Python & Data - Week 8

Python 解難賽

https://www.cpttm.org.mo/ist/it-competitions/%e5%85%a8%e6%be%b3%e4%b8%ad%e5%ad%b8%e7%94%9fpython%e8%a

新目標 (草案)

目的：通過編寫Python 程式，從網絡上取得資料、處理、輸出結果；題目自定，並與導師相討，共同制定目標。
形式：兩人一組共同解決一個難題。
時間：本年度學期完結前完成。

本期內容

目的：將培正網站 (http://www.puiching.edu.mo/news) 歷年新聞轉移到新平台 (Web Scraping)

為甚麼?

- 改善原有使用體驗（例如另設網站加入按年份搜尋的功能）
- 資料挖掘及統計 (例如找出某位同學出現在消息的次數)
- 轉移資料 (例如將資料轉移到新網站)

內容

Python 的用途：

1. "瀏覽" 網站讀取資料
2. 辨識網站上有用的資料，作進一步處理
3. 將資料按一定結構放到檔案或資料庫

步驟（預計）：

1. 評估可行性
2. 尋找規律
3. 編寫測試程式片段
4. 實作程式及運行
5. 其他注意事項

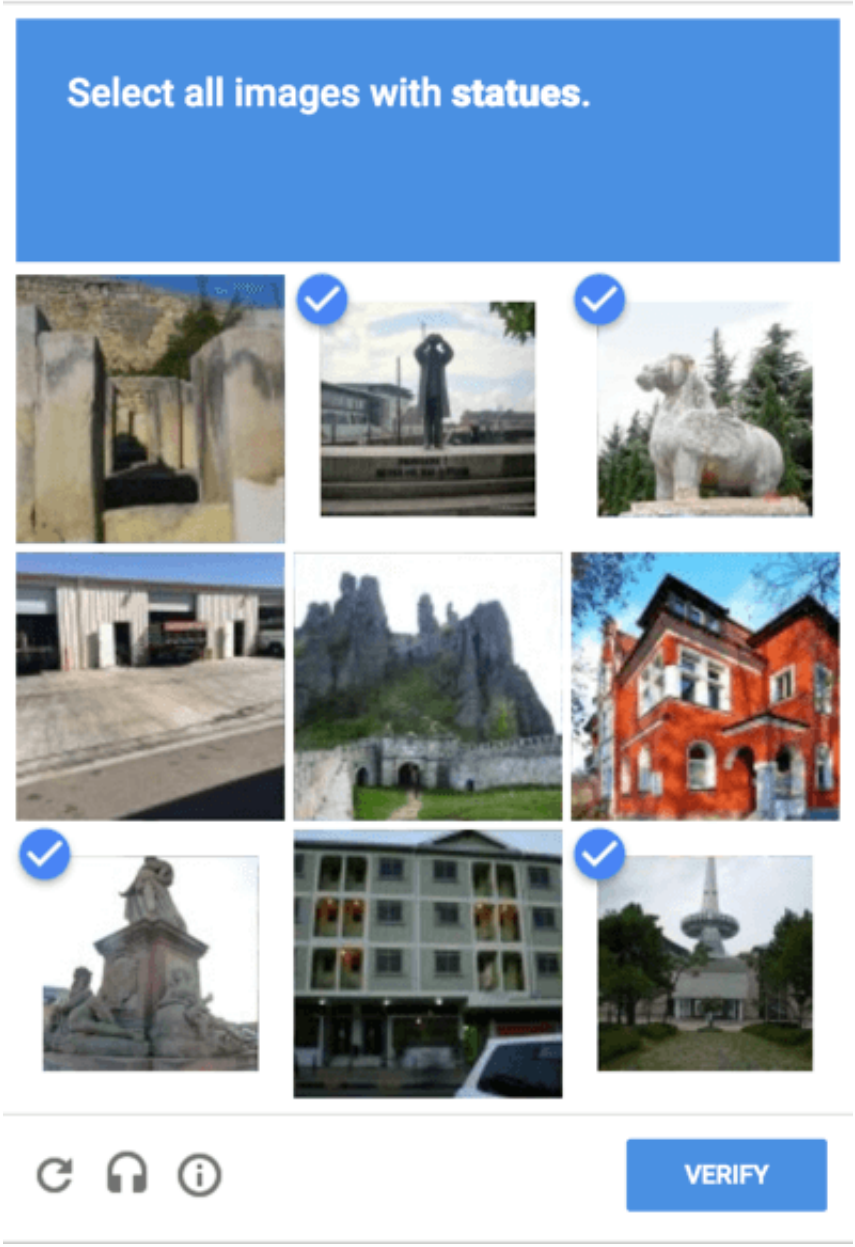
評估可行性

1. 是否有其他解決方法?

Web Scraping 應該作為解決問題的最後手段，如可行，應使用 API (Application Programming Interface)，獲取資料。現在很多機構 (例如：data.gov.mo) 都會提供 API，使用 API 比 Web Scraping 容易。

1. 網站的使用許可是否容許

1. 技術上是否可行?



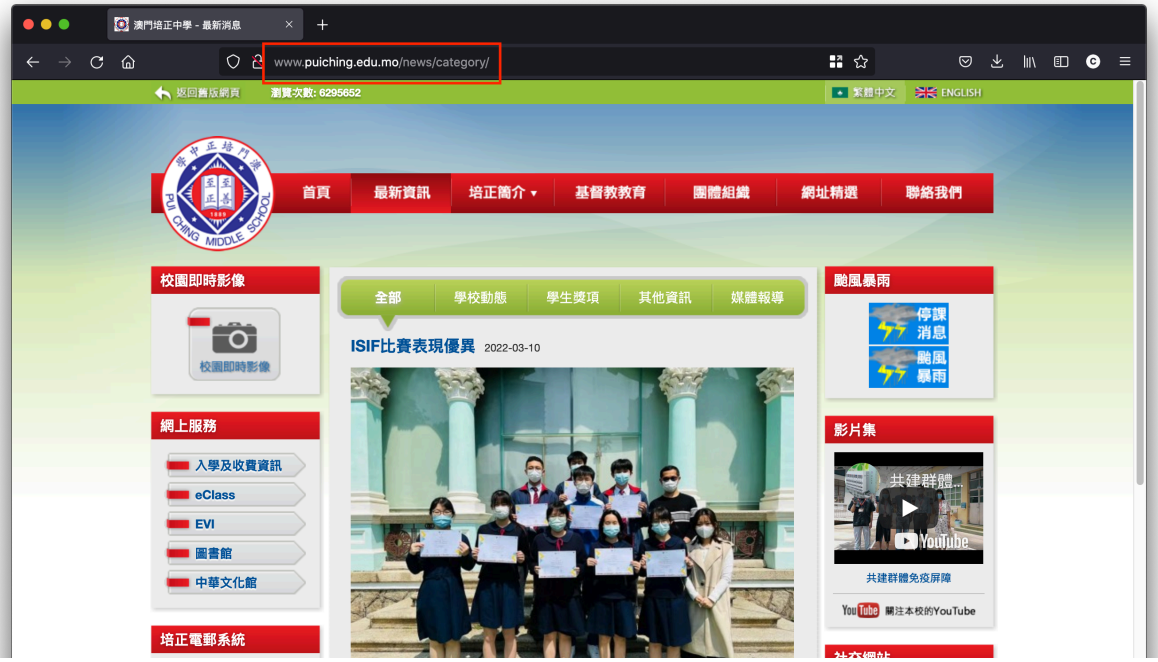
同時要注意的是，很多網站會限制一定時間內的訪問量 (例如一分鐘內有多少個連線)。而在道德上，我們也應該限制我們的連線速度，減少影響其他使用者。

尋找規律

1. 使用一般的方式瀏覽網站，辨識適合用於挖掘的部分

1. 找出網址的規律


返回舊版網頁
瀏覽次數: 6295646
繁體中文
ENGLISH
首頁
最新資訊
培正簡介
基督教教育
團體組織
網址精選
聯絡我們
校園即時影像
網上服務
入學及收費資訊
eClass
EVI
圖書館
中華文化館
培正電郵系統
帳號:
密碼:
登入
全部
學校動態
學生獎項
其他資訊
媒體報導
ISIF比賽表現優異 2022-03-10
第7屆 International Science and Invention Fair (ISIF) 比賽，以綫上形式進行，本校理化科組的項目參與是次賽事，來自30多個國家的400多個項目競逐數學、物理能源與工程、生命科學、環境科學、科技、社會科學六個範疇的獎項。
競賽形式包括海報設計展示、論文報告闡述、視頻短片介紹、視像解說及答辯，全部以英語進行。
本校參賽2個項目喜獲佳績，1銀、1銅。
物理、化學科組：
成員：
高二望 孔悅釗、黃東城、劉銘燁、鄧景恆、鄧欣欣
獎項：銀獎
指導老師：曾玉金老師、古成威老師
成員：
高二信 王丹怡、梁穎欣
高二望 吳曉汶、鄭麗汶
獎項：銅獎
指導老師：曾玉金老師、古成威老師
詳細
澳門郵電局徵文比賽成績優異 2022-03-06
颱風暴雨
停課消息
颱風暴雨
影片集
共建群體免疫屏障
關注本校的YouTube
社交網站
關注本校的Facebook
關注本校的Instagram
特別資訊RSS


澳門培正中學 - 最新消息
www.puiching.edu.mo/news/category/
返回舊版網頁
瀏覽次數: 6295652
繁體中文
ENGLISH
首頁
最新資訊
培正簡介
基督教教育
團體組織
網址精選
聯絡我們
校園即時影像
網上服務
入學及收費資訊
eClass
EVI
圖書館
中華文化館
培正電郵系統
全部
學校動態
學生獎項
其他資訊
媒體報導
ISIF比賽表現優異 2022-03-10
颱風暴雨
停課消息
颱風暴雨
影片集
共建群體免疫屏障
關注本校的YouTube

1. 辨識HTML碼中有用的部分

返回舊版網頁　瀏覽次數: 6295661　繁體中文

首頁 | 最新資訊 | 培正簡介 | 基督教教育 | 團體組織 | 網址精選 | 聯絡我們

校園即時影像

校園即時影像

網上服務

- 入學及收費資訊
- eClass
- EVI
- 圖書館
- 中華文化館

培正電郵系統

帳號:

密碼:

登入

全部 | 學校動態 | 學生獎項 | 其他資訊 | 媒體報導

CTF網絡安全比賽奪旗競賽獲獎者數　2022-03-02

2月19日本校資訊科技科老師盧聖生和黃燦霖再次組織了6名中學生（梁雅姿、莊曜聰、鄧浩峰、顏俊偉、黃天佑、葉俊濠）參加由澳門網絡安全暨奪旗競賽協會 (MOCTF)、澳大電腦學會及澳大學生會聯合舉辦的 第二屆CTF網絡安全比賽《菜鳥黑客松》奪旗競賽＊。

今年比賽以個人賽形式在澳門大學進行，本校學生與大學生們同場比試，參加選手需要在指定時間內從不同的資訊安全主題中選擇答題，包括網絡漏洞、密碼學、逆向工程、電腦法證學、電腦漏洞和其他雜項。經歷8小時連續不停解題下，最終各人以名列前茅成績獲得《優異獎》，各人名次如下：

顏俊偉 S5A（第4名）
黃天佑 S5A（第5名）
鄧浩峰 S6E（第6名）
莊曜聰 S6A（第9名）
葉俊濠 S5A（第10名）
梁雅姿 S6A（第12名）

＊CTF 奪旗賽簡介：CTF 奪旗賽起源於 1996 年 DEF CON 全球黑客大會，以代替之前黑客們通過互相發起真實攻擊進行技術比拼的方式。發展至今，已經成為全球範圍網絡安全圈流行的競賽形式，每年全球舉辦超過 50 場國際性 CTF 奪旗賽賽事。

返回

颱風暴雨

停課消息
颱風暴雨

影片集

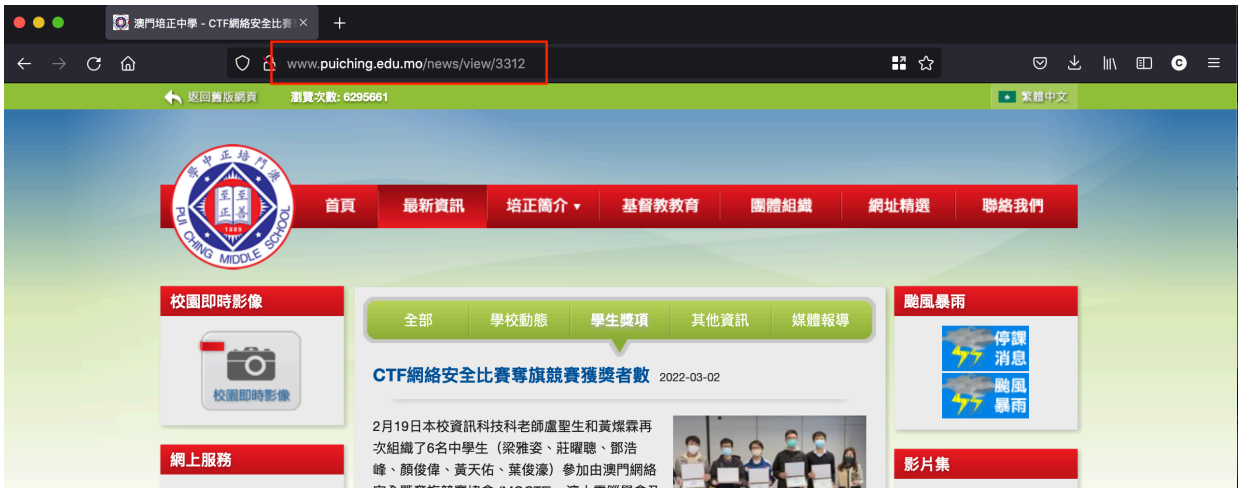
共建群體免疫屏障

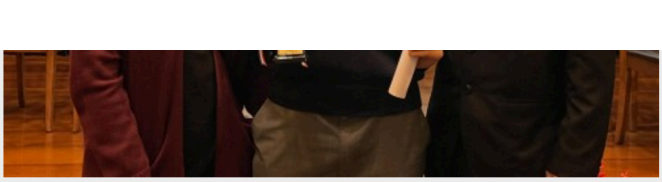
YouTube 關注本校的YouTube

社交網站

關注本校的Facebook
關注本校的Instagram
特別資訊RSS

澳門培正中學
PUI CHING MIDDLE SCHOOL MACAU
澳門高士德馬路七號
Avenida de Horta e Costa, No.7, Macau
TEL: 28529333　FAX: 28339056

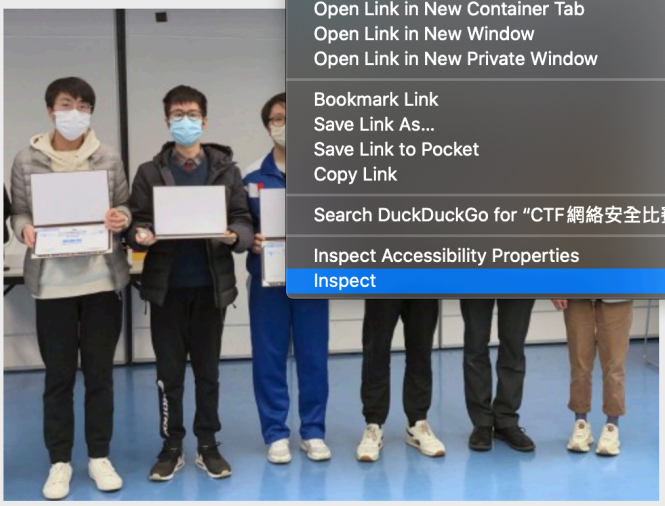




本校高二愛方良宇同學參加由澳門保安部隊高等學校主辦之「警民互助，小城有情」中文徵文比賽，獲得公開組（高中組別）優異獎，指導老師方秀娟。

詳細

CTF網絡安全比賽奪旗競賽



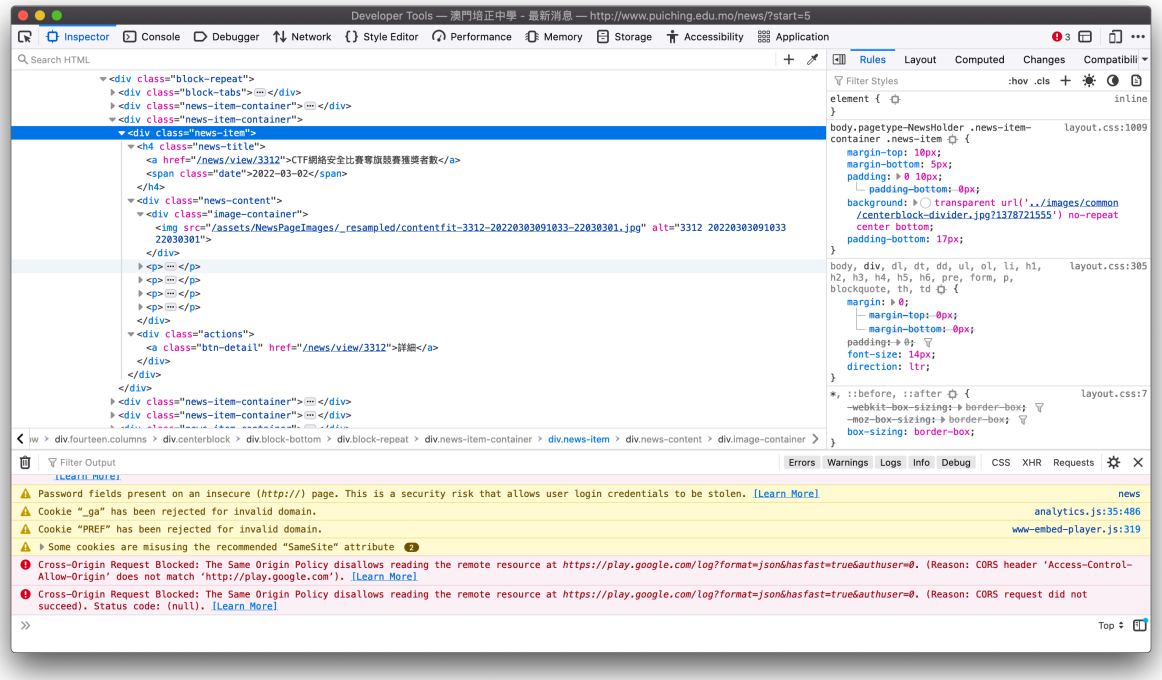
Open Link in New Tab
Open Link in New Container Tab ▶
Open Link in New Window
Open Link in New Private Window

Bookmark Link
Save Link As...
Save Link to Pocket
Copy Link

Search DuckDuckGo for “CTF網絡安全比賽奪旗競賽獲獎...”

Inspect Accessibility Properties
Inspect

2月19日本校資訊科技科老師盧聖生和黃燦霖再次組織了6名中學生（梁雅姿、莊



編寫測試程式片段

1. 規劃程式的組成
   1a. "瀏覽" 網站取得內容 (urlrequest, file open)
   1b. 解析及截取有的資料片段 (regex, xml parser)
   1c. 轉化資料及儲存 (資料庫、檔案)
   1d. 整合、統計，令資料產生價值 (map, filter, reduce)

1. 將每個組成以 Function 或 Class 的型式編寫

例如：

In [ ]:

```
# Example process

def run():
    scrap_result = step1_scrap()
    parsed_result = step2_parse(scrap_result)
    step3_save(parsed_result)
    step4_process(parsed_result)

def step1_scrap():
    print("1. Get some content from http://www.puiching.edu.mo/news")
    result = "Return some result"
    return result

def step2_parse(input):
    print(f"2. Get some data from step 1: {input}")
    parsed = "parsed result"
    return parsed

def step3_save(input):
    print(f"3. Save result: {input}")

def step4_process(input):
    print(f"4. Process, count etc: {input}")


run()
```

```
1. Get some content from http://www.puiching.edu.mo/news
2. Get some data from step 1: Return some result
3. Save result: parsed result
4. Process, count etc: parsed result
```