

# 建立預測定址空氣汙染數據之自動化平台

國立高雄應用科技大學 資訊工程系

指導教授：鐘文鈺 教授

1104108107 高煌勝

1104108112 王拓翔

1104108113 林子傑

1104108114 蔡信德

## 摘要

隨著時代的演進，人類的科技越來越發達。從以農業為主軸發展到工業，到了現代又因已無法脫離。因為工業的增長，產生了大量的有害物質，以致於大氣中的懸浮粒子日益增加，間接影響了我們的生活品質與健康。為了加強控管，設置了許多觀測站測量空氣中的懸浮粒子。但觀測站的個數與位置均有限，無法達到個人住家的精準量值，因此需由程式進行預測，但由於數據龐大，若使用人工統計實為困難，我們可以透過自動化程式來解決這個問題。

本專題目標：設計一套便於一般民眾及研究者利用與研究的軟體。透過將長年以來的數據匯整於資料庫，並利用R語言加以分析，得出細部座標的空汙數值。一般民眾可以使用這套軟體，便可輕鬆得知自家與工作地點附近的空氣品質。研究人員更可以利用這套軟體，獲得大量的座標的各懸浮微粒的數值，大幅度減少人工查閱以及運算的消耗。

## 研究動機與研究問題

近年來，由於空氣品質問題日益的嚴重，又因為老舊的核電廠逐漸停用，導致重汙染的台中火力發電廠增加機組運作，導致台灣中部以南的空氣汙染持續加重。在這樣的環境之下，開始有許多人也關心並且想要研究空氣汙染及健康的關聯性，並希望向政府提出相關的研究數據，使政府能夠規範降低空氣汙染。但是目前沒有方便的工具使研究人員和一般民眾能輕鬆的查詢自己所在處或是某一區的空氣汙染的數值，我們想到若能夠利用我們科系的專長，幫研究人員建一個簡易的資料庫並且能夠代入地址和日期便可以知道該日期及該地址的空氣汙染值，以便研究人員研究或一般民眾查詢及注意，如此一來就能夠增加研究人員作業上的效率也使得之後可望使政府降低空氣汙染的排放。

首先是資料的彙整，我們發現到要求的格式並沒有固定下來，甚至來源資料的格式都不一，所以導致擷取至資料庫程式要不斷的修改。

## 研究方法及步驟

### 蒐集資料、彙整數據

撰寫Python語言程式自動擷取環保署公開的空氣污染物的觀測測值，並將擷取的資料加入本地端資料庫將資料按照年份、測站、測值分類，以便檢索。

利用Python中的Selenium函式庫加上Chrome WebDriver能夠使程式自動的進入環保署的網站並選取測站、測項以及時間以獲得選取範圍內的資料數據。再利用HTMLParser函式庫將剛剛選取的資料轉為XML文件檔讀取，接著利用Python來判斷文件中的哪一些部分為選取的資料範圍，經整理過後，利用pymysql函式庫來連結MariaDB，便可以將剛剛的擷取並整理好的資料加入進資料庫。

對於相較於早期的資料，環保署有歷年資料的文件可供下載，我們也設計了一個能更有效率將.csv檔及.xls檔匯入的程式。

如此一來，便可將過去幾年來的空汙數據資料存入資料庫做更進一步的利用了。

Unamed\mydb\2018\_仁武 - HeidiSQL 9.4.0.5125

文件 編輯 搜索 工具 Go to 幫助

資料庫: mydb 表: 2018\_仁武

mydb.2018\_仁武: 24 條記錄 (大約)

日期	SO2	CO	O3	PM10	NOx	NO	NO2	THC	NMHC	WIND_SPEED	WIND_DIRECT
2018/03/08/00	2.3	0.35	21	111	19	1.8	17	2.2	0.18	2.2	328
2018/03/08/01	2.3	0.33	20	97	17	0.7	17	2.2	0.14	1.8	318
2018/03/08/02	2.5	0.36	21	89	14	0.5	14	2.2	0.15	3.9	336
2018/03/08/03	4.3	0.36	16	86	20	1.9	18	2.3	0.19	2.7	335
2018/03/08/04	6.2	0.38	12	80	26	3.6	22	2.3	0.26	2.1	323
2018/03/08/05	2.9	0.4	8.5	84	22	2.1	20	2.2	0.16	1.6	326
2018/03/08/06	3	0.45	8.1	77	27	5.3	21	2.3	0.2	2.2	334
2018/03/08/07	4.8	0.56	6.8	84	41	16	25	2.3	0.27	2.4	330
2018/03/08/08	3	0.5	12	72	32	9.2	22	2.2	0.21	4	335
2018/03/08/09	2.6	0.43	18	71	29	8.2	21	2.2	0.19	4.3	331
2018/03/08/10	2.1	0.31	24	57	23	6.4	17	2.1	0.16	3.4	330
2018/03/08/11	2	0.32	27	52	20	5.2	15	2.1	0.14	5.3	338
2018/03/08/12	2.3	0.36	24	40	23	5.8	17	2.1	0.15	5.9	335
2018/03/08/13	2.5	0.37	24	39	23	6.2	17	2.1	0.15	4.1	329
2018/03/08/14	2.2	0.38	19	40	27	7.2	20	2.1	0.19	5.1	336
2018/03/08/15	1.9	0.4	14	40	29	7.4	22	2.2	0.22	4	330
2018/03/08/16	1.7	0.43	12	42	29	5.4	24	2.2	0.21	4.9	335
2018/03/08/17	1.7	0.46	13	43	28	3.5	25	2.2	0.21	3.8	335
2018/03/08/18	1.6	0.46	12	44	28	3.8	24	2.2	0.2	4.7	334
2018/03/08/19	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2018/03/08/20	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2018/03/08/21	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2018/03/08/22	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2018/03/08/23	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

34 SHOW CREATE TABLE `mydb`.`2018\_仁武`;  
 35 SHOW COLLATION;  
 36 SHOW ENGINES;  
 37 SELECT \* FROM `mydb`.`2018\_仁武` LIMIT 1000;  
 38 SHOW CREATE TABLE `mydb`.`2018\_仁武`;

1:1 已連接: 00:04 h MariaDB 10.2.12 運行時間: 10:39 h UTC: 2018-03-08 上午 11:55 空閒

↑上圖為擷取後結果範例，擷取後每一個表格命名規則為"年份\_測站"，表格內的第一列為日期(年/月/日/時)，後面的列為各測項的測值。

## 運算分析

使用R語言程式結合資料庫將希望觀察的時間、地點、測值的資料取出，利用內插法預測出此地的空氣品質。

使用Python中geocoder函式庫便可讓使用者輸入地址便能輕鬆得轉換為經緯度，在找尋附近的測站以內插法的方式計算出結果並顯示出來，目前計算需要一段時間才能完成，有待改善。

## 系統需求

### 1. 安裝Python

打開cmd安裝&取得函式庫:

a.Selenium : pip install selenium

b.取得Chromedriver :

<https://sites.google.com/a/chromium.org/chromedriver/downloads>

c.pymysql:pip install pymysql

d.xlrd:pip install xlrd

e.geocoder:pip install geocoder

### 2. 安裝MariaDB

## 結論

(空)

## 未來展望

**伺服器：**架設雲端伺服器，讓使用者並不局限於單機。

**地圖：**結合google map，讓使用者能夠使用地圖來選擇要觀測的地方，能夠更方便的查詢。

## 參考文獻

行政院環境保護署 - 光化測站：<http://taqm.epa.gov.tw/taqm/tw/PamsDaily.aspx>

行政院環境保護署 - 空氣品質測站：<http://taqm.epa.gov.tw/taqm/tw/HourlyData.aspx>

行政院環境保護署 - 細懸浮微粒(PM2.5)：

<http://taqm.epa.gov.tw/pm25/tw/HourlyData.aspx>

行政院環境保護署 - 地方監測站資料：

<http://taqm.epa.gov.tw/taqm/tw/EpbDataHourly.aspx>