



爬蟲專題分享

分析

- 網址: <http://www.joyglobal.cc/joyday1.asp>
- 原始碼分析: 經分析後發現所需要的資訊原始碼內有, 用html分析即可

1. 一開始登入狀態

免費算命

免費今、明日 / 本月運勢

姓名 生辰 年 月 日 (陰曆閏月)

☐ 男生 / ☒ 女生

☒ 陽曆 / ☐ 陰曆

系統時間: 21:00-22:59

正確的生辰也有可能排出錯誤的紫微斗數命盤，
運算前請先確認您的真太陽時生辰。

資料正確，送出

範例

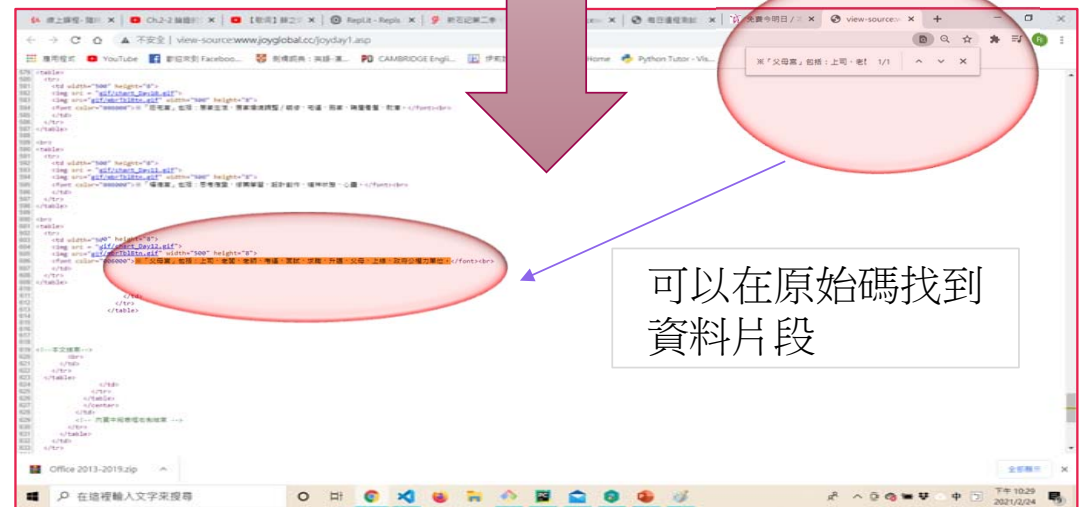
1. 女命
2. 姓名: 李百合
3. 生辰: 陽曆 1981年7月16日卯時
農曆 1981年6月15日卯時

【Part.1 今日運勢】

李百合 2011年1月7日運勢吉凶圖



2. 選取網頁資訊片段



分析(續)

- 程式原始碼雖然可以找到片段，但是我們需要在表格填入個人資訊並且點及下一頁，我需要通过一步一步的點擊表格和操作，所以我需要利用python中的selenium函式庫的功能去模擬表格的操作輸入，最後在利用其內建的函示庫去做網頁分析。

免費今、明日／本月運勢

姓名 生辰 年
☒ 男生 / ☐ 女生
☒ 陽曆 / ☐ 陰曆
 月 (☐ 陰曆閏月)
 日

⚠️ 正確的生辰也有可能排出錯誤的紫微斗數命盤，
運算前請先確認您的真太陽時生辰。

表單填寫和送出



分析(續)

【明日運勢】



- 所得的重要資訊都為圖檔，預計將把圖片檔下載下來，



程式編寫

抓取資料

- 利用SELENIUM, 下拉式選單,

存取資料

- 用月份建立資料夾且用當天日期存取圖片

寄件

- 寄送圖片和內文

排程

- 每天早上06:00寄出


```
def save_my_pic(url):
    import requests
    from datetime import datetime as d
    import os
    import shutil

    # https://www.itread01.com/content/1545450012.html
    my_dir = f"my_{{d.today().month}}_month_fortune"
    my_pic_dir = f"./{my_dir}/my_day_{{d.date(d.today())}}.jpg"

    if not os.path.exists(my_dir):
        os.mkdir(my_dir)
    # urlretrieve(url, f"./{my_dir}/my_day_{{d.date(d.today())}}.jpg")

    response = requests.get(url, stream=True, verify=False)
    with open(my_pic_dir, "wb") as f:
        # f.write(response.raw.read())
        shutil.copyfileobj(response.raw, f)
    return my_pic_dir
```

儲存圖片



```
def send_mail(my_pic_dir):
    import os
    from datetime import datetime as d
    import smtplib
    from email.mime.multipart import MIMEMultipart # email內容載體
    from email.mime.text import MIMEText # 用於製作文字內容
    from email.mime.image import MIMEImage

    account = "rasptest@gmail.com"

    # 環境變量設定https://cuiqingcai.com/8947.html
    # https://www.youtube.com/watch?v=IolxqkL7cD8
    password = "
    recepoint = "rasptest@gmail.com"

    mine = MIMEMultipart()
    mine["Content-Type"] = "image/jpeg"
    mine["Content-Disposition"] = f'attachment; filename="{my_pic_dir}"' # 寫你的檔案名讓他可以找到
    mine["Subject"] = f"這是{{d.date(d.today())}}的運勢" # 撰寫郵件標題
    mine["From"] = f"Rasper[{{account}}]" # 撰寫你的暱稱或是信箱
    mine["To"] = f"Rasper[{{recepoint}}]" # 撰寫你要寄的人
    # mime["Cc"] = "@gmail.com, @gmail.com" # 寫本收件人
```

寄發郵件

查找運勢



```
main.py × practice.py × function.py
22 # import re
23
24 def search_my_fortune():
25     from selenium import webdriver
26     import time
27     from selenium.webdriver.support.ui import Select
28     from selenium.webdriver import ActionChains
29     import pyautogui
30
31     driver = webdriver.Chrome("chromedriver")
32     driver.get("http://www.joyglobal.cc/joyday1.asp")
33     time.sleep(1)
34
35     name = driver.find_element_by_name("Name_1")
36     name.send_keys("王大明")
37     time.sleep(1)
38
39     gender = driver.find_element_by_css_selector('input[type="radio"][value="1"]')
40     gender.click()
41     time.sleep(1)
42
43     calender = driver.find_elements_by_name("Option1_1")[0]
```

主程式

```
main.py x errorhandler.py x practice.py x function.py x
1 import time
2 import schedule
3 from function import search_my_fortune, save_my_pic, send_mail
4 from practice import find_my_fortune
5
6 #https://schedule.readthedocs.io/en/stable/examples.html
7
8 def job():
9     my_url = search_my_fortune()
10    my_url = find_my_fortune()
11
12    path = save_my_pic(my_url)
13
14    send_mail(path)
15
16    schedule.every(5).seconds.do(job)
17
18 while True:
19     schedule.run_pending()
20     time.sleep(5)
21
```

引入自訂函數'

定義函數

主函數

找尋所有需要元素節點

免費今、明日／本月運勢

姓名 1

2 ☐ 男生 / ☒ 女生
3 ☒ 陽曆 / ☐ 陰曆

生辰 [2021 ▾] 年 4

3 ▾ 月 (☐ 陰曆閏月)
5 25 ▾ 日 丑時(01:00~02:59) ▾ 6

正確的生辰也有可能排出錯誤的紫微斗數命盤，
運算前請先確認您的真太陽時生辰。

資料正確，送出

範 例

1. 女命

2. 姓名：李百合

3. 生辰：陽曆 1981年7月16日卯時
 農曆 1981年6月15日卯時

【Part. 1 今日運勢】

李百合
2011年1月7日運勢吉凶圖

運勢 (Fortune)	命宮 (Destiny Palace)	兄弟宮 (Siblings Palace)	夫妻宮 (Spouse Palace)	子女宮 (Children Palace)	財帛宮 (Wealth Palace)	疾厄宮 (Illness Palace)	遷移宮 (Migration Palace)	奴僕宮 (Servants Palace)	官祿宮 (Official Rank Palace)	田宅宮 (Real Estate Palace)	福德宮 (Fortune Palace)	父母宮 (Parents Palace)
大凶 -100	命宮	兄弟宮	夫妻宮	子女宮	財帛宮	疾厄宮	遷移宮	奴僕宮	官祿宮	田宅宮	福德宮	父母宮

[illegible]


```
from selenium.webdriver.support.ui import Select
```

```
select_year = Select(driver.find_element_by_name("IYear_1"))
year = select_year.select_by_visible_text("1991")

time.sleep(1)

select_month = Select(driver.find_element_by_name("IMonth_1"))
month = select_month.select_by_index(5)

time.sleep(1)

select_date = Select(driver.find_element_by_name("IDay_1"))
date = select_date.select_by_visible_text("21")

time.sleep(1)

select_hour = Select(driver.find_element_by_name("IHour_1"))
hour = select_hour.select_by_index(4)

time.sleep(2)

submit = driver.find_element_by_name("M10a")
submit.click()

time.sleep(2)
```

免費今、明日／本月運勢

姓名

生辰 2021 ▾ 年

☐ 男生 / ☒ 女生

3 ▾ 月 (☐ 陰曆閏月)

☒ 陽曆 / ☐ 陰曆

23 ▾ 日

丑時(01:00~02:59) ▾

👤 正確的生辰也有可能排出錯誤的紫微斗數命盤，

運算前請先確認您的真太陽時生辰。

資料正確，送出

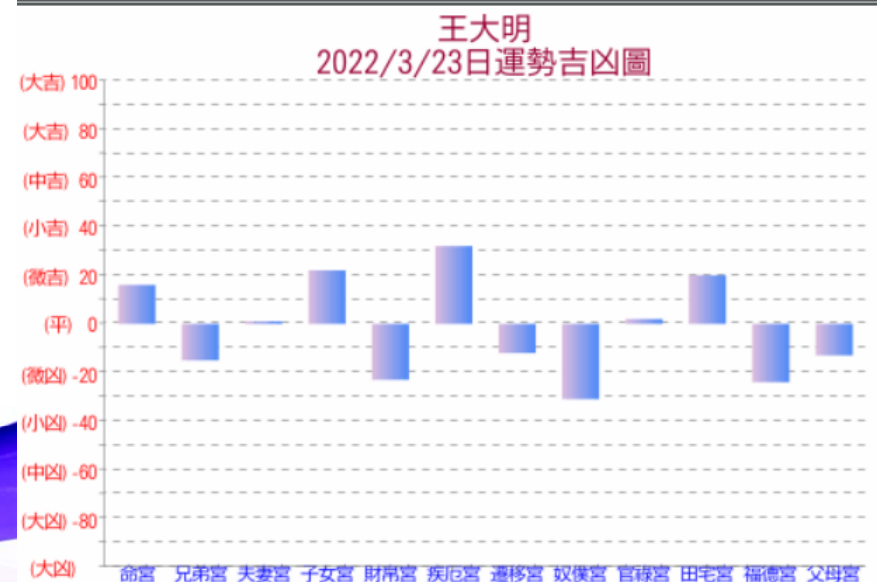
```
imgs = driver.find_elements_by_css_selector(".s2 img")

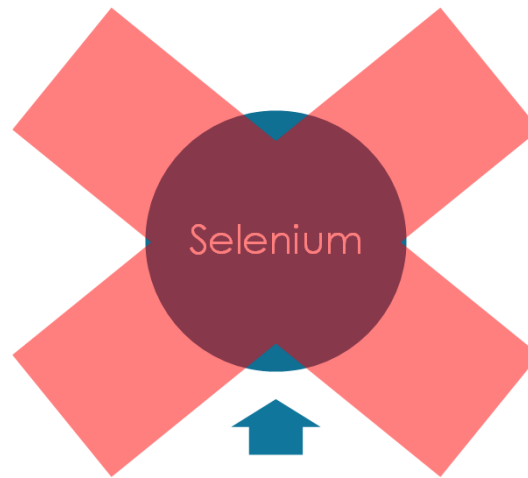
pic_1 = imgs[0].get_attribute("src")
print(pic_1)
# response = requests.get(pic_1, headers=headers, stream=True,
# verify=False)
# print(response.raw.read())

driver.quit()
```

得到了網址，點開來檢查看看吧

預期狀況





既然Selenium不能抓取，那換另一種方式呢



求取目前網址

取得html碼

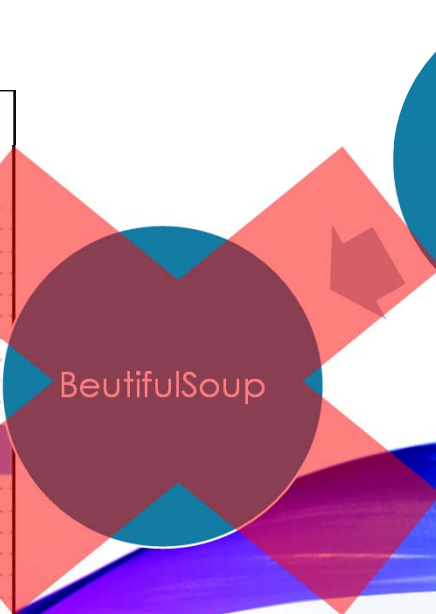
```
92
93 response = requests.post(url=url, data=post_1, headers=headers)
94 print(response)
95
96
97 html = BeautifulSoup(response.text)
98 # html = BeautifulSoup(driver.page_source, "html.parser")
99
100 td = html.find("td", class_="s2")
101 imgs = td.find_all("img")
102
103
104 # imgs = driver.find_elements_by_css_selector(".s2 img")
105
```

another_way_1.py

```
62
63 url = driver.current_url
64
65 post_1 = {
66     'buy': '1',
67     'joyday': 'DAYgbrwoth4',
68     'Name_1': '王大明',
69     'OptGdr_1': '1',
70     'Option1_1': '3',
71     'IYear_1': '1991',
72     'IMonth_1': '6',
73     'IDay_1': '21',
74     'IHour_1': '辰時(07:00~08:59)',
75     'M10a': '資料正確，送出',
76 }
77
78 headers={
79     'Accept': 'text/html,application/xhtml+xml,application/xml;q=0.9,
80     image/avif,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;
81     v=b3;q=0.9',
82     'Accept-Encoding': 'gzip, deflate',
83     'Accept-Language': 'zh-TW,zh;q=0.9,en-US;q=0.8,en;q=0.7',
84     'Cache-Control': 'max-age=0',
85     'Connection': 'keep-alive',
86     'Content-Length': '235',
87     'Content-Type': 'application/x-www-form-urlencoded',
88     'Cookie': 'ASPSESSIONIDQARSDCRR=CBNDLGKDNMIJHMIJEDJNDHAE;
89     gads=ID=18d236e627e6739c-2272f3ae6dc60077:T=1615868890:RT=1615868890:S=
90     ALNI_MavgjY8bhr3p1fRcFCptsAJ8AysHg',
91     'Host': 'www.joyglobal.cc',
92     'Origin': 'http://www.joyglobal.cc',
93     'Referer': 'http://www.joyglobal.cc/joyday1.asp',
94     'Upgrade-Insecure-Requests': '1',
95     'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64)
96     AppleWebKit/537.36 (KHTML, like Gecko) Chrome/89.0.4389.82
97     Safari/537.36'
```


The Selenium logo, which consists of a dark blue circle with the word "Selenium" in white text. The logo is centered on a background of four overlapping, semi-transparent squares in shades of blue and green, arranged in a cross pattern.

```
C:\Users\Casper\PycharmProjects\fortune\practice.py:90: GssuedAtParserWarning: No parser was explicitly specified, so I'm using the best available HTML parser for this system ("h..."). This usually isn't a problem, but if you run this code on ano...
```

[illegible]

BeutifulSoup

滑鼠右鍵，開 新分頁

```
from selenium.webdriver import ActionChains
import pyautogui
```

引入兩個函式模組，用Selenium模擬滑鼠行為

```
73     imgs = driver.find_elements_by_css_selector(".s2 img")
74     pic_1 = imgs[0]
75
76     action = ActionChains(driver).move_to_element(pic_1)
77     action.context_click(pic_1).perform()
78     time.sleep(2)
79
80     # pyautogui
81     pyautogui.typewrite(['down', 'enter'], interval=1)
82
83     # https://ithelp.ithome.com.tw/articles/10230717
84     driver.switch_to.window(driver.window_handles[1])
85     # driver.switch_to.window(driver.window_handles[0])
86
87     url = driver.current_url
88     print(url)
89     driver.quit()
90     return url
91
92
93 def save_my_pic(url):
94     import requests
95     search_my_fortune()
```

在瀏覽器下操作滑鼠，把滑鼠移動到想要的圖片

在圖片點選右鍵，執行

向下，按下enter鍵

切換到新視窗

獲取當前網址

• • •

<https://chart.apis.google.com/chart?cht=bvs&chf=h&l=0.0.DEBDE.0.4189P.1%&cnc=C6D9FD&whhh=a.15.5&chs=0.5&nbs=-100.100&chs=500x330&xht=x.v&yhs=0.0000ff.FF0000&chl=0.1E5E01E7.00AEE77C4B5E5E24.1E5AEAB7C7E5E44ABE5A6BBE5AEAE77>

```
C:\Users\Casner\PycharmProjects\fortune\practice.py:90: GussedAtParserWarning: No parser was explicitly specified, so I'm using the first one I can find. This usually isn't a prob...
```

28% E5% A4% A7% E5% 87% B6% 29 % 7C% 7C% 28% E5% A4% A7% E5% 87% B6% 29 % 2D80% 7C%

抓取圖片網址

放棄

BeautifulSoup

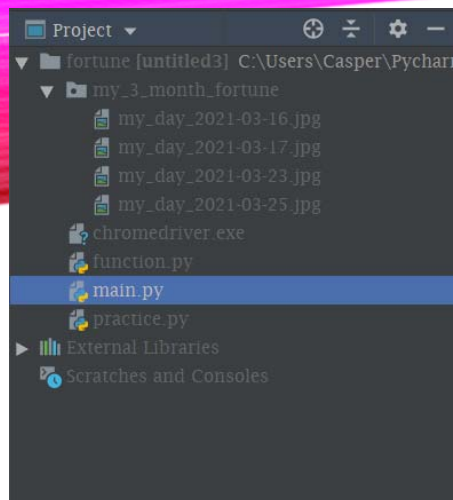
既然沒辦法下載到好圖片，那就把得到的結果下載起來也好

```
def save_my_pic(url):  
    import requests  
    from datetime import datetime as d  
    import os  
    import shutil  
  
    # https://www.itread01.com/content/1545450012.html  
    my_dir = f"my_{{(d.today().month)}}_month_fortune"  
    my_pic_dir = f"./{my_dir}/my_day_{{d.date(d.today())}}.jpg"  
  
    if not os.path.exists(my_dir):  
        os.mkdir(my_dir)  
    # urlretrieve(url, f"./{my_dir}/my_day_{{d.date(d.today())}}.jpg")  
  
    response = requests.get(url, stream=True, verify=False)  
    with open(my_pic_dir, "wb") as f:  
        # f.write(response.raw.read())  
        shutil.copyfileobj(response.raw, f)  
    return my_pic_dir
```

引入時間函數

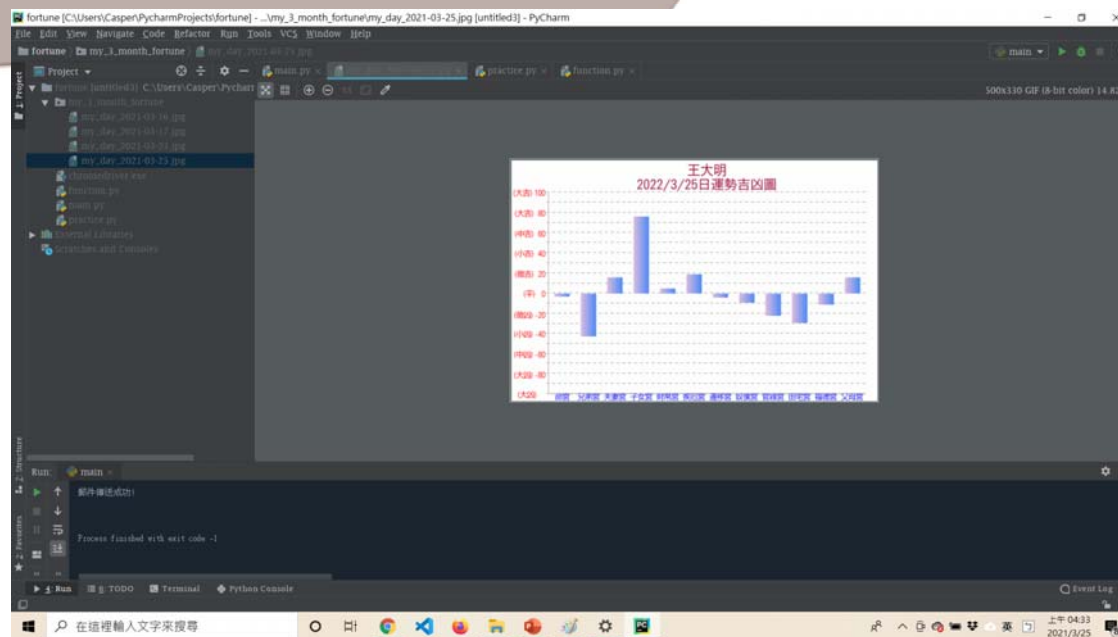
用月份來建立資料夾

用日期來儲存圖片

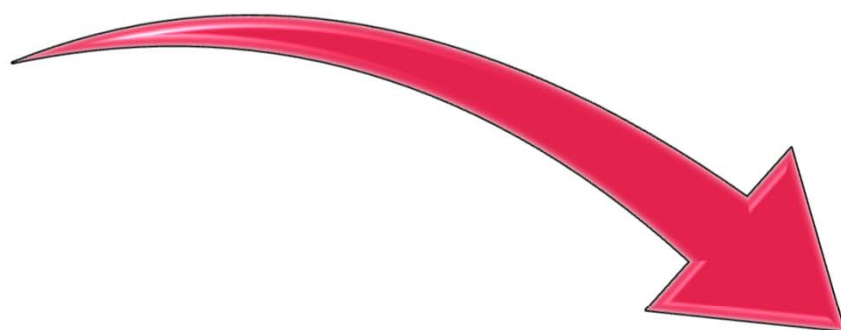


檢查圖片結果發現

下載完畢，建立資料夾和圖片下載



後來詢問老師後，結果發現因為直接點進去的圖像不會帶入header，所以會失敗，所以我只要下載我得到的網址就好那我其實這樣寫就好啦

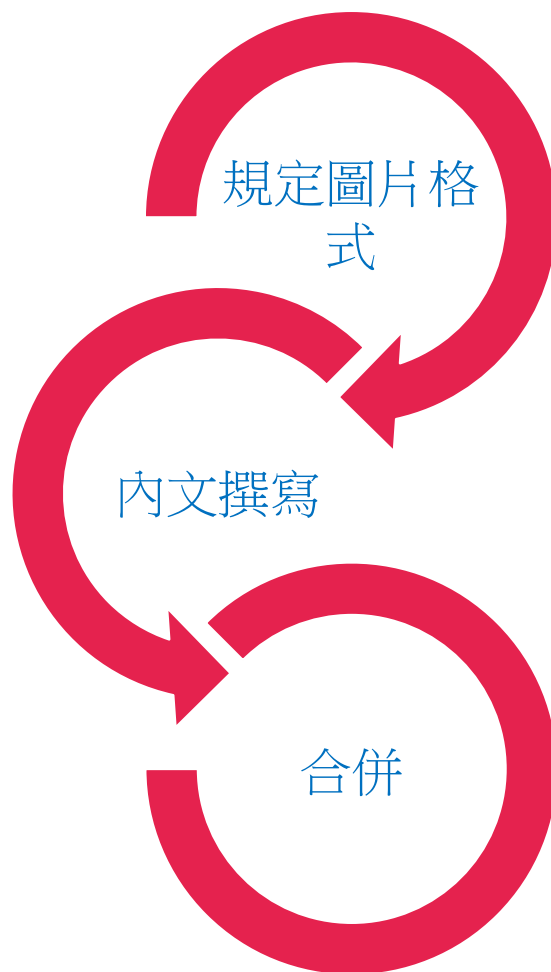


```
imgs = driver.find_elements_by_css_selector(".s2 img")

pic_1 = imgs[0].get_attribute("src")
print(pic_1)
# response = requests.get(pic_1, headers=headers,
# stream=True, verify=False)
# print(response.raw.read())

driver.quit()
```

解決了圖片問題，接下來就寄信了，過程有點繁複



```
def send_mail(my_pic_dir):
    import os
    from datetime import datetime as d
    import smtplib
    from email.mime.multipart import MIMEMultipart # email內容載體
    from email.mime.text import MIMEText # 用於製作文字內文
    from email.mime.image import MIMEImage
```

導入相關函式

```
account = "rasptest@gmail.com"
```

寄件人帳號

```
# 環境變量設定https://cuiqingcai.com/8947.html
# https://www.youtube.com/watch?v=lo1xqkL7cD8
password =
recepoint = "rasptest@gmail.com"
```

密碼**os.environ.get
("environment_variable")

```
mine = MIMEMultipart()
mine["Content-Type"] = "image/jpeg"
mine["Content-Disposition"] = f'attachment; filename="{my_pic_dir}"' # 寫你的檔案名讓他可以找到
mine["Subject"] = f"這是{d.date(d.today())}的運勢" # 撰寫郵件標題
mine["From"] = f"Rasper[{account}]" # 撰寫你的暱稱或是信箱
mine["To"] = f"Rasper[{recepoint}]" # 撰寫你要寄的人
# mime["Cc"] = "@gmail.com, @gmail.com" # 副本收件人
image = MIMEImage(open(rf'{my_pic_dir}', 'rb').read(), _subtype="application")
image.add_header('Content-Disposition', 'attachment', filename='img.jpg')
mine.attach(image)
# 撰寫內文內容，以及指定格式為plain，語言為中文
mine.attach(MIMEText("這是今日的運勢圖，請查看", "plain", "utf-8"))
msg = mine.as_string() # 將msg將text轉成str
print(msg)
```

收件人

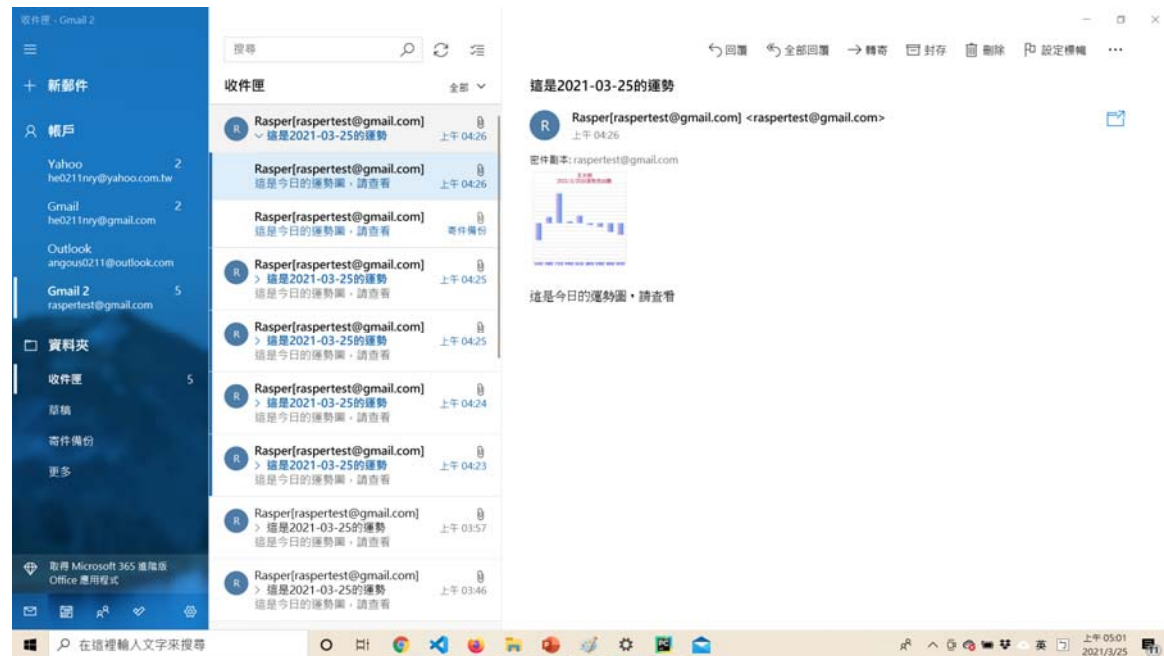
郵件相關格式

```
smtp = smtplib.SMTP("smtp.gmail.com") # google的ping
smtp.ehlo() # 申請身分
smtp.starttls() # 加密文件，避免私密信息被截取
smtp.login(account, password)
status = smtp.sendmail(from_addr=account, to_addrs=recepoint, msg=msg)
```

模擬SMTP(Simple Mail Transfer Protocol)協定
寄出信件

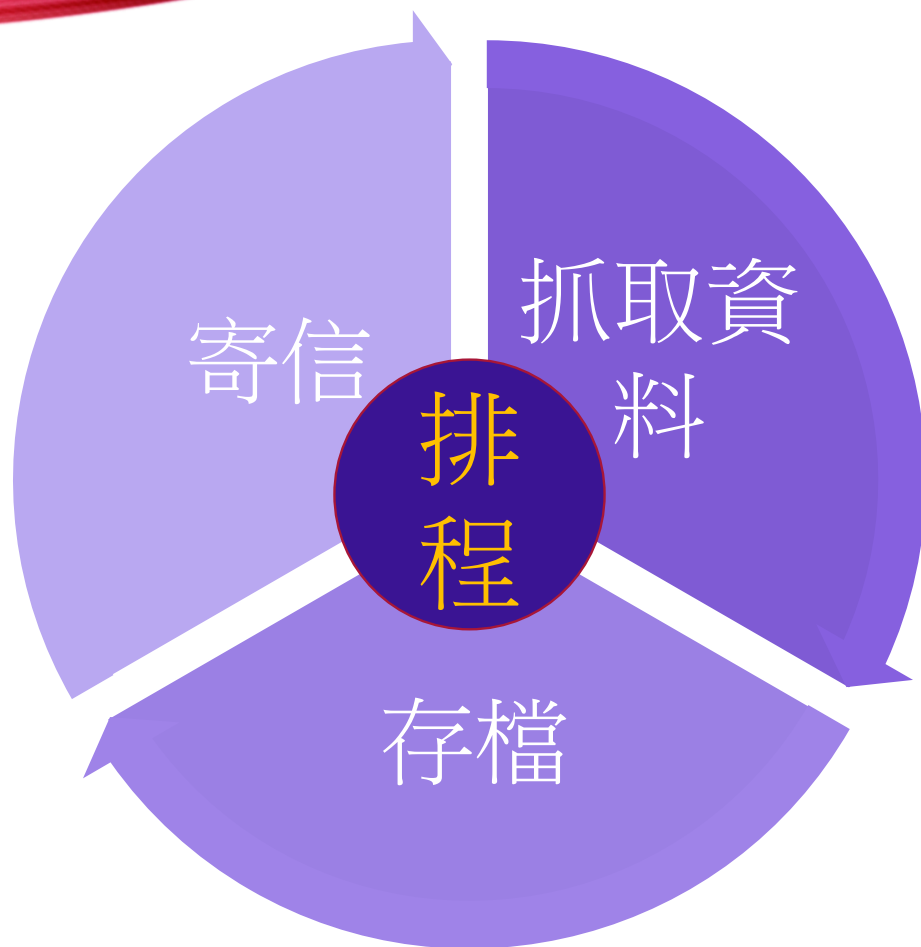
```
if status == {}:
    print("郵件傳送成功!")
else:
    print("郵件傳送失敗!")
smtp.quit()
```

信件也確實送出後，



把整段程式碼分成三個，查詢，存檔，寄信

最後一步就是排程，把我的每日運勢圖在早上六點寄出



```
import time
import schedule
from function import
search_my_fortune, save_my_pic, send_mail
from practice import find_my_fortune

# https://schedule.readthedocs.io/en/stable/examples.html

def job():  # 定義每日排程工作
    my_url = search_my_fortune()
    # my_url = find_my_fortune()

    path = save_my_pic(my_url)

    send_mail(path)

schedule.every(5).seconds.do(job)  # 規定排程頻率

while True:  # 迴圈一直循環
    schedule.run_pending()
    time.sleep(5)
```


Selenium在執行情況下可以不要有瀏覽器一直跑來嗎？

滑鼠右鍵

不行
必須開著瀏覽器讓它能夠
切換新視窗

用SELENIUM或
BEUTIFULSOUP

- 可以喔
- key in 以下:

```
from selenium.webdriver.chrome.options import Options
options = Options()
options.add_argument("--headless")
driver = webdriver.Chrome("chromedriver", options=options)
```

1.命理網站:

<http://www.joyglobal.cc/joyday1.asp>

2. Select 下拉式選單作:

<https://jzchangmark.wordpress.com/2015/03/05/%E9%80%8F%E9%81%8E-selenium-%E6%93%8D%E4%BD%9C%E4%B8%8B%E6%8B%89%E5%BC%8F%E9%81%B8%E5%96%AE-select/>

3. Pyautogui操作:

<https://ithelp.ithome.com.tw/articles/10230717>

4.視窗切換:

<https://www.itread01.com/content/1545450012.html>

5.郵件操作: <https://medium.com/jeasee%E9%9A%A8%E7%AD%86/pythonemail%E5%AF%84%E4%BF%A1-ba2b5eb05d6b>

<https://www.itread01.com/content/1550010974.html>

6.定時排程:

<https://schedule.readthedocs.io/en/stable/examples.html>

7.Selenium在不開啟瀏覽器下執行:

<https://iter01.com/555516.html>

8.環境變量設定:

<https://cuiqingcai.com/8947.html>

<https://www.youtube.com/watch?v=lo1xqkL7cD8>

