

Python60 系列 地址處理



</PLUS>

推動大學(程式設計)教學
MOE/Ministry of National Education

推動大學程式設計教學計畫。分項六：資料分析領域與學習評量推動團隊（德明財經科大資科系江政杰老師主編）

地址處理：學習內容

▶ 程式設計概念

- 函式與模組
- CSV 格式檔案讀取與寫入
- 字典
- 字串處理

▶ 資料分析概念

- 搜尋與比對
- 資料替換

地址處理

▶ 程式思考

```
# 開啟資料輸入檔
# 讀取輸入檔內容
# 針對每一筆客戶的資料
#     檢查縣市與住址並且修正

# 開啟資料輸出檔
# 針對每一筆的修正後資料
#     寫入資料到檔案
```

寫程式前請先想想要做甚麼事!!

	A	B	C
1	姓名	縣市	住址
2	陳大明	台北市	汀州路四段88號
3	張小華	臺北市	和平東路一段162號
4	李國華	台北市	復興南路一段6號8樓
5	朱志清	台中市	大仁路38巷4號5F
6	朱婷婷	台中市	中港路29號2樓
7	周杰倫	台南市	中港路四段390-1號
8	劉德明	高雄市	三多路49巷2弄1號14F
9	紀小月	台東市	中台五路95號



	A	B	C	D
1	姓名	縣市	住址	
2	陳大明	臺北市	汀州路四段88號	
3	張小華	臺北市	和平東路一段162號	
4	李國華	臺北市	復興南路一段6號8樓	
5	朱志清	臺中市	大仁路38巷4號5樓	
6	朱婷婷	臺中市	臺灣大道29號2樓	
7	周杰倫	臺南市	中港路四段390-1號	
8	劉德明	高雄市	三多路49巷2弄1號14樓	
9	紀小月	臺東市	中台五路95號	

地址處理

完整程式展示

```
import csv

with open('地址資料.csv') as infile:
    data = list(csv.DictReader(infile))
    for e in data:
        print('原始資料:', e['姓名'], e['縣市'], e['住址'])
        if e['縣市'][0] == '台':
            e['縣市'] = '臺'+e['縣市'][1:]
        if 'F' in e['住址']:
            e['住址'] = e['住址'].replace('F', '樓')
        if e['縣市'] == '臺中市' and '中港路' in e['住址']:
            e['住址'] = e['住址'].replace('中港路', '臺灣大道')

        print('    更新資料:', e['姓名'], e['縣市'], e['住址'])

with open('新地址資料.csv', 'w', newline='') as outfile:
    writer = csv.DictWriter(outfile, fieldnames = data[0].keys())
    writer.writeheader()
    for e in data:
        writer.writerow(e)
```

輸入檔

	A	B	C
1	姓名	縣市	住址
2	陳大明	台北市	汀州路四段88號
3	張小華	臺北市	和平東路一段162號
4	李國華	台北市	復興南路一段6號8樓
5	朱志清	台中市	大仁路38巷4號5F
6	朱婷婷	台中市	中港路29號2樓
7	周杰倫	台南市	中港路四段390-1號
8	劉德明	高雄市	三多路49巷2弄1號14F
9	紀小月	台東市	中台五路95號

程式執行輸出畫面

```
原始資料: 陳大明 台北市 汀州路四段88號
更新資料: 陳大明 臺北市 汀州路四段88號
原始資料: 張小華 臺北市 和平東路一段162號
更新資料: 張小華 臺北市 和平東路一段162號
原始資料: 李國華 台北市 復興南路一段6號8樓
更新資料: 李國華 臺北市 復興南路一段6號8樓
原始資料: 朱志清 台中市 大仁路38巷4號5F
更新資料: 朱志清 臺中市 大仁路38巷4號5樓
原始資料: 朱婷婷 台中市 中港路29號2樓
更新資料: 朱婷婷 臺中市 臺灣大道29號2樓
原始資料: 周杰倫 台南市 中港路四段390-1號
更新資料: 周杰倫 臺南市 中港路四段390-1號
原始資料: 劉德明 高雄市 三多路49巷2弄1號14F
更新資料: 劉德明 高雄市 三多路49巷2弄1號14樓
原始資料: 紀小月 台東市 中台五路95號
更新資料: 紀小月 臺東市 中台五路95號
>>> |
```

程式碼解說

先理解整個程式的大致流程，後面再逐步閱讀細節。

```
01 import csv
02
03 with open('地址資料.csv') as infile:
04     data = list(csv.DictReader(infile))
05     for e in data:
06         print('原始資料:', e['姓名'], e['縣市'], e['住址'])
07         if e['縣市'][0] == '台':
08             e['縣市'] = '臺'+e['縣市'][1:]
09         if 'F' in e['住址']:
10             e['住址'] = e['住址'].replace('F', '樓')
11         if e['縣市'] == '臺中市' and '中港路' in e['住址']:
12             e['住址'] = e['住址'].replace('中港路', '臺灣大道')
13
14     print('    更新資料:', e['姓名'], e['縣市'], e['住址'])
15
16 with open('新地址資料.csv', 'w', newline='') as outfile:
17     writer = csv.DictWriter(outfile, fieldnames = data[0].keys())
18     writer.writeheader()
19     for e in data:
20         writer.writerow(e)
```

匯入方便讀寫 csv 檔的模組

讀取地址資料後，
逐筆顯示檢查並修正。

將資料逐筆寫入輸出檔。

程式概念：函式與模組

- ▶ 可以想像函式是程式中一個特定功能的零件，可執行諸如「列印」、「開啟檔案」、「計算平均數」等功能。
- ▶ Python 語言方便之處在於它擁有大量的函式套件與模組。
 - 函式模組就像是工具箱，內含一系列相關功能的零件（函式）。
 - 使用函式庫之前，需要先安裝並匯入，例如

```
import csv
```

會匯入 `csv` 這個函式庫，裡面會有方便讀寫 **CSV** 檔案的函式，例如 `DictReader()`。

```
data = list(csv.DictReader(infile))
```

程式碼解說

```
01 import csv
02
03 with open('地址資料.csv') as infile:
04     data = list(csv.DictReader(infile))
05     for e in data:
06         print('原始資料:', e['姓名'], e['縣市'], e['住址'])
07         if e['縣市'][0] == '台':
08             e['縣市'] = '臺'+e['縣市'][1:]
09         if 'F' in e['住址']:
10             e['住址'] = e['住址'].replace('F', '樓')
11         if e['縣市'] == '臺中市' and '中港路' in e['住址']:
12             e['住址'] = e['住址'].replace('中港路', '臺灣大道')
13
14     print('    更新資料:', e['姓名'], e['縣市'], e['住址'])
```

- ▶ 第 01 行匯入 csv 模組，以便第 04 行可使用模組內的 DictReader() 函式。
- ▶ 第 03 行使用 open() 函式開啟「地址資料.csv」這個檔案，並以 infile 變數代表之。
 - 此處使用到 with 敘述，它可以自動關閉檔案。我們不深究細節，只要知道在哪裡填入檔名和變數名即可。
- ▶ 第 04 行使用 csv 模組內的函式 DictReader()，從 infile 代表的檔案中讀出資料；接著把資料轉換成 list 型態，最後以 data 變數代表之，以接續 05-14 行的操作。
 - 這裡我們看到一程式內可以包含讀取檔案資料、資料型態轉換和指定變數三個動作。在撰寫這樣複合的程式命令時，要注意資料流動的順序：檔案資料 → list 型態資料 → data 變數。

程式概念：字典

```
01 e2c = {'apple': '蘋果', 'book': '書本', 'program': '程式'}
02 c2e = {'蘋果': 'apple', '書本': 'book', '程式': 'program'}
03 print(e2c['apple'], c2e['書本'])
04
05 chen = {'姓名': '陳大明', '縣市': '台北市'}
06 chu = {'姓名': '朱志清', '縣市': '台中市'}
07 chou = {'姓名': '周杰倫', '縣市': '台南市'}
08 data = [chen, chu, chou]
09 for p in data:
10     print(p['姓名'])
```

蘋果 book
陳大明
朱志清
周杰倫

- ▶ Python 中有一種資料型態稱為字典，顧名思義，字典可以儲存多組字串的對應關係。
- ▶ 上面程式的第 01 行建立了一個（英漢）字典 e2c，這個字典記錄了 'apple' 對應到 '蘋果'，'book' 對應到 '書本'，'program' 對應到 '程式'。第 02 行建立一個（漢英）字典 c2e。
- ▶ 第 03 行示範如何取出字典中對應的結果，例如 e2c['apple'] 會取到 '蘋果'。
- ▶ 第 05-07 行建立了三個字典，代表三個人的資料。第 08 行我們以這三個字典建立一個列表 data。第 09-10 行則使用了 for ... in ... 句型取出列表中的每個字典並列印每個字典中 '姓名' 對應的值。

程式概念：字典

```
01 chen = {'姓名': '陳大明', '縣市': '台北市'}
02 for k in chen.keys():
03     print(k)
04 for v in chen.values():
05     print(v)
06
07 for k in chen:
08     print(k, chen[k])
```

```
姓名
縣市
陳大明
台北市
姓名 陳大明
縣市 台北市
```

- ▶ 除了以 `d[k]` 取得字典 `d` 中 `k` 對應的值之外，字典還可以用 `keys()` 取得所有的鍵和 `values()` 取得所有對應到的值。
- ▶ 我們可以用 `for ... in ...` 句型逐一取出鍵或值，如第 **02-05** 行所示。
- ▶ 當我們以 `for ... in ...` 句型直接取用字典時，意思就如同取用字典的鍵。因此，第 **07** 行我們逐一取出字典 `chen` 的鍵，然後第 **08** 行列印鍵 `k` 以及 `chen` 中 `k` 對應到的值。

程式碼解說

```
01 import csv
02
03 with open('地址資料.csv') as infile:
04     data = list(csv.DictReader(infile))
05     for e in data:
06         print('原始資料:', e['姓名'], e['縣市'], e['住址'])
07         if e['縣市'][0] == '台':
08             e['縣市'] = '臺'+e['縣市'][1:]
09         if 'F' in e['住址']:
10             e['住址'] = e['住址'].replace('F', '樓')
11         if e['縣市'] == '臺中市' and '中港路' in e['住址']:
12             e['住址'] = e['住址'].replace('中港路', '臺灣大道')
13
14     print('    更新資料:', e['姓名'], e['縣市'], e['住址'])
```

	A	B	C
1	姓名	縣市	住址
2	陳大明	台北市	汀州路四段88號
3	張小華	臺北市	和平東路一段162號
4	李國華	台北市	復興南路一段6號8樓
5	朱志清	台中市	大仁路38巷4號5F
6	朱婷婷	台中市	中港路29號2樓
7	周杰倫	台南市	中港路四段390-1號
8	劉德明	高雄市	三多路49巷2弄1號14F
9	紀小月	台東市	中台五路95號

- ▶ 第 04 行建立的列表 data 和前頁程式中的 data 相近，都是由多個字典物件構成。這裡我們透過 csv.DictReader() 函式把資料檔中的每一行都讀成一個字典。
- ▶ 第 05 行使用 for ... in ... 句型取出 data 中的每個字典，以 e 代表之；第 06-14 行則對每個字典作同樣的操作。
- ▶ 第 06 行顯示每個字典中 '姓名'、'縣市' 和 '住址' 對應的值，第 07-12 行依之前說過的三個規則處理，第 14 行再次顯示修正後的資料內容。

程式概念：字串處理

```
01 s = '程式設計'  
02 print(s, s[0])  
03 print(s[1:3], s[1:], s[:3], s[-2:])  
04 city = '台'+ '北市'  
05 if city[0]=='台':  
06     city = '臺'+city[1:]  
07 print(city)  
08 if '北' in city:  
09     city = city.replace('北', '南')  
10 print(city)
```

程式設計 程
式設 式設計 程式設 設計
臺北市
臺南市

- ▶ 這次的主題有一個重點是字串處理，這裡我們來看看如何在 Python 中進行字串的串接以及搜尋。
- ▶ 第 01 行我們建立字串 s，第 02-03 行則示範如何取出字串的單一字元和子字串。
 - 我們在 Python60 系列的單元 II 也作過這樣的操作。
- ▶ 第 04 行我們示範運用 + 運算子來把兩個字串串接起來。第 05-07 行則示範如何將 '台北市' 改為 '臺北市'。
- ▶ 第 08 行示範運用 in 運算判斷字串中是否存在指定的子字串，第 09 行示範運用字串的 replace() 操作。

程式碼解說

```
03 with open('地址資料.csv') as infile:
04     data = list(csv.DictReader(infile))
05     for e in data:
06         print('原始資料:', e['姓名'], e['縣市'], e['住址'])
07         if e['縣市'][0] == '台':
08             e['縣市'] = '臺'+e['縣市'][1:]
09         if 'F' in e['住址']:
10             e['住址'] = e['住址'].replace('F', '樓')
11         if e['縣市'] == '臺中市' and '中港路' in e['住址']:
12             e['住址'] = e['住址'].replace('中港路', '臺灣大道')
13
14     print('    更新資料:', e['姓名'], e['縣市'], e['住址'])
```

```
{'姓名': '陳大明',
'縣市': '台北市',
'住址': '汀州路四段88號'}
```

輸入檔

	A	B	C
1	姓名	縣市	住址
2	陳大明	台北市	汀州路四段88號
3	張小慧	臺北市	和平南路一段162號

程式執行輸出畫面

```
原始資料: 陳大明 台北市 汀州路四段88號
更新資料: 陳大明 臺北市 汀州路四段88號
原始資料: 張小慧 臺北市 和平南路一段162號
```

- ▶ 回顧本單元的任務，我們需要對資料檔中的資料作下列的修正：
 - 縣市中的「台」改為「臺」。
 - 住址中的「F」改為「樓」。
 - 住址中臺中市的中港路改為「臺灣大道」。
- ▶ 綜合前面幾頁所學，我們知道 `e` 代表資料檔中一行的資料，而且是一個字典型態的物件。
`e['縣市']` 該行欄位對應的值，是一個字串。
- ▶ 第 07 行中，`e['縣市'][0]` 代表單筆資料的「縣市」欄位中的第 1 個字。若該字為「台」，第 08 行會將該欄位改為「臺」和原資料第二個字開始的子字串串接的結果。
 - 請注意：若想取出第 `n` 個字，在 Python 中需以 `[n-1]` 取用，而非 `[n]`。

程式碼解說

```
03 with open('地址資料.csv') as infile:
04     data = list(csv.DictReader(infile))
05     for e in data:
06         print('原始資料:', e['姓名'], e['縣市'], e['住址'])
07         if e['縣市'][0] == '台':
08             e['縣市'] = '臺'+e['縣市'][1:]
09         if 'F' in e['住址']:
10             e['住址'] = e['住址'].replace('F', '樓')
11         if e['縣市'] == '臺中市' and '中港路' in e['住址']:
12             e['住址'] = e['住址'].replace('中港路', '臺灣大道')
13
14     print('    更新資料:', e['姓名'], e['縣市'], e['住址'])
```

輸入檔

	A	B	C
1	姓名	縣市	住址
2	陳大明	台北市	汀州路四段88號
3	張小華	臺北市	和平東路一段162號
4	李國華	台北市	復興南路一段6號8樓
5	朱志清	台中市	大仁路38巷4號5F
6	朱婷婷	台中市	中港路29號2樓
7	周杰倫	台南市	中港路四段390-1號
8	劉德明	高雄市	三多路49巷2弄1號14F
9	紀小月	台東市	中台五路95號

程式執行輸出畫面

```
原始資料: 陳大明 台北市 汀州路四段88號
更新資料: 陳大明 臺北市 汀州路四段88號
原始資料: 張小華 臺北市 和平東路一段162號
更新資料: 張小華 臺北市 和平東路一段162號
原始資料: 李國華 台北市 復興南路一段6號8樓
更新資料: 李國華 臺北市 復興南路一段6號8樓
原始資料: 朱志清 台中市 大仁路38巷4號5F
更新資料: 朱志清 臺中市 大仁路38巷4號5樓
原始資料: 朱婷婷 台中市 中港路29號2樓
更新資料: 朱婷婷 臺中市 臺灣大道29號2樓
原始資料: 周杰倫 台南市 中港路四段390-1號
更新資料: 周杰倫 臺南市 中港路四段390-1號
原始資料: 劉德明 高雄市 三多路49巷2弄1號14F
更新資料: 劉德明 高雄市 三多路49巷2弄1號14樓
原始資料: 紀小月 台東市 中台五路95號
更新資料: 紀小月 臺東市 中台五路95號
>>> |
```

- ▶ 第 09 行判斷 e['住址'] 是否包含 'F'，若有，則第 10 行將 e['住址'] 更改為把 'F' 取代為 '樓' 的新字串。
- ▶ 第 11-12 行則是當 e['縣市'] 是 '臺中市' 而且 e['住址'] 中包含 '中港路' 時，將 e['住址'] 更改為把 '中港路' 取代為 '臺灣大道' 的新字串。

程式碼解說

```
16 with open('新地址資料.csv', 'w', newline='') as outfile:
17     writer = csv.DictWriter(outfile, fieldnames = data[0].keys())
18     writer.writeheader()
19     for e in data:
20         writer.writerow(e)
```

- ▶ 利用前面的程式，我們已經將輸入檔的資料讀入且依規則修正完成。最後我們要將修正過的資料存成一個新的 **CSV** 檔。
- ▶ 第 16 行我們以 `with` 敘述和 `open()` 函式開啟一個檔案，和之前開啟不同的是我們在 `open()` 函式的檔名後面加上了 'w'。這個 'w' 代表我們會開啟一個全新的檔案，而且這個檔案是用來寫入的。
- ▶ 第 17 行我們運用 `csv` 模組的 `DictWriter()` 函式，這個函式建立一個可輕易將字典內容輸出到檔案的物件，我們以 `writer` 代表之。
 - 此處我們指定輸出檔是第 16 行建立的 `outfile` 所代表的檔案。
 - 呼叫 `DictWriter()` 函式時，`fieldnames` 參數用來設定將來處理的字典的欄位名稱。`data` 列表中所有字典的鍵都是一樣的，這裡我們以 `data[0].keys()`，也就是第一個字典物件的鍵來設定。
- ▶ 在第 17 行設定 `fieldnames` 後，第 18 行才能以 `writeheader()` 在檔案中寫入表頭。
- ▶ 最後，第 19-20 行逐一取出 `data` 列表中的每個字典，令 `e` 代表之，然後以 `writerow()` 將字典輸出成一行文字。

姓名	縣市	住址
陳大明	臺北市	汀州路四段88號
張小華	臺北市	和平東路一段162號
李國華	臺北市	復興南路一段6號3樓
朱志清	臺中市	大仁路38巷4號5樓
朱婷婷	臺中市	臺灣大道29號2樓
周杰倫	臺南市	中港路四段390-1號
劉德明	高雄市	三多路49巷2弄1號14樓
紀小月	臺東市	中台五路95號

地址處理

完整程式展示

```
import csv

with open('地址資料.csv') as infile:
    data = list(csv.DictReader(infile))
    for e in data:
        print('原始資料:', e['姓名'], e['縣市'], e['住址'])
        if e['縣市'][0] == '台':
            e['縣市'] = '臺'+e['縣市'][1:]
        if 'F' in e['住址']:
            e['住址'] = e['住址'].replace('F', '樓')
        if e['縣市'] == '臺中市' and '中港路' in e['住址']:
            e['住址'] = e['住址'].replace('中港路', '臺灣大道')

        print('    更新資料:', e['姓名'], e['縣市'], e['住址'])

with open('新地址資料.csv', 'w', newline='') as outfile:
    writer = csv.DictWriter(outfile, fieldnames = data[0].keys())
    writer.writeheader()
    for e in data:
        writer.writerow(e)
```

姓名	縣市	住址
陳大明	台北市	汀州路四段88號
張小華	臺北市	和平東路一段162號
李國華	台北市	復興南路一段6號8樓
朱志清	台中市	大仁路38巷4號5F
朱婷婷	台中市	中港路29號2樓
周杰倫	台南市	中港路四段390-1號
劉德明	高雄市	三多路49巷2弄1號14F
紀小月	台東市	中台五路95號

```
-----
原始資料: 陳大明 台北市 汀州路四段88號
更新資料: 陳大明 臺北市 汀州路四段88號
原始資料: 張小華 臺北市 和平東路一段162號
更新資料: 張小華 臺北市 和平東路一段162號
原始資料: 李國華 台北市 復興南路一段6號8樓
更新資料: 李國華 臺北市 復興南路一段6號8樓
原始資料: 朱志清 台中市 大仁路38巷4號5F
更新資料: 朱志清 臺中市 大仁路38巷4號5樓
原始資料: 朱婷婷 台中市 中港路29號2樓
更新資料: 朱婷婷 臺中市 臺灣大道29號2樓
原始資料: 周杰倫 台南市 中港路四段390-1號
更新資料: 周杰倫 臺南市 中港路四段390-1號
原始資料: 劉德明 高雄市 三多路49巷2弄1號14F
更新資料: 劉德明 高雄市 三多路49巷2弄1號14樓
原始資料: 紀小月 台東市 中台五路95號
更新資料: 紀小月 臺東市 中台五路95號
>>>|
```

姓名	縣市	住址
陳大明	臺北市	汀州路四段88號
張小華	臺北市	和平東路一段162號
李國華	臺北市	復興南路一段6號8樓
朱志清	臺中市	大仁路38巷4號5樓
朱婷婷	臺中市	臺灣大道29號2樓
周杰倫	臺南市	中港路四段390-1號
劉德明	高雄市	三多路49巷2弄1號14樓
紀小月	臺東市	中台五路95號

挑戰時刻

```
01 import csv
02
03 with open('地址資料.csv') as infile:
04     data = list(csv.DictReader(infile))
05     for e in data:
06         print('猜猜看結果會顯示甚麼資料', e['縣市'][0:2])
```

- ▶ 請問上面程式會顯示那些資料？
- ▶ 請修改今天的地址程式，使得它能在最後輸出如圖的資料
 - 住址的部分，直接複製縣市名稱加入。

姓名	縣市	住址
陳大明	台北市	台北市汀州路四段88號
張小華	臺北市	臺北市和平東路一段162號
李國華	台北市	台北市復興南路一段6號8樓
朱志清	台中市	台中市大仁路38巷4號5F
朱婷婷	台中市	台中市港路29號2樓
周杰倫	台南市	台南市中港路四段390-1號
劉德明	高雄市	高雄市三多路49巷2弄1號14F
紀小月	台東市	台東市中台五路95號

挑戰時刻2

```
01 s = '1999/1/2'  
02 t = s.split('/')  
03 print(t[0], t[1], t[2])  
04 y = int(t[0])  
05 if y>110:  
06     y -= 1911  
07 print(str(y)+'.'+t[1]+'.'+t[2])
```

```
1999 1 2  
88.1.2
```

- ▶ 現在資料檔中又多了「生日」欄位，但每個人填寫的生日格式仍舊不一致，需要我們撰寫程式來修正。
- ▶ 主要修正規則有二：
 - 將 '/' 都改為 '.'。
 - 將西元年份（超過 110）的改為民國（減 1911）。
- ▶ 本頁上方的程式碼提供你一些提示：
 - 第 02 行示範如何將字串以 '/' 作為分割符號切成三段。
 - 第 04 行示範如何將字串轉換成數字，以便作大小比較（第 05 行）。
 - 第 07 行示範如何將數字轉換成字串，並和其它字串以 + 運算串接起來。

	A	B	C	D
1	姓名	生日	縣市	住址
2	陳大明	1999/1/2	臺北市	汀州路四段88號
3	張小華	100.4.30	臺北市	和平東路一段162號
4	李國華	101.5.15	臺北市	復興南路一段6號3樓
5	朱志清	102.12.1	臺中市	大仁路38巷4號5樓
6	朱婷婷	100.6.6	臺中市	臺灣大道29號2樓
7	周杰倫	1979/1/18	臺南市	中港路四段390-1號
8	劉德明	101.7.7	高雄市	三多路49巷2弄1號14樓
9	紀小月	103/7/8	臺東市	中台五路95號

	A	B	C	D
1	姓名	生日	縣市	住址
2	陳大明	88.1.2	臺北市	汀州路四段88號
3	張小華	100.4.30	臺北市	和平東路一段162號
4	李國華	101.5.15	臺北市	復興南路一段6號3樓
5	朱志清	102.12.1	臺中市	大仁路38巷4號5樓
6	朱婷婷	100.6.6	臺中市	臺灣大道29號2樓
7	周杰倫	68.1.18	臺南市	中港路四段390-1號
8	劉德明	101.7.7	高雄市	三多路49巷2弄1號14樓
9	紀小月	103.7.8	臺東市	中台五路95號