

Project#4

即時空汙品質查詢

IM1201301 – Computer Programming
Instructor: Chao-Lung Yang

即時空汙品質查詢 (project)

- ▶ 空氣污染越來越嚴重了，若能即時知道空氣污染的指數那就太好了！
- ▶ 身邊有**Android**裝置的同學，有注意到**Google Now**最近不時的跳出空汙緊告通知，但**Google**到底是怎麼做到這點的？
- ▶ 其實**Google**可能是利用政府開放的資料來達到通知的效果

空污資料下載

- ▶ 請先到下列網站下載資料，了解一些資料長得什麼樣子
- ▶ 開放資料來源：
[http://opendata.epa.gov.tw/ws/Data/REWXQA/?\\$orderby=SiteName&\\$skip=0&\\$top=1000&format=csv](http://opendata.epa.gov.tw/ws/Data/REWXQA/?$orderby=SiteName&$skip=0&$top=1000&format=csv)
- ▶ 這個資料每個小時都會更新，但如何讓python自動的讀取這個資料呢？

Practice#1:線上存取檔案

► 使用 urllib2

讀取網頁資料

```
import urllib.request
rsp = urllib.request.urlopen('http://opendata.epa.gov.tw/ws/Data/REWXQA/?$orderby=SiteName&$skip=0&$top=1000&format=csv')
data=rsp.read()
csvdata = data.decode('utf8')
print(csvdata)
```

轉成unicode編碼

```
SiteName,County,PSI,MajorPollutant,Status,SO2,CO,O3,PM10,PM2.5,NO2,WindSpeed,WindDirec,FPMI,NOx,NO,PublishTime
二林,彰化縣,61,懸浮微粒,普通,7.9,0.51,28,66,26,17,2.2,42,3,18.04,.84,2017-05-07 00:00
三重,新北市,41,,良好,1.6,0.92,,27,4,23,,2,36.09,12.94,2017-05-07 00:00
三義,苗栗縣,46,,良好,2.3,0.37,28,38,17,7.4,3.8,40,2,8.73,1.34,2017-05-07 00:00
土城,新北市,32,,良好,1,0.35,39,17,6,11,1.3,302,2,12.14,1.16,2017-05-07 00:00
士林,臺北市,34,,良好,1.1,0.24,40,16,13,6.3,2.6,112,2,6.56,.28,2017-05-07 00:00
大同,臺北市,35,,良好,1.9,0.89,,22,3,27,,1,45.17,17.75,2017-05-07 00:00
大里,臺中市,54,懸浮微粒,普通,3.5,0.58,23,64,22,21,2.1,39,3,22.09,1.34,2017-05-07 00:00
大園,桃園市,39,,良好,4.2,0.26,37,32,9,16,6.2,43,2,17.92,1.72,2017-05-07 00:00
大寮,高雄市,56,懸浮微粒,普通,2.3,0.36,24,58,24,14,0.8,63,3,14.91,1.08,2017-05-07 00:00
小港,高雄市,60,懸浮微粒,普通,4.6,0.48,31,77,31,16,1.2,59,3,17.42,1.4,2017-05-07 00:00
中山,臺北市,34,,良好,1.3,0.4,33,28,9,15,2.8,118,2,16.81,1.77,2017-05-07 00:00
中壢,桃園市,45,,良好,11,0.65,22,42,20,25,1.9,68,2,31.04,6.23,2017-05-07 00:00
仁武,高雄市,58,懸浮微粒,普通,3.4,0.48,27,69,19,19,0.7,11,3,19.74,.71,2017-05-07 00:00
斗六,雲林縣,58,懸浮微粒,普通,3.3,0.49,27,67,41,18,2.1,22,4,19.59,1.14,2017-05-07 00:00
```

Unicode? 先了解ASCII

- ▶ 每個Byte能儲存的範圍是0~255
- ▶ ASCII中要表達A其實是65這個數字，B則是66
- ▶ 但中文字這麼多,怎麼辦?

Ctrl	Dec	Hex	Char	Code	Dec	Hex	Char	Dec	Hex	Char	Dec	Hex	Char
^@	0	00		NUL	32	20	!	64	40	@	96	60	'
^A	1	01		SOH	33	21	"	65	41	A	97	61	a
^B	2	02		STX	34	22	#	66	42	B	98	62	b
^C	3	03		ETX	35	23	\$	67	43	C	99	63	c
^D	4	04		EOT	36	24	%	68	44	D	100	64	d
^E	5	05		ENQ	37	25	&	69	45	E	101	65	e
^F	6	06		ACK	38	26	'	70	46	F	102	66	f
^G	7	07		BEL	39	27	(71	47	G	103	67	g
^H	8	08		BS	40	28)	72	48	H	104	68	h
^I	9	09		HT	41	29	*	73	49	I	105	69	i
^J	10	0A		LF	42	2A	+	74	4A	J	106	6A	j
^K	11	0B		VT	43	2B	,	75	4B	K	107	6B	k
^L	12	0C		FF	44	2C	-	76	4C	L	108	6C	l
^M	13	0D		CR	45	2D	.	77	4D	M	109	6D	m
^N	14	0E		SO	46	2E	/	78	4E	N	110	6E	n
^O	15	0F		SI	47	2F	0	79	4F	O	111	6F	o
^P	16	10		DLE	48	30	1	80	50	P	112	70	p
^Q	17	11		DC1	49	31	2	81	51	Q	113	71	q
^R	18	12		DC2	50	32	3	82	52	R	114	72	r
^S	19	13		DC3	51	33	4	83	53	S	115	73	s
^T	20	14		DC4	52	34	5	84	54	T	116	74	t
^U	21	15		NAK	53	35	6	85	55	U	117	75	u
^V	22	16		SYN	54	36	7	86	56	V	118	76	v
^W	23	17		ETB	55	37	8	87	57	W	119	77	w
^X	24	18		CAN	56	38	9	88	58	X	120	78	x
^Y	25	19		EM	57	39	:	89	59	Y	121	79	y
^Z	26	1A		SUB	58	3A	;	90	5A	Z	122	7A	z
^[27	1B		ESC	59	3B	<	91	5B	[123	7B	{
^\	28	1C		FS	60	3C	=	92	5C	\	124	7C	
^]	29	1D		GS	61	3D	>	93	5D]	125	7D	}
^^	30	1E	▲	RS	62	3E	?	94	5E	~	126	7E	~
^~	31	1F	▼	US	63	3F		95	5F	_	127	7F	~*

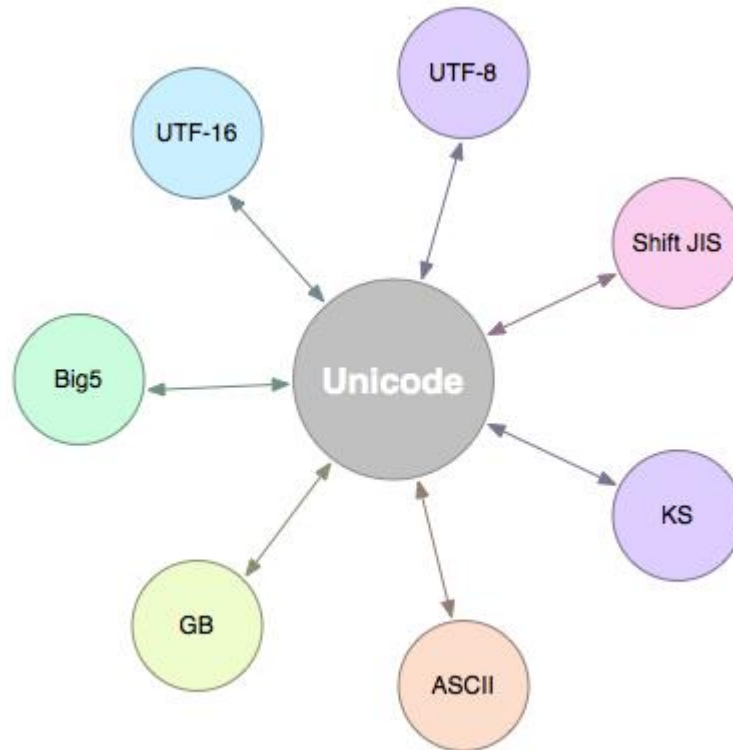
* ASCII 碼 127 具有代碼 DEL。在 MS-DOS 下，這個代碼與 ASCII 8 (BS) 的效果相同。DEL 代碼可以由 CTRL + BKSP 鍵產生。

Big5 碼

- ▶ 幾家台灣的電腦廠商一起討論決定出一套編碼系統就叫Big5，我們知道一個byte可以表示0~255，而中文字不像英文字母只有26個字，而是上萬個字，因此一個byte肯定不夠表示，所以 Big5 是用兩個數字來表示一個中文
 - (164, 164) -> 中
 - (164, 229) -> 文
 - (166, 72) -> 字
- ▶ 除了台灣以外，不同的地區都紛紛自己定出適合當地語系的文字編碼系統，例如簡體中文的 GB，日文的 Shift-JS 等等

不同編碼的困擾

- ▶ 軟體要為了不同的編碼,重新開發一次
- ▶ 如果能有一個統一的方法就好了!



Python的Unicode支援

- ▶ Python支援Unicode
- ▶ 在Python裡的字串有兩種，一種是 `str`，另一種則是 `unicode`
- ▶ 在Python表示`unicode`字串的方式，是在字串前面加上一個`u`，而且很重要的一點是，為了讓Python直譯器能正確地解析我們在程式碼裡打的中文字，我們得告訴它我們用的是什麼編碼，因此得加上
 - `# -*- coding: utf8 -*-`

Practice#2: 中文 unicode

- ▶ 試著編寫下列程式(Python3)

```
msg='今天天氣真好'  
print(msg)  
print('字數',len(msg))
```

```
今天天氣真好  
字數 6
```

Unicode vs. String (1)

- ▶ 對於 **unicode** 來說，基本單位是 “字”，而且每個字的儲存空間較大，它可以容納任何的字，在 **unicode** 裡一個字母就只佔一個位置
- ▶ 但是 **str** 就不一樣了，**str** 的基本單位是 **byte**，而因為文字編碼是以多個 **byte** 來表示一個字，例如 **Big5** 即是以2個 **byte** 來表示一個中文字，因此如果 **Unicode** 和 **Big5** 編碼的中文字相比就會像這個樣子

Unicode	今	天	天	氣	真	好						
	0	1	2	3	4	5						
String	今	天	天	氣	真	好						
	0	1	2	3	4	5	6	7	8	9	10	11

Unicode vs. String (2)

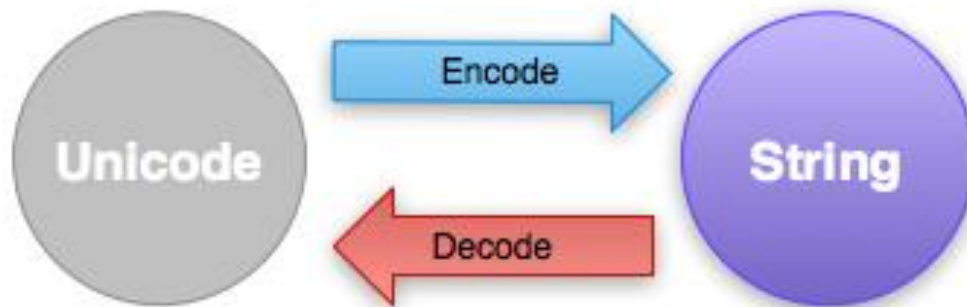
- ▶ 在上面我們學到了如何表示 `unicode` 字串，但是事實上是，`unicode` 字串只能存在程式的內部，並沒有一個統一的表達方式，並沒有辦法和外界溝通，因此當我們想把字串存到檔案裡，或著透過網路傳給別人，得先將 `unicode` 字串編碼成 `str` 字串，相對地，當我們想開啟某種編碼的檔案時，我們得進行解碼

Python 2 vs Python 3

- ▶ 在Python3因為字串已經全部統一成 unicode ，所以不必加上 u ，這是Python2和Python3的重要差別之一，需要特別注意

Encode and Decode

- ▶ Python編碼或解碼的方式很簡單，透過encode與decode的函數呼叫，我們可以在 unicode 和 str 兩種之間進行轉換



.encode

- 由於 UTF-8 可以編碼任何字集，同時還有兼容 ASCII 的優點，因此通常我們使用的編碼都是 UTF-8，編碼只要呼叫unicode的.encode函數即可，以下是編碼的簡單的範例

```
msg='今天天氣真好'  
encoded = msg.encode('utf8')  
print(encoded)  
  
b'\xe4\xbb\x8a\xe5\xa4\xa9\xe5\xa4\xa9\xe6\xb0\xa3\xe7\x9c\x9f\xe5\xa5\xbd'
```

.decode

- ▶ 相反的，解碼一樣簡單

```
encoded = b'\xe4\xbb\x8a\xe5\xa4\xa9\xe5\xa4\xa9\xe6\xb0\xa3\xe7\x9c\x9f\xe5\xa5\xbd'  
decoded = encoded.decode('utf8')  
print(decoded)
```

今天天氣真好

Practice#3: Sorted Dictionary

- ▶ Create a dict structure to store students shown in Lab #6, and show the information

```
file=open('practice_data.csv')

student_dicts={}
for line in file:
    name,grade,age = line[:-1].split(",")
    student_dicts[name] = [grade,int(age)]

file.close()

print(student_dicts)

{'john': ['A', 15], 'april': ['C', 12], 'dave': ['B', 10]}
```

Practice#3: Sorted Dictionary (cont')

- ▶ Sort the dictionary by key, and items

```
print('-----Sort by students\'s name (key)-----')
for key, value in sorted(student_dicts.items(), key = lambda item: (item[0]]):
    print(key,':',value)

print('-----Sort by students\'s grade-----')
for key, value in sorted(student_dicts.items(), key = lambda item: (item[1][0], item[0]]):
    print(key,':',value)

print('-----Sort by students\'s age-----')
for key, value in sorted(student_dicts.items(), key = lambda item: (item[1][1], item[0]]):
    print(key,':',value)
```

```
-----Sort by students's name (key)-----
april : ['C', 12]
dave  : ['B', 10]
john  : ['A', 15]
-----Sort by students's grade-----
john  : ['A', 15]
dave  : ['B', 10]
april  : ['C', 12]
-----Sort by students's age-----
dave  : ['B', 10]
april  : ['C', 12]
john  : ['A', 15]
```

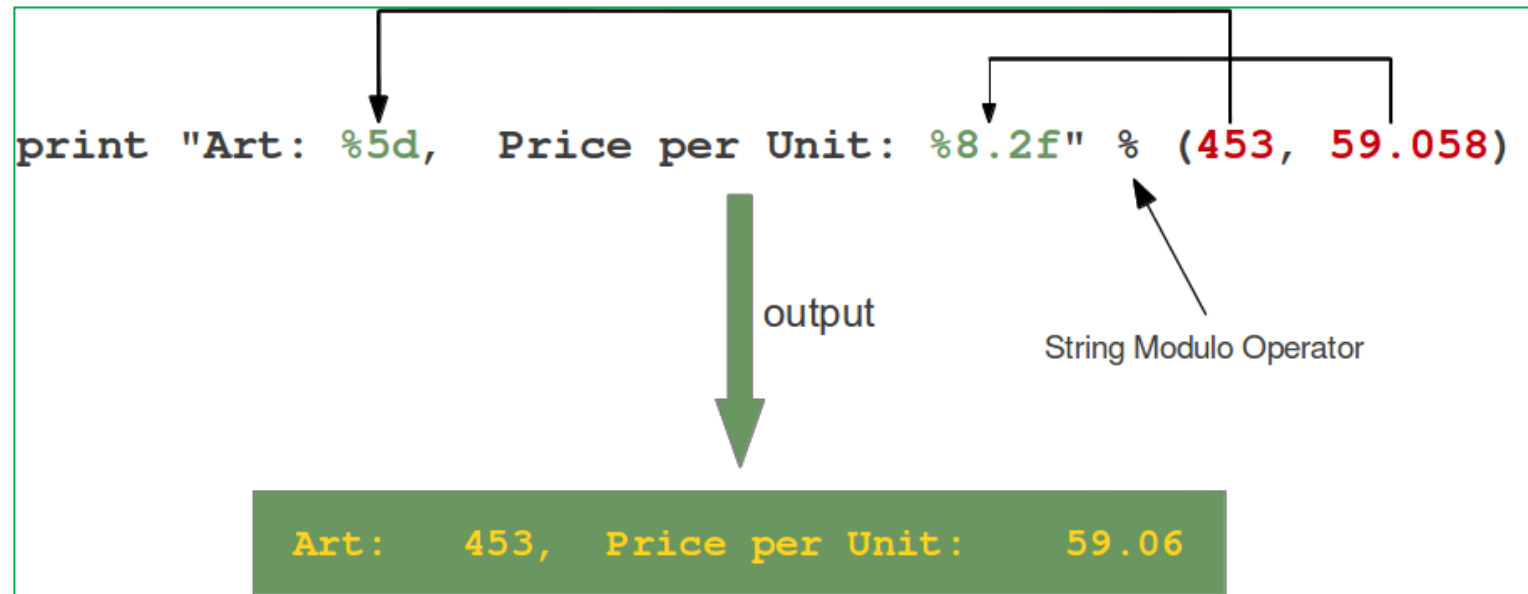
PSI 說明

► <http://taqm.epa.gov.tw/taqm/tw/b0201.aspx>

空氣品質指標 (AQI)							
AQI指標	O ₃ (ppm) 8小時平均值	O ₃ (ppm) 小時平均值 ⁽¹⁾	PM _{2.5} ($\mu\text{g}/\text{m}^3$) 24小時平均值	PM ₁₀ ($\mu\text{g}/\text{m}^3$) 24小時平均值	CO (ppm) 8小時平均值	SO ₂ (ppb) 小時平均值	NO ₂ (ppb) 小時平均值
良好 0~50	0.000 - 0.054	-	0.0 - 15.4	0 - 54	0 - 4.4	0 - 35	0 - 53
普通 51~100	0.055 - 0.070	-	15.5 - 35.4	55 - 125	4.5 - 9.4	36 - 75	54 - 100
對敏感族群 不健康 101~150	0.071 - 0.085	0.125 - 0.164	35.5 - 54.4	126 - 254	9.5 - 12.4	76 - 185	101 - 360
對所有族群 不健康 151~200	0.086 - 0.105	0.165 - 0.204	54.5 - 150.4	255 - 354	12.5 - 15.4	186 - 304 ⁽³⁾	361 - 649
非常不健康 201~300	0.106 - 0.200	0.205 - 0.404	150.5 - 250.4	355 - 424	15.5 - 30.4	305 - 604 ⁽³⁾	650 - 1249
危害 301~400	⁽²⁾	0.405 - 0.504	250.5 - 350.4	425 - 504	30.5 - 40.4	605 - 804 ⁽³⁾	1250 - 1649
危害 401~500	⁽²⁾	0.505 - 0.604	350.5 - 500.4	505 - 604	40.5 - 50.4	805 - 1004 ⁽³⁾	1650 - 2049

Print 格式化輸出

► Print output format



```
sp = ' '*5
print "[%s|%s|%s|%s]" % (sp,sp,sp,sp)
print "0    25    50    75    100"
```

```
>>>
[    |    |    |    ]
0    25    50    75    100
```

Format

Conversion	Meaning
d	Signed integer decimal.
i	Signed integer decimal.
o	Unsigned octal.
u	Unsigned decimal.
x	Unsigned hexadecimal (lowercase).
X	Unsigned hexadecimal (uppercase).
e	Floating point exponential format (lowercase).
E	Floating point exponential format (uppercase).
f	Floating point decimal format.
F	Floating point decimal format.
g	Same as "e" if exponent is greater than -4 or less than precision, "f" otherwise.
G	Same as "E" if exponent is greater than -4 or less than precision, "F" otherwise.
c	Single character (accepts integer or single character string).
r	String (converts any python object using repr()).
s	String (converts any python object using str()).
%	No argument is converted, results in a "%" character in the result.

Practice#4: Format Print Output

- ▶ Print a grade from 0 to 100 by visual bar

```
r1, r2, r3, r4 = ' '*5, ' '*5, ' '*5, ' '*5

grade = 72

if grade <= 25:
    num = int(round(grade/5,0))
    r1 = '='*num + ' '*(5-num)
elif grade <= 50:
    num = int(round((grade-25)/5,0))
    r1 = '='*5
    r2 = '='*num + ' '*(5-num)
elif grade <= 75:
    num = int(round((grade-50)/5,0))
    r1, r2 = '='*5, '='*5
    r3 = '='*num + ' '*(5-num)
elif grade <= 100:
    num = int(round((grade-75)/5,0))
    r1, r2, r3 = '='*5, '='*5, '='*5
    r4 = '='*num + ' '*(5-num)

print("[%s|%s|%s|%s]" % (r1,r2,r3,r4))
print("0      25      50      75      100")
```

```
[=====|=====|=====|      ]
0          25          50          75          100
```

Project #4

- ▶ 利用urllib取得即時空氣污染指數
- ▶ 設計一個dict來存取空污指數
- ▶ 我們需要的欄位：
 - 測站名稱(SiteName)
 - 縣市(County)
 - 空氣污染指標(PSI)
 - 指標污染物(MajorPollutant)
 - 狀態(Status)
 - 發布時間(PublishTime)

程式要求

- ▶ 你們的程式要呈現下列功能
 - PSI 排序(由高至低, 顯示**PSI 數字**, SiteName, County, Status)
 - 地區查詢功能
 - 輸入 SiteName 去查詢,顯示**PSI 數字**, SiteName, County, Status, MajorPollutant
 - PSI 以圖形化比例尺方式取代數字顯示
 - 請依下列比例計算

```
>>>
58: 古亭 臺北市 懸浮微粒
[====|      |      |      |      |      ]
0      50     100     150     200     300     500
```

Project #4

- ▶ Deadline: 5/19
- ▶ Teamwork project, every group has to submit “**ipynb**” file
- ▶ Hint
 - 你需要用到 `csvdata.splitlines()`, 及 `split(',')` 來將空污資料切開並放入 `dict` 結構

References

- ▶ Python Unicode

- <http://python.ez2learn.com/basic/unicode.html>
- http://www.python-course.eu/python3_formatted_output.php

- ▶ 行政院環保署空氣品質監測網

- <http://taqm.epa.gov.tw/taqm/tw/b0201.aspx>

- ▶ 行政院環保署環境資源資料開放平台(測試版)

- <http://opendata.epa.gov.tw/>