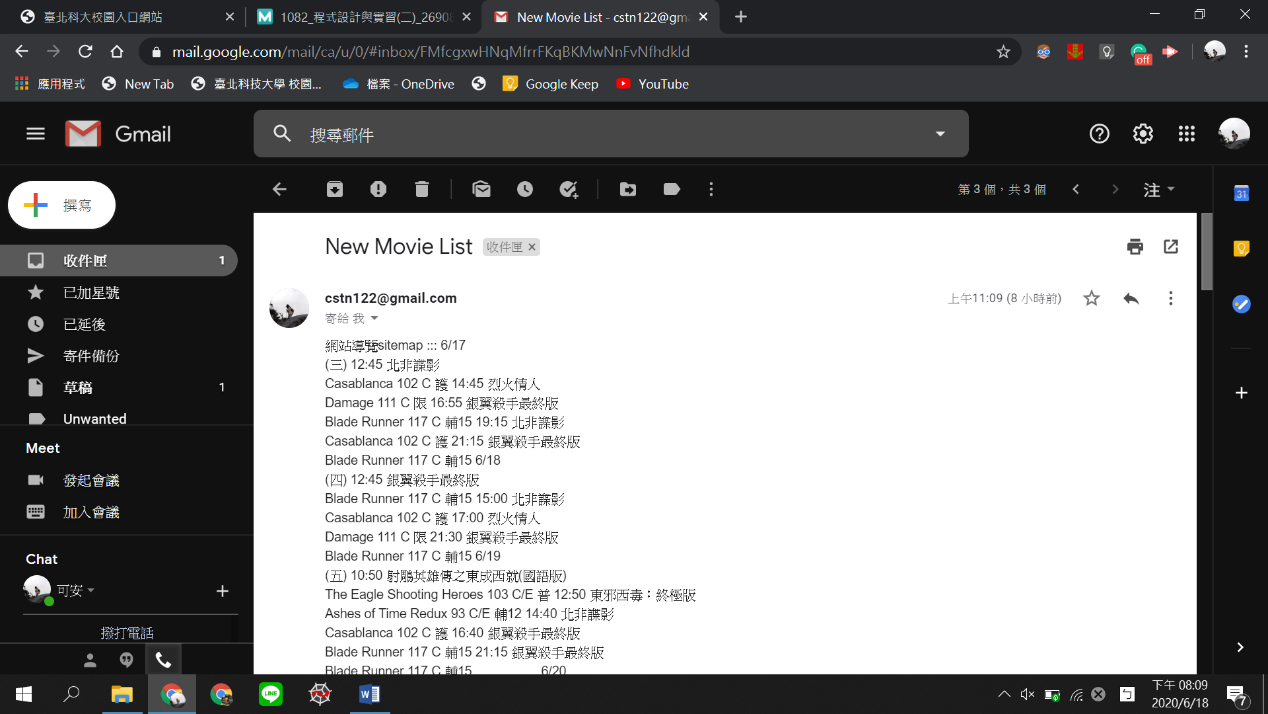
無資料更新

→無email寄出

第一次爬取，

有資料更新

→寄出email



執行程式後編譯器會輸出電影院本週最新的上映資訊及訊息「New movielist sent!」，並作為內文寄信給指定信箱。隔5秒後再執行一次，若無更新資訊則編譯器輸出「No difference updated!」（不會寄信），若有更新資訊則同上在編譯器輸出結果及寄信。

四、貢獻

改寫第11到16行，具體貢獻如下：

1. 更換爬蟲網址

2. 將ISO-8859-1解碼成UTF-8以正常顯示中文字(UNICODE)。

五、參考資料

-----------------原型--------------------

PYTHON期末專題 電影訂票

<https://www.youtube.com/watch?v=p8UeIDupR5Y&list=WL&index=2&t=0s>

-----------------爬蟲用法參考------------

pyquery 網路爬蟲

<https://flysatellite.pixnet.net/blog/post/188930692-pyquery>

爬蟲入門之爬取靜態網頁表格資料

<https://www.itread01.com/content/1542369667.html>

06.爬蟲工具pyquery用法

<https://ithelp.ithome.com.tw/articles/10203551>

-----------------SMTPLIB函式庫用法參考------------

python smtplib 練習--寄電子郵件

<https://self.jxtsai.info/2016/09/python_22.html>

-----------------亂碼問題解決參考------------

常見亂碼問題分析和總結

<https://www.itread01.com/content/1544558406.html>

python中把ISO-8859-1编码转化为UTF-8

<https://www.jianshu.com/p/e487e89c2a4e>

Python字串的encode與decode研究心得亂碼問題解決方法

<https://codertw.com/%E7%A8%8B%E5%BC%8F%E8%AA%9E%E8%A8%80/375085/>