

Python & Data - Week 8

Python 解難賽

<https://www.cpttm.org.mo/ist/it-competitions/>

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)

為什麼？

- 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）



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，以線上形式進行，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：

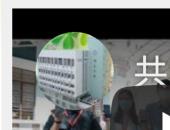
成員：
高二望 孔悅釗、黃東城、劉銘輝、鄧景恆、鄧欣欣
獎項：銀獎
指導老師：曾玉金老師、古成威老師

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

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](#)

澳門郵電局徵文比賽成績優異 2022-03-06

[詳細](#)

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

About ▾

Careers ▾

Diversity



Blog

Contact

Press

- Access the Services using any interface other than ours;
- Maintain any link to the Services that we ask you to remove, in our sole discretion;
- Frame the Services or Content, make the Services or Content available via in-line links, otherwise display the Services or Content in connection with an unauthorized logo or mark, or do anything that could falsely suggest a relationship between us or our affiliates and any third party or potentially deprive us of revenue (including, without limitation, revenue from advertising, branding, or promotional activities);
- Threaten, defame, stalk, abuse, or harass other persons or engage in illegal activities, or encourage conduct that would constitute a criminal offense or give rise to civil liability;
- Transmit any material that is inappropriate, profane, vulgar, offensive, false, disparaging, defamatory, obscene, illegal, sexually explicit, racist, that promotes violence, racial hatred, or terrorism, or that we deem, in our sole discretion, to be otherwise objectionable;
- Violate any person's or entity's legal rights (including, without limitation, intellectual property, privacy, and publicity rights), transmit material that violates or circumvents such rights, or remove or alter intellectual property or other legal notices;
- Transmit files that contain viruses, spyware, adware, or other harmful code;
- Advertise or promote goods or services without our permission (including, without limitation, by sending unsolicited email);
- Remove, modify, disable, block or otherwise impair any advertising in connection with the Services;
- Interfere with others using the Services or otherwise disrupt the Services;
- Disassemble, decompile or otherwise reverse engineer any software or other technology included in the Content or used to provide the Services;
- Transmit, collect, or access personally identifiable information about other users without the consent of those users and us;
- Engage in unauthorized spidering, "scraping," data mining or harvesting of Content, or use any other unauthorized automated means to gather data from or about the Services;
- Impersonate any person or entity or otherwise misrepresent your affiliation or the origin of materials you transmit;
- Remove, avoid, interfere with, or otherwise circumvent any access control measures for the Services or Content, including password-protected areas and geo-filtering mechanisms, or any digital rights management measures used in connection with Content; or
- Access any portion of the Services that we have not authorized you to access (including password-protected areas), link to password-protected areas, attempt to access or use another user's account or information, or allow anyone else to use your account or access credentials.

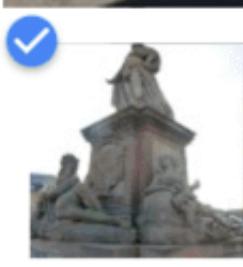
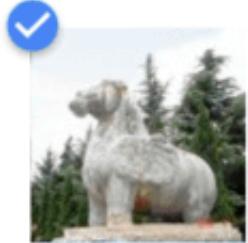
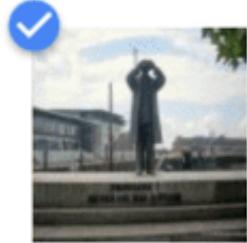
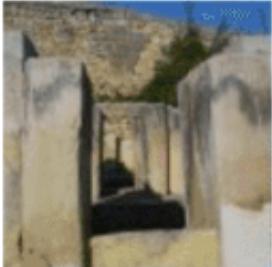
If you violate this Agreement, we may terminate your access to the Services without notice, and take any other actions or seek any remedies permitted by law.

If we terminate your access to any of the Services, you must immediately stop using such Service. However, if you have paid for a subscription to a paid Service, and we discontinue the Service before the end of a paid subscription period, or we terminate your account before the end of a paid subscription period for reasons other than your breach of this Agreement, we will refund a prorated portion of the applicable subscription fee corresponding to the portion of the paid subscription period for which our action caused you not to have access to the relevant Service. If we terminate your access to a paid Service because you breached this Agreement, you will not be entitled to any refund.

[Fee-Based Services and Single Purchases](#)

1. 技術上是否可行?

Select all images with statues.



VERIFY

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

The screenshot shows the homepage of Pui Ching Middle School Macau. At the top, there is a banner featuring two people in medical scrubs holding a small award or certificate. Below the banner, the school's name is partially visible: "Macau Pui Ching Middle School". A green navigation bar contains links for "全部", "學校動態", "學生獎項", "其他資訊", and "媒體報導". To the left, there is a sidebar with sections for "校園即時影像" (featuring a camera icon), "網上服務" (with links to "入學及收費資訊", "eClass", "EVI", "圖書館", and "中華文化館"), and "培正電郵系統" (with fields for "帳號" and "密碼", and a "登入" button). The main content area features a large image of students holding certificates in front of a classical building. Below this image is a list of news items, each with a date and a title. Two red arrows point from the text "日期" and "標題" to the first item in the list. The news items are:

- 2022-03-06 澳門郵電局徵文比賽成績優異
學界田徑比賽破紀錄；數人奪冠
- 2022-03-05 2022高三學生活動供應酒店餐飲招標
- 2022-03-05 本校體育科組參與國際網上研討會
- 2022-03-04 「警民互助，小城有情」徵文比賽獲獎
- 2022-03-02 CTF網絡安全比賽奪旗競賽奪冠者數
- 2022-03-01 全澳青少年劍擊錦標賽奪冠者數
- 2022-03-01 我校學生在央視演唱獲高度評價
- 2022-02-26 兒科醫生講座籲盡早接種疫苗

On the right side of the page, there are additional sections: "颱風暴雨" (with icons for lightning and rain), "影片集" (with a thumbnail of a video showing a group of people), and "社交網站" (links to the school's Facebook, Instagram, and RSS feed).

1. 找出網址的規律

返回舊版網頁 濱覽次數: 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



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:

ISIF比賽表現優異 2022-03-10



第7屆 International Science and Invention Fair (ISIF) 比賽，以線上形式進行，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、劉銘輝、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

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

 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

返回舊版網頁 濱覽次數: 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



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

ISIF比賽表現優異 2022-03-10



第7屆 International Science and Invention Fair (ISIF) 比賽，以線上形式進行，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、劉銘輝、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

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

社大網站

今天晚上，本校體育科組全體老師參與一場有關“Game Contribution Assessment Instrument (GCAI)”體育評估工具國際網上研討會，在會議中集合多國體育教學研究人員，互相交流討論。

詳細

1 2 3 4 ...



澳門培正中學

The screenshot shows the homepage of Puiching Middle School's website. At the top, there is a navigation bar with links for 首頁 (Home), 最新資訊 (Latest Information), 培正簡介 (About Pui Ching), 基督教教育 (Christian Education), 團體組織 (Organizations), 網址精選 (Selected Websites), and 聯絡我們 (Contact Us). Below the navigation bar, there is a banner featuring the school's logo and a photograph of three people wearing face masks. To the left of the banner, there is a "School Live Stream" section and a "Online Services" section with links for 入學及收費資訊 (Admission and Tuition Information), eClass, EVI, 圖書館 (Library), and 中華文化館 (Chinese Culture Museum). On the right side, there are sections for "颱風暴雨" (Typhoon and Rain) and "影片集" (Video Collection) with a YouTube link.

1. 辨識HTML碼中有用的部分

返回舊版網頁 濟覽次數: 6295661 繁體中文



首頁 最新資訊 培正簡介 基督教育 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](#)

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](#)



返回舊版網頁 濟覽次數: 6295661 繁體中文



首頁 最新資訊 培正簡介 基督教育 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

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](#)

電郵系統

登入



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](#)

社交網站

- 關注本校的Facebook
- 關注本校的Instagram
- 特別資訊RSS

CTF网络安全比賽奪旗競賽



[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网络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

Developer Tools — 澳門培正中學 - 最新消息 — http://www.pulching.edu.mo/news/?start=5

Inspector Console Debugger Network Style Editor Performance Memory Storage Accessibility Application

Rules Layout Computed Changes Compatibility

Search HTML

```
<div class="block-repeat">
  <div class="block-tabs">::></div>
  <div class="news-item-container">::></div>
  <div class="news-item-container">
    <div class="news-item">
      <h4 class="news-title">
        <a href="/news/view/3312">CTF网络安全比賽奪旗競賽獲獎者數</a>
        <span class="date">2022-03-02</span>
      </h4>
      <div class="news-content">
        <div class="image-container">
          
        </div>
        <p>...</p>
        <p>...</p>
        <p>...</p>
        <p>...</p>
      </div>
      <div class="actions">
        <a class="btn-detail" href="/news/view/3312">詳細</a>
      </div>
    </div>
    <div class="news-item-container">::></div>
    <div class="news-item-container">::></div>
    <div class="news-item-container">::></div>
  </div>
</div>
```

Errors Warnings Logs Info Debug CSS XHR Requests

Learn More

⚠ Password fields present on an insecure (<http://>) page. This is a security risk that allows user login credentials to be stolen. [\[Learn More\]](#)

⚠ Cookie “_ga” has been rejected for invalid domain.

⚠ Cookie “PREF” has been rejected for invalid domain.

⚠ Some cookies are misusing the recommended “SameSite” attribute [\[?\]](#)

❗ Cross-Origin Request Blocked: The Same Origin Policy disallows reading the remote resource at <https://play.google.com/log?format=json&hasFast=true&authUser=0>. (Reason: CORS header ‘Access-Control-Allow-Origin’ does not match ‘<http://play.google.com>’). [\[Learn More\]](#)

❗ Cross-Origin Request Blocked: The Same Origin Policy disallows reading the remote resource at <https://play.google.com/log?format=json&hasFast=true&authUser=0>. (Reason: CORS request did not succeed). Status code: (null). [\[Learn More\]](#)

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