

計算機概論期末報告

# 以爬蟲程式與tkinter

實作

# 期刊論文引注格式

司法三 B08103013 林崇堯

法學三 B08A01164 李易儒

ONE

# 動機與目的

某天，我們滿懷剛的意志

在整理民事訴訟法筆記時



感慨於  
要查找的文章  
太多.....

ONE

# 動機與目的

做註腳整理時，想到老師上課教的爬蟲與視窗程式

於是挑選了

月旦知識庫 來爬 期刊論文  
華藝圖書館 來爬 學位論文

格式部分採用

《臺大法學論叢》（新修）格式範本

# 《臺大法學論叢》（新修）格式範本

## 期刊論文

(5) 期刊論文：王皇玉（2004），〈論施用毒品行為之犯罪化〉，《臺大法學論叢》，33 卷 6 期，頁 39-56。

## 學位論文

(6) 學位論文：詹森林（1984），《物之抽象使用利益之損害賠償》，頁 30，國立臺灣大學法律學研究所碩士論文。

說明：論文出處單位請用全稱（需包括校及院／系名稱）

TWO 方法與素材

Step 1.

爬蟲的方法

Step 2.

作成視窗的方法  
(tkinter)



# Step 1. 爬蟲：月旦

 月旦知識庫 文獻檢索  

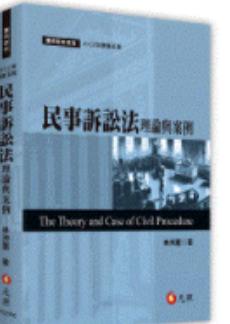
熱門：憲法訴願 公司法修正 洗錢防制 人工智慧與大數據 大法庭

首頁 | 臺灣期刊 法律 公行政治 醫學 財經 社會學 教育 其他 | 大陸期刊 核心 非核心 | DOI文章

首頁 > 臺灣期刊 > 法律 > 月旦法學雜誌 > 202112 (319期)

 查看詳細全文

篇名	廢棄物清理法非法清理廢棄物罪構成要件之解釋與適用——簡評最高法院109年度台上大字第3338號刑事裁定
作者	徐育安
中文摘要	大法庭裁定認為，廢棄物清理法第46條第4款非法清理廢棄物罪之犯罪主體，所謂未得許可而從事廢棄物清理者，毋須以清理廢棄物作為其業務者為限，亦即毋庸為反覆為之者，若僅偶然為之者，亦屬適格的犯罪主體。本文從本罪在我國的發展軌跡出發，亦即以歷史解釋及文義解釋為起始點，並從規範目的、體系解釋的角度，並參酌比較法，逐步探討對於本罪主體所應採取的解釋為何。本文認為，應連結本罪與同法第41條第1項及其相關規定，將未得許可而從事廢棄物清理者，限縮在抽象危險犯的框架之下，將其理解為受委託清理廢棄物為業或具有同等營運規模者，此種行為人非法清理之數量已達到環境無法輕易承受的程度，其危險性方與第46條其他幾款犯罪相當。

 新書閱讀 

民事訴訟法理論與案例  
林洲富

- 聯合行為要件之評析及競爭法與智財法之交錯 / 廖義男、黃銘傑 主編
- 研之得法——中央研究院法律學研究所成立十週年文集 / 李建良

起訟直擊：民商事訴訟實務研究

# Step 1. 爬蟲：月旦網頁原始碼

```
自動換行   
1 <HTML>  
2   <HEAD>  
3  
4     <TITLE>我國專利權無效雙軌制之互動與調和—以「一事不再理」法效之詮釋為中心- 月旦知識庫</TITLE>  
5     <META http-equiv=Content-Language content=zh-tw>  
6     <META http-equiv=Content-Type content="text/html; charset=utf-8">  
7     <meta name="title" content="我國專利權無效雙軌制之互動與調和—以「一事不再理」法效之詮釋為中心- 月旦知識庫?">  
8     <META name="description" content="沈宗倫,雙軌制,專利權,複審,禁反言,一事不再理,舉發,無效抗辯,近來,專利法進行重大修正,計畫將專利法之舉發法制朝向「爭議對審」方向調整,在程序上強化專利權效力挑戰者與專利權人就專利,月旦知識庫,整合十大資料庫交叉檢索搜尋,是法律學者,實務工作者,法律學子學習好幫手!">  
9     <META name="keywords" content="沈宗倫,雙軌制,專利權,複審,禁反言,一事不再理,舉發,無效抗辯,tssci,月旦法學,月旦知識庫,裁判時報,大陸法學,法律詞典,法學資料庫,法規,判解,教學,論文,法律題庫">  
10    <meta name="robots" content="index,follow">  
11  
12  
13    <meta name="citation_title" content="我國專利權無效雙軌制之互動與調和—以「一事不再理」法效之詮釋為中心">  
14    <meta name="citation_author" content="沈宗倫">  
15  
16    <meta name="citation_publication_date" content="2022/01/01">  
17    <meta name="citation_publisher" content="元照出版公司">  
18    <meta name="citation_journal_t??title" content="月旦法學雜誌">  
19    <meta name="citation_issue" content="320">  
20  
21    <meta name="citation_firstpage" content="43">  
22    <meta name="citation_lastpage" content="60">  
23  
24    <link href=".file/css/default.css" rel="stylesheet" type="text/css">  
25    <script src=".file/js/default.js" type="text/javascript"></script>  
26  </HEAD>  
27  <BODY>  
28    <div id="BodyContainer">  
29      <div id="BodyHeader"><META HTTP-EQUIV="PRAGMA" CONTENT="NO-CACHE">  
30      <!-- IE可能不見得有效 -->  
31      <META HTTP-EQUIV="EXPIRES" CONTENT="0">  
32      <!-- 請勿刪除此行，以免影響IE的執行 -->
```

# Step 1. 爬蟲內容

```
<meta name="citation_title" content="我國專利權無效雙軌制之互動與調和—以「一事不再理」法效之詮釋為中心">
<meta name="citation_author" content="沈宗倫">

<meta name="citation_publication_date" content="2022/01/01">
<meta name="citation_publisher" content="元照出版公司">
<meta name="citation_journal_title" content="月旦法學雜誌">
<meta name="citation_issue" content="320">

<meta name="citation_firstpage" content="43">
<meta name="citation_lastpage" content="60">
```

# Step 1.

## 程式撰寫

```
def r_news():
    form = form_entry.get()
    url = form
# webpage 可以自己取 總之只要把檔案抓下來
    r = requests.get(url)
    soup = BeautifulSoup(r.content, "html.parser")
# 用 html5lib 這個解析器去解析上面抓到的網址

    #meta1=文章名稱
    meta1 = soup.findAll(attrs={'name':'citation_title'})
    #meta2=作者名稱
    meta2 = soup.findAll(attrs={'name':'citation_author'})
    #meta3=時間
    meta3 = soup.findAll(attrs={'name':'citation_publication_date'})
    #meta4=出版公司
    meta4 = soup.findAll(attrs={'name':'citation_publisher'})
    #meta5=雜誌
    meta5 = soup.findAll(attrs={'name':'citation_journal_title'})
    #meta6=雜誌期數
    meta6 = soup.findAll(attrs={'name':'citation_issue'})
    #meta7=起始頁數
    meta7 = soup.findAll(attrs={'name':'citation_firstpage'})
    #meta8=終止頁數
    meta8 = soup.findAll(attrs={'name':'citation_lastpage'})

    content = meta2[0]["content"] + "(" + meta3[0]["content"][0:4] + ")" + " , "
    text = tk.Text(window, height=2, borderwidth=0)
    text.insert(1.0, content)
    text.pack()
    text.configure(state="disabled")

get_news = tk.Button(new_frame, text='期刊引注格式查詢', command=r_news)
```

# Step 1.

## 程式撰寫 (進階)

```
<tr>
    <th nowrap="nowrap">刊名</th>
    <td><a href="journal_list.aspx?no=141" id="SearchItem">月旦法學雜誌</a>
</tr>
<tr>
    <th nowrap="nowrap">期數</th>
    <td><a href="journal.aspx?no=141&pno=70759">202201 (320期)</a></td>
</tr>

<tr>
    <th nowrap="nowrap">出版單位</th>
    <td>
        <a href="http://www.angle.com.tw/" target="_blank">
        元照出版公司
        </a>
    </td>
</tr>
```

# Step 1.

## 程式撰寫 (進階)

```
import requests
from bs4 import BeautifulSoup
url = "http://lawdata.com.tw/tw/detail.aspx?no=902496"
r = requests.get(url)
soup = BeautifulSoup(r.content, "html.parser")
# 因為我們要找的資料是
# <tr>
#   <th nowrap="nowrap">起訖頁</th>
#   <td>6-24</td>
# </tr>
# 我們知道他藏在 tr 標籤裡面
# 然後我們會用到 th 和 td
# 所以我們直接找在程式碼中，所有的 tr 標籤，再慢慢比對
table_datas = soup.findAll("tr")
# 這是一種 for 迴圈的用法，可以跑過陣列裡面所有的資料
for table_data in table_datas:
    # 這行是為了怕說有 tr 標籤中，沒有 th 標籤
    # 如果不加上這行或出現錯誤
    if(table_data.th != None):
        # 既然我們的目標是找到 "起訖頁"
        # 那我們可以直接找 th 裡面文字是 "起訖頁" 的就好了
        if table_data.th.text == "起訖頁":
            print(f"{table_data.th.text}: {table_data.td.text}")
            break
```

# Step 1. 爬蟲：華藝線上圖書館



# Step 1. 爬蟲：華藝線上圖書館原始碼

```
<!DOCTYPE html>
<html>
<head>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
    <meta content="期刊, 論文, 期刊論文, 中文, 電子期刊, 華藝, CEPS, 中文期刊, 學術期刊, 期刊文章, 文章, 全文下載, 論文下載, CETD, 碩博士論文, 博碩士論文, 學位論文, 會議論文, 研討會, GCScholar, GCS, 華文, 電子全文, PDF, 免費PDF, 免費全文, 學術文獻, 圖書館, 電子資料庫, 全文資料庫, 資料庫, Airiti, article, 台大論文, 國家圖書館, 思博網, 查資料, 國圖, Journal, Airiti Library, Library, 華藝線上圖書館, DOI, DOI申請, DOI註冊 " name="keywords" />
    <meta content="Airiti Library華藝線上圖書館" name="description" />
    <meta http-equiv="imagetoolbar" content="no" />
    <meta http-equiv="Expires" content="Wed, 26 Feb 1950 08:21:57 GMT" />
    <meta http-equiv="Pragma" content="no-cache" />
    <meta name="robots" content="nofollow,noarchive" />
    <title>Airiti Library華藝線上圖書館</title>
    <link id="shortcut_icon" type="image/x-icon" href="/Content/images/favicon.ico" rel="shortcut icon" />

    <meta name="apple-itunes-app" content="app-id=1486515085" />
```

# Step 1.

## 程式撰寫

```
def r_othernews():
    form = form_entry.get()
    url = form
    r = requests.get(url)
    soup = BeautifulSoup(r.content, "html.parser")

    #meta1=文章名稱
    meta1 = soup.findAll(attrs={'name':'citation_title'})
    #meta2=作者名稱
    meta2 = soup.findAll(attrs={'name':'citation_author'})
    #meta3=時間
    meta3 = soup.findAll(attrs={'name':'citation_issue'})
    #meta4=出版公司
    meta4 = soup.findAll(attrs={'name':'citation_publisher'})

    content = meta2[0]["content"] + "(" + meta3[0]["content"] + ")" + " , " +
    text = tk.Text(window, height=2, borderwidth=0)
    text.insert(1.0, content)
    text.pack()
    text.configure(state="disabled")
```

# Step 1.

## 程式撰寫

```
def r_othernews():
    form = form_entry.get()
    url = form
    r = requests.get(url)
    soup = BeautifulSoup(r.content, "html.parser")

    #meta1=文章名稱
    meta1 = soup.findAll(attrs={'name':'citation_title'})
    #meta2=作者名稱
    meta2 = soup.findAll(attrs={'name':'citation_author'})
    #meta3=時間
    meta3 = soup.findAll(attrs={'name':'citation_issue'})
    #meta4=出版公司
    meta4 = soup.findAll(attrs={'name':'citation_publisher'})

    content = meta2[0]["content"] + "(" + meta3[0]["content"] + ")" + " , " +
    text = tk.Text(window, height=2, borderwidth=0)
    text.insert(1.0, content)
    text.pack()
    text.configure(state="disabled")
```

# Step 2.

## 視窗 tkinter

```
import tkinter as tk
import requests
from bs4 import BeautifulSoup

window = tk.Tk()

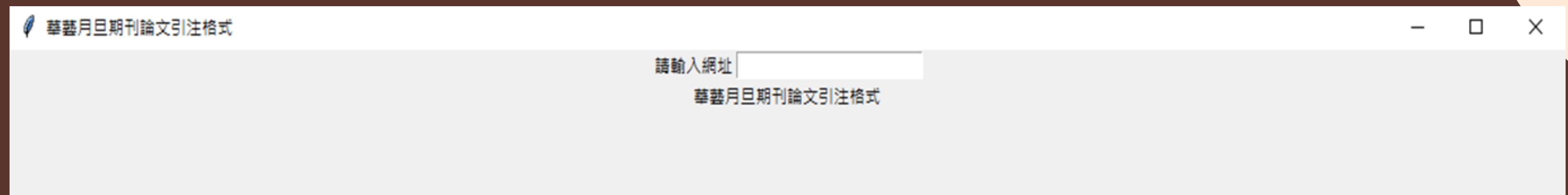
window.title('華藝月旦期刊論文引注格式')
window.geometry('1200x500')

new_frame = tk.Frame(window)
new_frame.pack(side = 'bottom')

form_frame = tk.Frame(window)
form_frame.pack(side=tk.TOP)
form_label = tk.Label(form_frame, text='請輸入網址')
form_label.pack(side=tk.LEFT)
form_entry = tk.Entry(form_frame)
form_entry.pack(side=tk.LEFT)

header_label = tk.Label(window, text='華藝月旦期刊論文引注格式')
header_label.pack()
```

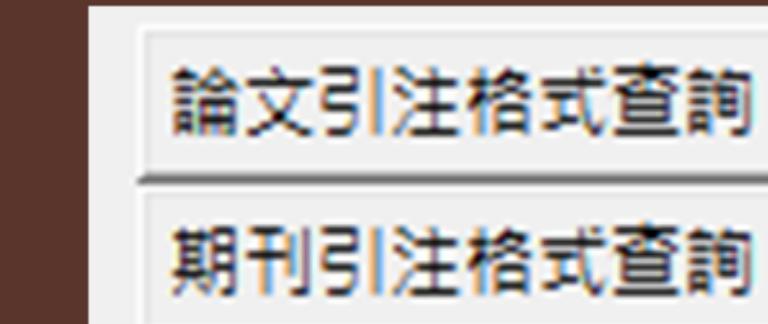
# Step 2. 視窗 tkinter



## Step 2. 視窗 tkinter

```
get_news = tk.Button(new_frame, text='期刊引注格式查詢', command=r_news)
get_news.pack(side = 'bottom')
```

```
get_news = tk.Button(new_frame, text='論文引注格式查詢', command=r_othernews)
get_news.pack(side = 'bottom')
```



# Step 2.

## 視窗 tkinter

```
def r_news():
    form = form_entry.get()
    url = form
#wbpage可以自己取 總之只要把檔案抓下來
    r = requests.get(url)
    soup = BeautifulSoup(r.content, "html.parser")
#用html5lib這個解析器去解析上面抓到的網址

    #meta1=文章名稱
    meta1 = soup.findAll(attrs={'name':'citation_title'})
    #meta2=作者名稱
    meta2 = soup.findAll(attrs={'name':'citation_author'})
    #meta3=時間
    meta3 = soup.findAll(attrs={'name':'citation_publication_date'})
    #meta4=出版公司
    meta4 = soup.findAll(attrs={'name':'citation_publisher'})
    #meta5=雜誌
    meta5 = soup.findAll(attrs={'name':'citation_journal_t??title'})
    #meta6=雜誌期數
    meta6 = soup.findAll(attrs={'name':'citation_issue'})
    #meta7=起始頁數
    meta7 = soup.findAll(attrs={'name':'citation_firstpage'})
    #meta8=終止頁數
    meta8 = soup.findAll(attrs={'name':'citation_lastpage'})
```

# Step 2. 視窗 tkinter

```
content = meta2[0]["content"] + "(" + meta3[0]["content"][0:4] + ")" + ", " + "<" +
(meta1[0]["content"]) + "> " + ", " + "《" + (meta5[0]["content"]) + "》 " + ", " +
meta6[0]["content"] + '期' + ", " + "頁" + meta7[0]["content"] + "-"
+meta8[0]["content"] + "。"
```

```
text = tk.Text(window, height=2, borderwidth=0)
text.insert(1.0, content)
text.pack()
text.configure(state="disabled")
get_news = tk.Button(new_frame, text='期刊引注格式查詢', command=r_news)
get_news.pack(side = 'bottom')
```

# Step 2.

## 視窗 tkinter

```
def r_othernews():
    form = form_entry.get()
    url = form
    r = requests.get(url)
    soup = BeautifulSoup(r.content, "html.parser")

#meta1=文章名稱
meta1 = soup.findAll(attrs={'name':'citation_title'})
#meta2=作者名稱
meta2 = soup.findAll(attrs={'name':'citation_author'})
#meta3=時間
meta3 = soup.findAll(attrs={'name':'citation_issue'})
#meta4=出版公司
meta4 = soup.findAll(attrs={'name':'citation_publisher'})
```

# Step 2. 視窗 tkinter

```
content = meta2[0]["content"] + "(" + meta3[0]["content"] + ")" + ", " + "《" +  
(meta1[0]["content"]) + "》 " + ", " + "頁" + ", " + meta4[0]["content"] + '法律學  
研究所學位論文' + "。 "
```

```
text = tk.Text(window, height=2, borderwidth=0)  
text.insert(1.0, content)  
text.pack()  
text.configure(state="disabled")  
get_news = tk.Button(new_frame, text='論文引注格式查詢', command=r_othernews)  
get_news.pack(side = 'bottom')
```

```
window.mainloop()
```

THREE

# 結果：作品呈現



THREE

# 結果：作品呈現

The screenshot shows a search results page for legal citations. At the top, there is a header with a magnifying glass icon and the text "華藝月旦期刊論文引注格式". Below the header is a search bar containing the URL "/detail.aspx?no=902723". The main content area displays two entries:

李佳樺(2021年),《跨國關係企業移轉訂價稅制之研究——以集團內部服務費為中心》,頁  
,國立臺灣大學法律學研究所學位論文。

李素華(2022),〈初探專利法修正草案之兩造對審制〉,《月旦法學雜誌》,320期,頁25-  
42。

At the bottom right of the page, there are two buttons: "論文引注格式查詢" and "期刊引注格式查詢".



ONE

<HTTP://LAWDATA.COM.TW/TW/DETAIL.ASPX?NO=902723>

TWO

<HTTPS://WWW.AIRITILIBRARY.COM/PUBLICATION/ALDETAILEDMESH1?DOCID=U0001-3008202122214400>

THREE

# 結果：作品呈現

【實測網址】

# 效用實測 | 開啟十個期刊論文 紀錄受試者完成十組引注格式所需的時間

	使用程式	未使用程式	所費時間差 (只取整數)
測試者A	80秒	410秒	5倍
測試者B	77秒	464秒	6倍
測試者C	100秒	502秒	5倍
測試者D	84秒	410秒	4倍

# 結論

1. 達成引注月旦與華藝期刊論文格式之目標
2. 上手容易，僅須貼上網址+按下查詢即可完成
3. 節省大量時間（實測結果）
4. 對須引注之法律系報告均有幫助

# 參考資料

1. 如何利用TKINTER做一個的簡單文字介面  
網址：[HTTPS://REURL.CC/VE32OJ](https://reurl.cc/VE32OJ)
2. 計算機概論WEEK4 TK\_WINDOW例示
3. 計算機概論WEEK4 TK\_BMI例示