

# Python Web Crawler Lab



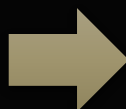
每日寄送電視台西片時刻表

# 抓取內容

HBO電影台

03-17 (三)

時間	片名
23:25	賓漢 2016 護
01:25	燈塔 護
03:10	好萊塢好好看 普
03:35	暗黑天使 輔15
05:10	蜘蛛人:新宇宙 普
07:05	班傑明的奇幻旅程 護
09:45	賓漢 2016 護
11:45	全面啟動 護
14:10	詐騙女神 護
15:40	史酷比：第十三幽靈的詛咒 普
17:00	好萊塢好好看 901 普
17:25	瞞天機密 普
19:15	魔鬼終結者3 護
21:00	託陰2：布拉姆回來了 護
22:25	歌喉讚 護
00:15	怒海劫 護
02:25	班傑明的奇幻旅程 護
05:05	瞞天機密 普
時間	片名



2021-03-17電視台洋片時刻表.csv

選擇開啟工具 ▾

A	B	C	D	E
頻道	播出時間	結束時間	電影片名	電影分級
HBO電影台	23:25	1:25	賓漢 2016	護
HBO電影台	1:25	3:10	燈塔	護
HBO電影台	3:10	3:35	好萊塢好好看	普
HBO電影台	3:35	5:10	暗黑天使	輔15
HBO電影台	5:10	7:05	蜘蛛人:新宇宙	普
HBO電影台	7:05	9:45	班傑明的奇幻旅程	護
HBO電影台	9:45	11:45	賓漢 2016	護
HBO電影台	11:45	14:10	全面啟動	護
HBO電影台	14:10	15:40	詐騙女神	護
HBO電影台	15:40	17:00	史酷比：第十三幽靈的詛咒	普
HBO電影台	17:00	17:25	好萊塢好好看	普
HBO電影台	17:25	19:15	瞞天機密	普
HBO電影台	19:15	21:00	魔鬼終結者3	護
HBO電影台	21:00	22:25	託陰2：布拉姆回來了	護
HBO電影台	22:25	0:15	歌喉讚	護
HBO電影台	0:15	2:25	怒海劫	護
東森洋片台	23:15	2:00	駭客任務完結篇：最後武士	護
東森洋片台	2:00	6:00	甜蜜冤家	普
東森洋片台	6:00	6:30	洋片幕後趣	護

# 使用套件

+ BeautifulSoup

+ Requests

+ Pandas

+ Time

+ os

+ email.mime.multipart

+ email.mime.text

+ email.mime.application

+ smtplib

+ schedule

# 寄送email

```
from email.mime.multipart import MIMEMultipart
from email.mime.text import MIMEText
from email.mime.application import MIMEApplication
import smtplib

content = MIMEMultipart()
content[ "subject" ] = "信箱標題" #標題
content[ "from" ] = "寄件人信箱" #寄件人
content[ "to" ] = "收件人信箱" #收件人
content.attach(MIMEText( "信箱內文" )) #內容
pdfload = MIMEApplication(open(filename, 'rb' ).read())
pdfload.add_header( "Content-Disposition" , "attachment" ,filename = filename)
content.attach(pdfload)
with smtplib.SMTP(host = "smtp.gmail.com" , port= "587" ) as smtp: #gmail的smtp和port
    try:
        smtp.ehlo() # 驗證SMTP伺服器
        smtp.starttls() # 建立加密傳輸
        smtp.login( "寄件人信箱" ,my_password)
        smtp.send_message(content)
        print("傳送成功")
    except Exception as e:
        print("傳送失敗,Error message:",e)
```



# 定時排程

+ 設定每天的凌晨1點執行程式

```
import schedule
```

```
def job():
```

```
    #程式碼(網頁爬蟲、建立csv檔、傳送email並附件csv檔)
```

```
schedule.every().day.at("01:00").do(job)
```

# 運作模式



**Every day at 01:00**

**Run the program**



**CSV**



**File**



**Sending mail**



# 問題點

- + 設定低安全度
- + 顯示密碼

## 低安全性應用程式存取權

由於您允許登入技術安全性較低的應用程式和裝置存取您的帳戶，因此您的帳戶可能較容易受到侵害。如果您並未使用這項設定，Google 會自動關閉該權限，以維護您的帳戶安全。



開啟

關閉存取權 (建議)

# 問題點

- + 設定低安全度
- + 顯示密碼

因有填入信箱密碼，  
而設定環境變數

```
# 設定環境變數，不顯示密碼  
my_password = os.environ.get('PASSWORD')
```





# 呈現結果 Gmail

☐ ☆ 我

03-22電視台洋片時刻表 - 附件：今日電視台洋片時刻表2021-03-22 01:00:09 頻道:HBO電影台日期:03-...

上午1:00

 2021-03-22電視...

☐ ☆ 我

03-21電視台洋片時刻表 - 附件：今日電視台洋片時刻表2021-03-21 01:00:11 頻道:HBO電影台日期:03-...

3月21日

 2021-03-21電視...

☐ ☆ 我

03-20電視台洋片時刻表 - 附件：今日電視台洋片時刻表2021-03-20 01:00:10 頻道:HBO電影台日期:03-...

3月20日

 2021-03-20電視...

☐ ☆ 我

03-19電視台洋片時刻表 - 附件：今日電視台洋片時刻表2021-03-19 01:00:08 頻道:HBO電影台日期:03-...

3月19日

 2021-03-19電視...

☐ ☆ 我

03-18電視台洋片時刻表 - 附件：今日電視台洋片時刻表2021-03-18 17:52:23 頻道:HBO電影台日期:03-...

3月18日

 2021-03-18電視...



# Thanks!

My GitHub



Click here