

## 記錄及檢查防疫指引更新情況

Ka Ian 2021-09-30

因為近月 COVID-19 疫情，澳門均按照情況而發出相應指引。因為我需要舉辦活動，需要留意在有關指引下活動能否如期舉行。8 月初時發佈的「關於舉辦外訪和人群聚集活動的指引」一直至 9 月中才更新，這段期間不時都需要留意有否更新措施，且難以及時掌握資料，為此製作用作記錄及檢查防疫指引更新情況的 python 程序，希望減少人力問題以及可以掌握資訊。按照需求，程序中僅檢查「政府部門及公共設施」及「教育場所及社會住宿設施」兩個表格。

自動運行的方法：

1. 放到 server（但我的 server 不支持 Python）
2. 使用 Windows 的工作排程器，按時自動運行（已設定每天 23:55 運行一次）

### 1. 爬取資料（原設計及新設計）

以下說明以「政府部門及公共設施」部分為例。

原設計：進入相關頁面後，選取有關的表格部分，取其<tr>行的值，包括指引標題及發佈日期。隨後檢查程序當天的日期會否在所取<tr>行的值當中，若是即當天有新發佈，否則沒有。

```
# 是日新措施
today_guidelines = []

# 政府部門及公共設施
print('** 政府部門及公共設施 **')
table_gov = browser.find_element_by_id('grid_17674')
rows_gov = table_gov.find_elements_by_css_selector('tr')

for r in rows_gov[1:]:
    print(r.text)
    if today in r.text:
        today_guidelines.append(r.text)
```

新設計：考慮到僅儲存指引標題及發佈日期並不足夠，因有機會於日後發佈同樣標題的新指引，要是想查回指定日期的指引將稍有困難，因此把指引文件的連結也下載下來，以便查找。

如只查找文字，可找 css 中的<tr>即可；需查找文件的連結的話也需要找它的<a>再用 get\_attribute('href')去取得連結。

```
gov_data = table_gov.find_elements_by_css_selector('tr')_# 標題 + 日期
link_gov = table_gov.find_elements_by_css_selector('a')_# 為了get_attribute('href')
```

這存在查找需要的問題，gov\_data 可以找出標題及日期，link\_gov 可以找出標題及連結，而我需要的是標題、日期及連結。因此，第二次設計中均查找 gov\_data 及 link\_gov 的資料，及後以其相關的內容（標題）作比對，再組成新 list。

```
for i in link_gov:_# doc names
    for j in gov_data:_# doc names with date
        if i.text in j.text:_# 如超連結的文件也在指引標題中，則而者為同一項目
            gov_list.append(j.text)_# 標題 + 日期
            gov_list.append(i.get_attribute('href'))_# link
```

gov\_list 是儲存形式是['標題 日期', 'link', '標題 日期', 'link', ...]，因此需要為它們分組成每 2 個一組，即[['標題 日期', 'link'], ['標題 日期', 'link'], ...]。其分組結果會是在一個 list 中包含數個 list，那些小 list 中再包含標題+日期及文件連結。

```
_# 分組，儲存到edu_group
step = 2
gov_group = [gov_list[i:i+step] for i in range(0, len(gov_list), step)]
```

最後，檢查那些指引中的發佈日期會否是今天發佈，若是則儲存到 today\_guidelines，用作最後的存檔。

```
_# 與today有關的儲存至today_guidelines
for each in gov_group:
    if today in each[0]:_# each[0]是標題 日期，each[1]是link
        today_guidelines.append(each)
```

## 2. 記錄資料

資料呈現的寫法調整，純粹為了清晰好看一點。

```
# 將今日的指引及link用指定寫法save到results
results = []
for i in today_guidelines:
    results.append(f'{(i[0].split(" "))[0]}, 有關文件: {i[1]}\n')
```

無論有沒有新指引均在程序中列印。

```
# print是日資訊
print(f'今日公佈了{num_guidelines}個防疫指引')
for res in results:
    print(res)
```

如果有發佈新指引，標題、日期及文件連結將會儲存起來，並發電郵給自己。

```
# 有新指引才發email及儲存
if num_guidelines > 0:
    # send email
    email_content = '\n'.join(results)
    email_subject = f'{today_doc}_公佈了{num_guidelines}個防疫指引'
    send_simple_message(content=email_content, subject=email_subject)

    # save txt
    w_file = open(f'./{today_doc}.txt', 'w', encoding='utf-8')
    w_file.write(f'{today}公佈了{num_guidelines}個防疫指引' + '\n')
    w_file.writelines(results)
    w_file.close()
```

網頁上的日期是以'%d/%m/%Y'格式的，原先將儲存的 txt 名稱設為程序檢查日期中的 today，但因為檔案名稱中不可以包含斜號，故在獲取日期時，另外生成 today\_doc 用作文件名稱。

```
# Today
today_orig = datetime.datetime.today()
today = today_orig.strftime('%d/%m/%Y')
today_doc = today_orig.strftime('%d%m%Y')
```

### 3. 結果

print 當天指引（以真實時間為例，9 月 30 日沒有指引）

```
prevent_covid19_doc x
C:\Users\user\Desktop\pycharm_codes\CM540\venv\Scripts\python.exe
** Getting... 政府部門及公共設施 **
** Getting... 教育場所及社會住宿設施 **
*****
今日公佈了0個防疫指引
Process finished with exit code 0
```

print 當天指引（以 2020 年 12 月 23 日為例，有指引）

```
prevent_covid19_doc x
C:\Users\user\Desktop\pycharm_codes\CM540\venv\Scripts\python.exe C:\Users\user\Desktop\pycharm_codes\CM540\prevent_covid19_doc.py
** Getting... 政府部門及公共設施 **
** Getting... 教育場所及社會住宿設施 **
*****
今日公佈了13個防疫指引
體育設施的管理建議，有關文件：https://www.ssm.gov.mo/docs/17751/17751\_0c0900c4afaa45f2be811c4df81acabc\_000.pdf
公園的管理建議，有關文件：https://www.ssm.gov.mo/docs/17622/17622\_2bc92605be7f4185b0bb185a2a4cf5e4\_000.pdf
室內公共設施的管理建議，有關文件：https://www.ssm.gov.mo/docs/17623/17623\_2cb88ccc5b5d481aab8374bed599b243\_000.pdf
給公共機構的建議，有關文件：https://www.ssm.gov.mo/docs/16943/16943\_e244696be19a4b04928a12c90ab85eb0\_000.pdf
給高等院校及學生的建議，有關文件：https://www.ssm.gov.mo/docs/17506/17506\_e26f90978e8a4a738bdcca7e568dfc0\_000.pdf
給遊樂中心的管理建議，有關文件：https://www.ssm.gov.mo/docs/17592/17592\_bfb1eb0d744a4cd1922f782c6c5c3d5a\_000.pdf
給高等院校復課後的管理建議，有關文件：https://www.ssm.gov.mo/docs/17595/17595\_51c10a74d3134f34a2a1bf87ec6211ae\_000.pdf
給托兒所、學校及同類場所的建議，有關文件：https://www.ssm.gov.mo/docs/16949/16949\_e0ad9c6798aa4dc6bf0e42c3a1fb8eec\_000.pdf
給日間社會服務設施及同類場所的管理建議，有關文件：https://www.ssm.gov.mo/docs/17933/17933\_c9fc85a80b8442ec9914ed0511318f1c\_000.pdf
給私立補充教學輔助中心及持續教育機構等同類場所的建議，有關文件：https://www.ssm.gov.mo/docs/17594/17594\_add89645a5164336a052981d2db7e246\_000.pdf
給社會住宿設施及同類場所的建議，有關文件：https://www.ssm.gov.mo/docs/16950/16950\_03d2f0e34c7d435dbc0cf32fd3df687f\_000.pdf
給學校開學後的管理建議，有關文件：https://www.ssm.gov.mo/docs/17834/17834\_6af52d5cba4744318bf3c04fc2cf4ea1\_000.pdf
給家長開學的建議，有關文件：https://www.ssm.gov.mo/docs/17835/17835\_f2879c8a57f249b4bc9fcef3dfbd17f8\_000.pdf
Process finished with exit code 0
```

## 儲存檔案（以 2020 年 12 月 23 日為例，有指引）



## Send email（以 2020 年 12 月 23 日為例，有指引）

