東吳大學X理律學堂 網路爬蟲



何謂爬蟲?

爬蟲web-crewler

- ▼又稱網路蜘蛛Spider
- ●網路爬蟲就是讓程式模擬人操作電腦的行為,利用自動化大量 搜集特定資料的過程。再簡單一點說,就是寫一套程式,讓它幫 你自動抓取特定資料
- ●例子: Google、Yahoo大型爬蟲案例

網頁的運作方式



HTTP

●超文本傳輸協定 (HTTP) 是一種用來傳輸超媒體文件 (像是HTML文件) 的應用層協定, 被設計來讓瀏覽器和伺服器進行溝通, 但也可做其他用途。HTTP 遵循標準客戶端—伺服器模式, 由客戶端連線以發送請求, 然後等待接收回應。HTTP 是一種無狀態協定, 意思是伺服器不會保存任兩個請求間的任何資料 (狀態)。儘管作為 TCP/IP 的應用層, HTTP 亦可應用於其他可靠的傳輸層 (例如 UDP), 只要不會無聲無息地遺失訊息即可。

HTTPS

●https 全名 超文本傳輸安全協定,那個 s 就是 secure 的意思;https 透過 http 進行通訊,但通訊過程使用 ssl/tls 進行加密,藉由類似於前述的加密方式,在 http 之上定義了相對安全的資料傳輸方法。由於非對稱加密的運算量較高,傳遞回應較慢;實際的架構上,會透過公開金鑰加密傳遞出共用的金鑰,再透過共用金鑰加密進行後續的傳遞,兼顧了安全性及傳遞速度。

靜態網頁

- ●指網站完成一個請求 (request)與回應 (response)後, 用戶端即不再與伺服器有任何的交流, 所有的互動都只與瀏覽器的網頁互動, 資訊不會傳遞到後端伺服器。
- ●比較容易實作, 只要爬蟲已經閱讀完整份網頁, 就可以取得這個網頁所有的 資訊進行分析。



動態網頁

- ●指網站會依照使用者的行為不斷的與伺服器進行交流, 傳送資訊給伺服器, 資訊經過伺服器處理後, 才會回應相關資訊, 甚至不少動態網站甚至需要進行「登入」的動作。
- ●爬蟲實作比較複雜, 爬蟲必須要知道網站需要什麼「資訊」, 提供了正確的資訊, 才能取得所需要的資料。



請求方式(Requests)





HTTP方法	說明
GET	向指定資源提交請求,可額外設定 params 參 數字典。
POST	向指定資源提交請求,可額外設定 data 參數字典。
PUT	向指定資源提供最新內容,可額外設定 data 參數字典。
DELETE	請求刪除指定的資源。
HEAD	請求提供資源的回應標頭(不含內容)。
OPTIONS	請求伺服器提供資源可用的功能選項。

─→ 在標頭中傳送(公開)

── 在內容中傳送(隱私)

POST vs. GET

	GET	POST				
網址變化	因查詢而改變	不會改變				
查詢參數	直接寫在網址	在網路封包裡				
資料傳遞量	URL長度有限制,查詢參數會受限	查詢參數不會受限				
安全性	較差	較佳				

HTTP狀態碼



狀態代碼	說明
200	網頁正常
301	網頁搬家, 重新導向到新的網址
400	錯誤的要求
401	未授權,需要憑證
403	沒有權限
404	找不到網頁
500	伺服器錯誤
503	伺服器暫時無法處理請求 (附載過大)。
504	伺服器沒有回應

爬蟲常用套件

Requests

主要功能是HTTP請求,可以代替人手動進入網頁

●Urllib (較少用)

和Requests相似, 可以下載網頁上的資訊, 沒辦法反爬蟲

Beautiful Soup

快速解析 HTML/XML

•selenium

模擬瀏覽器操作資料

爬蟲回傳資料格式

- 1. HTML
- 2. XML
- 3. JSON
- 4.其他



針對格式使用工具提取所需資料

JSON 格式

一.儲存物件-object

{"regulations":"民法","court":"士林地方法院"}

二.陣列格式-array

["民法","士林地方法院"]



資料儲存格式

- 1. JSON
- 2. CSV
- 3. SQL (資料庫)
- 4. XLSX



使用爬蟲的注意事項



●不造成網站伺服器的負擔 time.sleep(秒數)

每次爬取資料時, 設定適當的等待延遲, 避免短時間內送出大量的請求而造成伺服器的負擔(DDoS攻擊, 根據刑法第360條可能會觸法)。

●確認網站是否有提供API

如果網站有提供API 供第三方直接取得資料,可以直接透過API 抓取資料, 節省讀取與分析網站HTML的時間。

●注意 robots.txt

robots.txt 會規範一個網站允許什麼樣的User-Agent 訪問, 也會規範 Crawl-delay 訪問間隔時間, 如果Crawl-delay 設定 1, 表示這個網站期望每次 訪問的時間間隔一秒鐘。

反爬蟲

- 判斷瀏覽器 headers 資訊
- 使用動態頁面
- 加入使用者行為判斷
- 模擬真實用戶登入授權
- 加入驗證碼機制
- 封鎖代理伺服器與第三方 IP



爬取所需資料之後呢?

流程

網路爬蟲



資料清整



資料視覺化



- ●模組化程式設計
- ●定期定時爬取
- ●異常處理



法律角度看爬蟲

靠挖掘別人家的資料數據來賺錢,「網路爬蟲」這個行為合法嗎?

範例實作

網站: 裁判家<u>https://www.lawplus.com.tw/</u>

關鍵字搜索: (車禍+交通事故)&(民法第217+過失比例)

一審判決

目標: 爬取前10頁內容

開啟開發人員模式



Chrome: FN+F12

Safari: 右鍵>顯示網頁原始碼

設定方式:

左上角的 Safari>偏好設定>進階>『在選單列中顯示「開發」選單』



F12

Windows

引用模組

```
In [3]:
import requests
import json
import time
from time import sleep
import pandas as pd
```



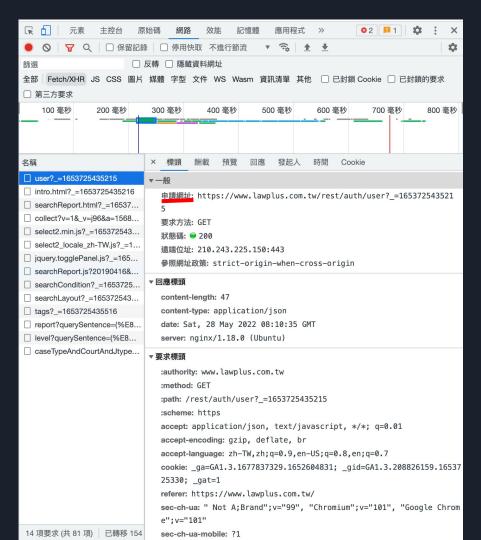
使用colab、jupyter不用特別安裝模組 正常狀況下需要pip install模組



該如何找尋爬蟲所需的網址?

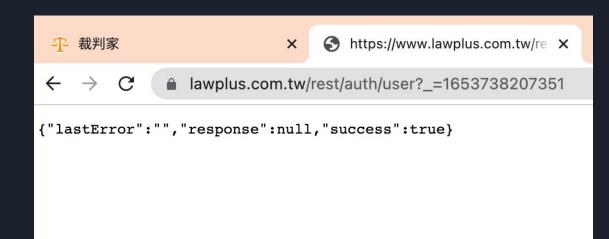
Step1

複製標頭-申請網址(紅色) 至新分頁開啟查看內容



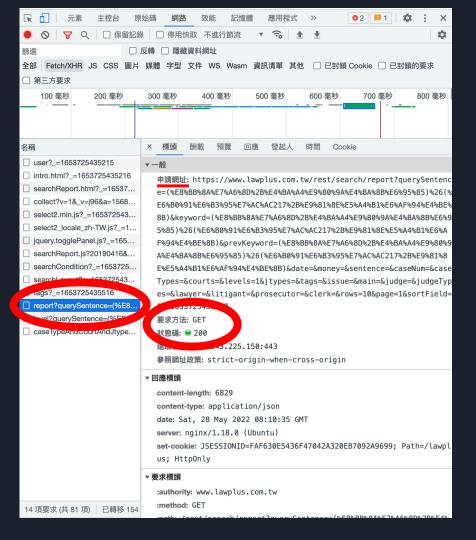
Step2

- ●內容須為我們所要查看 的中文文字才是正確
- ●依序往下找尋



Step3

- ●找到名稱report?query 申請網址 並複製在新分頁開啟
- 同時可以看到方法以及狀態碼



正確內容為以下

```
型 裁判家
                             https://www.lawplus.com.tw/re X
               |awplus.com.tw/rest/search/report?querySentence=(車禍%2B交涌事故)%26(民法第217%2B渦失比例)&keyword=(車禍%2B交涌事故)%26(民法第217%2B渦失比例)&pr... 🔯 🐧
                                                                                                                                  告 泰安產物保險股份有限公司\r\n法定代理人
{"page":1,"records":14017,"rows":[{"caseNum":"105,投小,319","caseType":"V","content":"臺灣南投地方法院民事小額判決
                                                                                                        105年度投小字第319號\r\n原
李松季\r\n訴訟代理人 陳建興\r\n
                                  藝松麟\r\n被
                                                 告 陳慶隆\r\n訴訟代...","court":"","identifier":"NTEV,105,投小,319,20161124,1-80378243-52d7-4154-96bf-59579c9a47e0","issue":"損害賠
償","jtype":"J","judgeDate":"2016/11/24","litiqant":null,"tags":null,"title":"NTEV,105,投小,319,20161124,1-80378243-52d7-4154-96bf-59579c9a47e0","type":"民事判決"},{"caseNum":"107,
訴,1641","caseType":"V","content":"臺灣桃園地方法院民事判決
                                                           107年度訴字第1641號\r\n原
                                                                                    告 新光產物保險股份有限公司桃園分公司\r\n法定代理人 徐慶霖\r\n訴訟代理人 葉書佑\r\n被
                                                                                                                                                             告 張凱雄
\r\n上列當事人間請求...", "court":"", "identifier":"TYDV,107,訴,1641,20180831,1-9d58c481-6ee6-4994-8a56-02183c8a3492", "issue":"損害賠
價","jtype":"J","judgeDate":"2018/08/31","litigant":null,"tags":null,"title":"TYDV,107,訴,1641,20180831,1-9d58c481-6ee6-4994-8a56-02183c8a3492","type":"民事判決"},{"caseNum":"106,士
小,787","caseType":"V","content":"臺灣士林地方法院小額民事判決
                                                         106年度士小字第787號\r\n原
                                                                                   告 新安東京海上產物保險股份有限公司\r\n法定代理人 陳忠鏗\r\n訴訟代理人 謝京燁\r\n被
\r\n上列當事人間侵權行為...", "court":"", "identifier": "SLEV, 106, 士小, 787, 20170811, 1-67289756-bbd8-47de-868e-d44d6d4a478f", "issue": "侵權行為損害賠
價","itype":"J","judgeDate":"2017/08/11","litigant":null,"tags":null,"title":"SLEV,106,士小,787,20170811,1-67289756-bbd8-47de-868e-d44d6d4a478f","type":"民事判決"},{"caseNum":"100,彰
                                                                                                    告 許正國\r\n上列當事人間100年度彰小字第192號損害賠償事件,於中華民國\r\n100年6月
小,192","caseType":"V","content":"臺灣彰化地方法院彰化簡易庭民事判決 100年度彰小字第192號\r\n原
                                                                                   告 郭雅致\r\n被
22日言詞辯...", "court": "", "identifier": "CHEV, 100, 彭小, 192, 20110704, 1-09efcef9-0558-48f0-897f-8312a03e4b2b", "issue": "損害賠
價","itype":"J","judgeDate":"2011/07/04","litigant":null,"tags":null,"title":"CHEV,100,彰小,192,20110704,1-09efcef9-0558-48f0-897f-8312a03e4b2b","type":"民事判決"},{"caseNum":"105,湖
小,595","caseType":"V","content":"臺灣士林地方法院小額民事訴訟判決 105年度湖小字第595號\r\n原
                                                                                   告 肝肝友聯產物保險股份有限公司\r\n法定代理人 孔令範\r\n訴訟代理人 李繼偉\r\n被
                                                                                                                                                       告 李宏裕\r\n h
列當事人間損害賠償事件...","court":"","identifier":"NHEV,105,湖小,595,20160808,1-b44e4001-52b5-4bd0-b309-fabd56dd59ab","issue":"損害賠
價","itype":"J","judgeDate":"2016/08/08","litigant":null,"tags":null,"title":"NHEV,105,湖小,595,20160808,1-b44e4001-52b5-4bd0-b309-fabd56dd59ab","type":"民事判決"},{"caseNum":"105,投
小,392", "caseType": "V", "content": "臺灣南投地方法院民事小額判決
                                                         105年度投小字第392號\r\n原
                                                                                   告 新光產物保險股份有限公司台中分公司\r\n法定代理人 簡國釧\r\n訴訟代理人 張光賓\r\n
                                                                                                                                                               唐永興
          告 李彩鳳...", "court":"", "identifier":"NTEV,105,投小,392,20161124,1-1d726d10-d8d6-4f86-bcc5-5fac352578e7", "issue":"損害賠
價","itype":"J","judgeDate":"2016/11/24","litigant":null,"tags":null,"title":"NTEV,105,投小,392,20161124,1-1d726d10-d8d6-4f86-bcc5-5fac352578e7","type":"民事判決"},{"caseNum":"106,沙
                                                                                                             告 豐原汽車客運股份有限公司\r\n法定代理人 林伯洋\r\n訴訟代理人 蔡閔清
小,424","caseType":"V","content":"臺灣臺中地方法院小額民事判決\r\n
                                                                                    106年度沙小字第424號\r\n原
          告 張秀...", "court": "", "identifier": "SDEV.106,沙小.424.20171024.1-5a012615-d86a-44a5-b517-89a7eb457182", "issue": "損害賠
價","itype":"J","judgeDate":"2017/10/24","litigant":null,"tags":null,"title":"SDEV,106,沙小,424,20171024,1-5a012615-d86a-44a5-b517-89a7eb457182","type":"民事判決"},{"caseNum":"106,屏
小,458","caseType":"V","content":"臺灣屏東地方法院民事判決
                                                         106年度屏小字第458號\r\n原
                                                                                   告 黄吳金桂\r\n訴訟代理人 黃勇勝\r\n被
                                                                                                                       告 黃文彬\r\n兼法定代理 黃清弘\r\n人\r\n上列當事人間請求
損害賠償事件,本院於民...","court":"","identifier":"PTEV.106.屏小.458.20180330.1-d72f2ad9-88fe-42b1-b61b-f6c9f042db0c","issue":"損害賠
價","itype":"J","judgeDate":"2018/03/30","litigant":null,"tags":null,"title":"PTEV,106,屏小,458,20180330,1-d72f2ad9-88fe-42b1-b61b-f6c9f042db0c","type":"民事判決"},{"caseNum":"106,花
小,221","caseType":"V","content":"宣 示 判 決
                                                筆 錄 106年度花小字第221號\r\n原
                                                                                   告 國泰世紀產物保險股份有限公司\r\n法定代理人 胡一敏\r\n訴訟代理人 呂紹偉\r\n被
                                                                                                                                                       告 林奐妤\r\n訴訟
代理人 林□成\r\n上...","court":"","identifier":"HLEV,106,花小,221,20170731,1-f3931972-8d22-4394-9703-dac0f8cf6540","issue":"侵權行為損害賠
僧","itype":"J","judgeDate":"2017/07/31","litigant":null,"tags":null,"title":"HLEV,106,花小,221,20170731,1-f3931972-8d22-4394-9703-dac0f8cf6540","type":"民事判決"},{"caseNum":"107,豐
小,707","caseType":"V","content":"臺灣臺中地方法院小額民事判決
                                                         107年度豐小字第707號\r\n原
                                                                                   告 臺灣產物保險股份有限公司台中分公司\r\n\r\n法定代理人 廖原益 \r\n訴訟代理人 方建閔 \r\n被
楊仁豪 \r\n\r\n上列...","court":"","identifier":"FYEV,107,豐小,707,20181213,1-6afccf14-22a5-4966-9c9a-49d8662e4525","issue":"損害賠
價","jtype":"J","judgeDate":"2018/12/13","litiqant":null,"tags":null,"title":"FYEV,107,豐小,707,20181213,1-6afccf14-22a5-4966-9c9a-49d8662e4525","type":"民事判
決"}],"total":1402,"userdata":{"series":[{"name":"2021","v":312},{"name":"2020","v":1419},{"name":"2019","v":1575},{"name":"2018","v":1585},{"name":"2017","v":1421},
{"name":"2016","y":1309},{"name":"2015","y":551},{"name":"2014","y":917},{"name":"2013","y":772},{"name":"2012","y":714},{"name":"2011","y":551},{"name":"2010","y":512},
{"name":"2009","y":442},{"name":"2008","y":435},{"name":"2007","y":414},{"name":"2006","y":331},{"name":"2005","y":230},{"name":"2004","y":5}],"categories":
["2021", "2020", "2019", "2018", "2017", "2016", "2015", "2014", "2013", "2012", "2011", "2010", "2009", "2008", "2007", "2006", "2005", "2004"]}}
```



- ●使用迴圈方式一頁爬取之後換下一頁
- ●爬蟲爬取網址為link=urlfirst+**str(page)**+urlsec

```
In [2]: case_num=[]
for page in range(1,11):
    urlfirst="https://www.lawplus.com.tw/rest/search/report?querySentence=(%E8%BB%8A%E7%A6%8D%2B%E4%BA%A4%E9%80%9A%E
    urlsec="&sortField=&_=1652714083801"
    link=urlfirst+str(page)+urlsec
    res = requests.get(link)
    data=json.loads(res.text)
    temp_data=data['rows']

    for i in temp_data:
        case_num.append(i['identifier'])
    print(case_num)
```



•res=resquests.get(link)

```
In [4]: ### 以此查看網頁是否正常爬取 res

Out[4]: <Response [200]>
```



●data=json.loads(res.text) 將資料轉換成json格式

```
In [5]: ###觀察data擷取資料
                           data
Out[5]: {'page': 10,
                               'records': 10891,
                               'rows': [{'caseNum': '105,沙小,413',
                                      'caseType': 'V'.
                                     'content': '臺灣臺中地方法院小額民事判決\r\n\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u
                           u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000
                           公司\r\n法定代理人\u3000簡國釧\r\n訴訟代理人\u3000張光賓\r\n被\u3000\u3000...',
                                     'court': '',
                                     'identifier': 'SDEV,105,沙小,413,20161025,1-aa0cc915-683d-49f7-a884-2308d5fc3b5c'.
                                     'issue': '損害賠償',
                                     'jtype': 'J',
                                     'judgeDate': '2016/10/25'.
                                     'litigant': None.
                                     'tags': None.
                                     'title': 'SDEV,105,沙小,413,20161025,1-aa0cc915-683d-49f7-a884-2308d5fc3b5c',
                                     'type': '民事判決'},
                                  {'caseNum': '106,沙小,103',
                                     'caseType': 'V',
                                      'content': '臺灣臺中地方法院小額民事判決\r\n\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u
```

In [20]: ###查看temp data的內容



●temp_data=data['rows'] p.s. ['rows']為各判決編號

```
temp data
Out[20]: [{'caseNum': '105,沙小,413',
                                  'caseType': 'V',
                                 'content': '臺灣臺中地方法院小額民事判決\r\n\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u
                           3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u300\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u
                           公司\r\n法定代理人\u3000簡國釧\r\n訴訟代理人\u3000張光賓\r\n被\u3000\u3000...',
                                  'court': ''.
                                  'identifier': 'SDEV,105,沙小,413,20161025,1-aa0cc915-683d-49f7-a884-2308d5fc3b5c',
                                 'issue': '損害賠償'.
                                 'jtype': 'J',
                                 'iudgeDate': '2016/10/25',
                                 'litigant': None,
                                 'tags': None,
                                 'title': 'SDEV.105,沙小.413.20161025.1-aa0cc915-683d-49f7-a884-2308d5fc3b5c'.
                                 'type': '民事判決'},
                              {'caseNum': '106,沙小,103',
                                  'caseType': 'V',
                                  'content': '臺灣臺中地方法院小額民事判決\r\n\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u
                           3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000106年度沙小字第103號\r\n原\u3000\u3000\u3000\u3000售\u3000富邦產物保險股份有限公司\r\n法
                           定代理人\u3000陳伯燿\r\n訴訟代理人\u3000楊立慈\r\n\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u
```

程式碼-觀察網址



裁判家網址 https://www.lawplus.com.tw/

以下舉「105,投小,319損害賠償」為例:

網頁網址

https://www.lawplus.com.tw/#reportDetail?id=NTEV,105,%E6%8A %95%E5%B0%8F,319,20161124,1-80378243-52d7-4154-96bf-5 9579c9a47e0

爬蟲網址

https://www.lawplus.com.tw/rest/search/report/NTEV,105,%E6%8A%95%E5%B0%8F,319,20161124,1-80378243-52d7-4154-96bf-59579c9a47e0?_=1653742306728



- ●each_url 這個list用於存放網頁真實網址
- ●case_num這個list用於存放各判決編號

```
##網頁真正網址
each_url=[]
for r in case_num:
    url="https://www.lawplus.com.tw/#reportDetail?id="+r
    each_url.append(url)
for i in each_url:
    print(i)
```



- ●each_link 這個list用於存放爬蟲所需網頁網址
- ●case_num這個list用於存放各判決編號

```
### 爬蟲爬取的網址
each_link=[]
for r in case_num:
    each="https://www.lawplus.com.tw/rest/search/report/"+r
    each_link.append(each)
for i in each_link:
    print(i)
```

</>>

- ●以迴圈方式重複執行獲得每個判決資料
- ●sleep(1) 避免被網頁阻擋爬蟲

```
content=[]
for i in each link:
    res=requests.get(i)
    resquests=json.loads(res.text)
### dictionary存放
    content.append({
    '日期':resquests['response']['reportBase']['identifier'].split(',')[4
    '案號':resquests['response']['reportBase']['caseNum'],
    '內文':resquests['response']['reportBase']['content'],
    '法規':resquests['response']['report']['statute'],
    '法官':resquests['response']['report']['presideJudge'],
    '裁判類型':resquests['response']['report']['type'],
    '案由':resquests['response']['reportBase']['issue']
    sleep(1)
```

</>>

- res=resquests.get(i)
- reaquests=json.loads(res.text)
- ●resquests此行用來查看所需擷取內容

```
In [34]: resquests
Out[34]: {'lastError': '',
        'response': {'reportBase': {'caseNum': '107.中簡.534'.
         'caseTvpe': 'V'.
         'codeNum': '中簡',
         'content': '臺灣臺中地方法院民事判決\u3000\u3000\u3000\u3000\u3000\u3000107年度中簡字第534號\n原告\u3000王清輝\n被告\u300
      0張家豪\n上列當事人間損害賠償事件,本院於民國107年3月27日言詞辯論\n終結,判決如下:\n
       自民國107年1月19日起至清\n償日止,按年息5%計算之利息。\n原告其餘之訴駁回。\n訴訟費用新臺幣342元由被告負擔,餘由原告負擔。\n本判決原告
                       事實及理由\n一、原告主張:被告於民國106年1月6日0時29分許,在臺中市旅\n
      有車輛(下稱系爭\n
                     車輛)受損,支出修理費用新臺幣(下同)158450元。又系爭\n
                                             日起至清償日止,按年息5%計算之利息。\n二、被告則以:我有原告車子的照片
      部分沒有修,但估價\n
                       單有等語置辯,請求駁问原告之訴。\n三、原告主張之事實,業據其提出道路交通事故現場圖、估價單\n
      件為證,且經調閱本件臺中市政府警察局道\n
      狀態,均為可能之肇\n
```



- ●以append方式並更改為字典格式存取內容
- ●content此行用來查看擷取內容

In [35]: content

Out[35]: [{'日期': '20161124',

'案號': '105,投小,319',

'內文': '臺灣南投地方法院民事小額判決\u3000\u3000\u3000\u3000\u3000105年度投小字第319號\n原\u3000\u3000\u3000\u3000告\u3000 股份有限公司\n法定代理人\u3000李松季\n訴訟代理人\u3000陳建興\n\u3000\u3000\u3000\u3000\u3000\u3000\u3000\$松麟\n被\u3000\u3000\u3000\u30 00告\u3000陳慶降\n訴訟代理人\u3000劉道明\n\u3000\u3000\u3000\u3000\u3000\u3000面到面上列當事人間損害賠償事件,本院於民國105 年11月3日言詞辯論\n終結,茲判決如下:\n 文\n被告應給付原告新臺幣陸萬參仟肆佰伍拾貳元,及自民國一百零\n五年六月三十日起至清 償日止按週年利率百分之五計算之利息。\n原告其餘之訴駁回。\n訴訟費用新臺幣壹仟元由被告負擔新臺幣捌佰參拾元,餘由原告\n負擔。\n本判決第一項 事實及理由\n一、原告起訴主張略以:原告承保訴外人陳頤瑄所有並由其駕駛\n 之車號000-0000號自用小客車(下稱系爭A 車 3 年12月7 日9 時許自西向東行駛至南投縣∞鄉∞巷00\n 號前時(下稱系爭時地),被告於同一時、地駕駛車號0000\n ,於民國10\n 彎車未禮讓百行車及未盡車前注意義務,致撞擊系爭A 車,\n 0 號自用小客車(下稱系爭B車)欲由北向西右轉,因轉\n A 車維修費用新臺幣(下同) 109,218 元(零件56,337元、\n 工資39,316元、塗裝13,565元)。本件 駛系爭A 車雖有未依規定減速、未盡車前注意義務之過失,\n 惟被告駕駛系爭B 車亦有右轉未依規定、未盡車前注意義務 之過失存在,且系爭B 車之過失比例應屬較高,因此原告與\n 被告間之過失比例應為3 成、7 成等語。爰依保險法第53條\n 第184 條第1 項前段、第191 條之2 規定提起\n 本件訴訟等語。並聲明:被告應給付原告76,453元,及自起\n 按调年利率5%計算之利息。\n二、被告則以:被告不爭執就本件車禍事故有過失責任存在,惟\n 意義務之疏失,因此雙方過失比例各應各為5 成等詞。並聲\n 告之訴駁回。\n三、本院之判斷:\n \uf67e被告應就本件車禍事故負損害賠償責任:\n 1. 按因故意或過失,不法侵害他人之權利者,負損害賠償責任 ; 汽車、機車或其他非依軌道行駛之動力車輛, 在使用中加\n 損害於他人者,駕駛人應賠償因此所生之損害。民法第184\n



- ●利用pandas將資料轉為dataframe(雙維度)
- ●將資料存為cs∨檔
- ●「lawcontent.csv」會和你存放此ipynb檔相同位置

```
### 資料輸出

df = pd.DataFrame(content)

df.to_csv('lawcontent.csv')
```

●df 此行用來查看結果



7]: d	f							
l: _	日期	案號		內文	法規	法官	裁判 類型	案由
	0 20161124	105,投 小,319	臺灣南投地方法院民事小額判決 \n原	105年度投小字第319號 告 泰安產物保險股…	[民法第213條, 民法第184條第1項, 民事訴訟法第436條 之8第1項, 民法第196條	張 家 豪	民事 判決	損害賠償
	1 20180831	107, 訴,1641	臺灣桃園地方法院民事判決 \n原	107年度訴字第1641號 告 新光產物保險	[民法第203條, 民法第196條, 民法第196條第1項, 民法第 233條第1項, 道路交	廖珮伶	民事 判決	損害賠償
	2 20170811	106,士 小,787	臺灣士林地方法院小額民事判決 \n原	106年度士小字第787號 告 新安東京海上產…	[民法第184條第1項,民法第196條,民事訴訟法第436條 之24第2項,民事訴訟法第…	張 明 儀	民事 判決	侵權行為損 害賠償
	3 20160808	105,湖 小,595	臺灣士林地方法院小額民事訴訟判決 \n原	105年度湖小字第595號 告 旺旺友聯產物保	[民法第196條,營利事業所得稅查核準則第95條第6項,民 法第217條第1項,民法第2	黃紀錄	民事 判決	損害賠償
	4 20161124	105,投 小,392	臺灣南投地方法院民事小額判決 \n原	105年度投小字第392號 告 新光產物保險股…	[民法第213條, 民法第184條第1項, 民事訴訟法第436條 之8第1項, 民法第196條	張家豪	民事 判決	損害賠償

練習時間



各組試著使用爬蟲蒐集所需資料

(網站可為裁判家,或自行選擇)