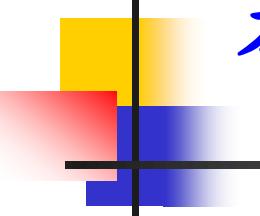


資料科學 檔案處理



國立宜蘭大學資訊工程系
吳政瑋 助理教授
蔡佳韋 課程助教

wucw@niu.edu.tw



本堂教學重點函式

- open()
- read()
- readlines()

open 函式

open(filename[,mode][,encode])

open 函式只有 filename 參數 必填

- filename(字串) 為讀寫的檔案名稱，包含檔案路徑(絕對或相對)
- mode(字串) 為檔案開啟模式(預設讀取)

模式	說明
r	讀取模式(預設)
w	寫入模式，若檔案已存在，內容將會被覆蓋。
a	附加模式，若檔案已存在，內容將會被附加至尾端。

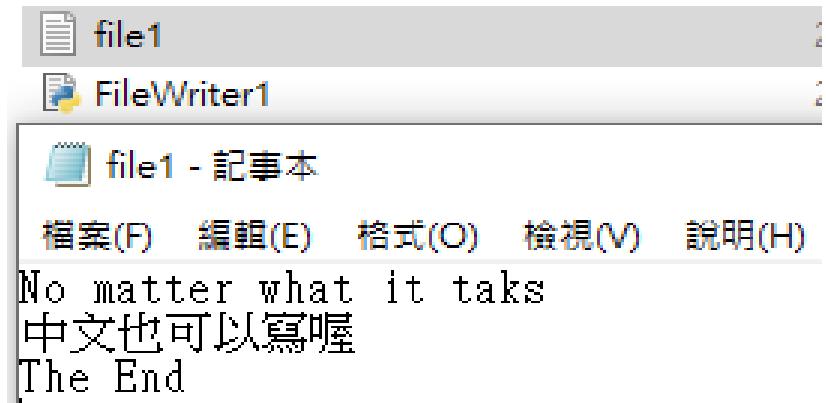
- encode 指定檔案編碼模式，cp950(Big-5)或 UTF-8

open 函式範例

```
FileWriter1.py

content = '''No matter what it takes
中文也可以寫喔
The End
'''

f = open('file1.txt', 'w')
f.write(content)
f.close()
```



open 函式範例-2

```
FileReader1.py

1 f = open('file1.txt', 'r')
2 for line in f:
3     print(line, end="")
4 f.close()

No matter what it takes
中文也可以寫喔
The End
[Finished in 0.1s]
```

使用 with...as 語法

- 可 with...as 語法 將 open() 開啟的檔案簡化 close()

```
FileReader1.py
1 f = open('file1.txt', 'r')
2 for line in f:
3     print(line, end="")
4 f.close()
```

No matter what it takes
中文也可以寫喔
The End
[Finished in 0.1s]

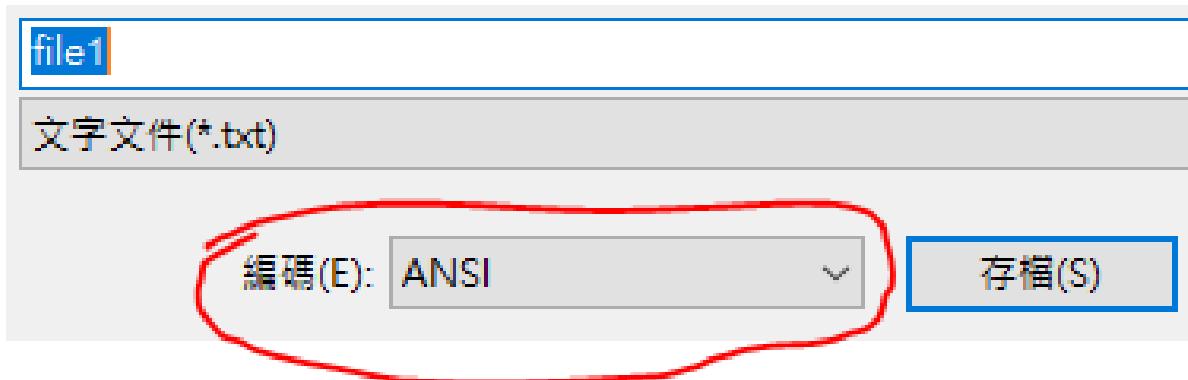
```
FileReader1.py
1 with open('file1.txt', 'r') as f:
2     for line in f:
3         print(line, end="")
```

No matter what it takes
中文也可以寫喔
The End
[Finished in 0.1s]

with 敘述內需縮排

設定 檔案編碼

- open() 預設的文件編碼會依作業系統而定，
 - Windows 系統預設編碼 cp950 (Big-5)



記事本編輯文件後儲存檔案預設編碼為ANSI，即是cp950(Big-5)

指定讀取檔案的編碼

指定編碼 cp950:

```
f = open('file1.txt', 'r', encoding = 'cp950')
```



預設 Big-5(cp950)
與編碼不同

```
FileReader1.py
```

```
1 f = open('file1.txt', 'r', encoding = 'cp950')
2 for line in f:
```

File "C:\PythonPracitce\FileReader1.py", line 2, in <module>

Traceback (most recent call last):

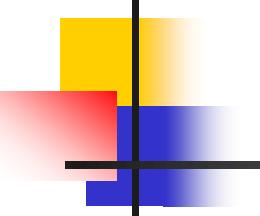
```
  File "C:\PythonPracitce\FileReader1.py", line 2, in <module>
    for line in f:
UnicodeDecodeError: 'cp950' codec can't decode byte 0x96 in position 22: multibyte sequence
```

[Finished in 0.1s with exit code 1]

```
[shell_cmd: python -u "C:\PythonPracitce\FileReader1.py"]
```

```
[dir: C:\PythonPracitce]
```

```
[path: C:\Windows\system32;C:\Windows;C:\Windows\System32\Wbem;C:\WindowsPowerShell\v1.0\;C:\Windows\System32\OpenSSH\;C:\Program Fi
```



檔案處理

方法	說明
close()	關閉檔案，檔案關閉後就不再進行讀寫的動作
flush()	一般情況，在檔案關閉時會將資料寫入檔案中，也可以使用 flush() 強迫將緩衝區資料 立即 寫入檔案中，並清除緩衝區
read([size])	讀取指定長度的字元(預設所有字元)
readable()	測試是否可讀取
readline([size])	讀取目前文字指標所在列中 size 長度的文字內容，若省略參數，則會讀取一整列，包含"\n"字符
readlines()	讀取所有列，它會回傳一個串列
next()	移動到下一列
seek(0)	將指標移到文件最前端
tell()	傳回文件目前位置
write(str)	將指定字串寫入文件中，它沒有返回值
writable()	測試是否可寫入

read()

- 讀取指定長度的字元(預設所有字元):
- 例如讀取 file1.txt 前 5 字元:

```
◀ ▶ fileread().py ×  
1 f = open('file1.txt', 'r')  
2 str1 = f.read(5)  
3 print(str1) # No ma  
4 f.close()  
  
No ma  
[Finished in 0.1s]
```

readlines()

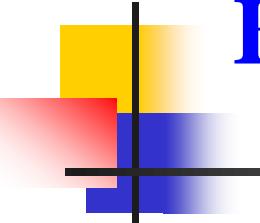
讀取所有列，它會回傳一個串列

每一列會成為串列中的一個元素

```
fileread0.py
```

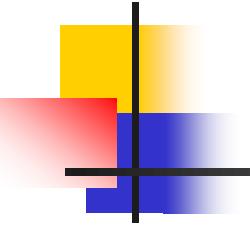
```
1 with open('file1.txt', 'r') as f:
2     rows = f.readlines() # 第一列為 rows[0] 第二列為 rows[1] ...
3     print(type(rows))   # <class 'list'>
4     print(rows)
```

```
<class 'list'>
['No matter what it takes\n', '中文也可以寫喔\n', 'The End\n']
[Finished in 0.1s]
```



BOM 的處理

- Windows 中的 UTF-8 文件，會在檔案最前方自動產生占了一個字元的 **BOM (Unicode Byte Mark)**，功能是讓軟體識別是否為 Unicode 的檔案
- 可在 open() 中加上 encoding = ‘UTF-8-sig’ 將 BOM 去除

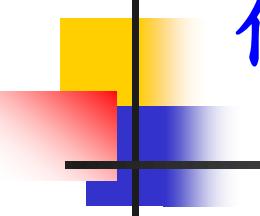


readline([size])

- 讀取目前文字指標所在列中 size 長度的文字內容，若省略參數，則會讀取一整列，包含”\n”字符

```
file = open('file2.txt', 'r', encoding = 'UTF-8-sig')
```

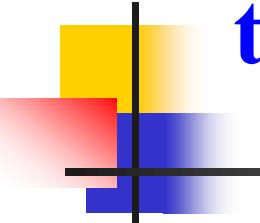
- 加上 `encoding = 'UTF-8-sig'` 將 **BOM** (`\nfeff`) 去除



例外處理

- Python 直譯器 (interpreter) 逐行執行發生錯誤時，會引發例外 (exception)
 - ex: 變數不存在，資料型別不符
- 有些錯誤，並非程式邏輯有錯
 - ex: 程式中開啟的檔案不存在

這時需引發 例外處理 `try...except...finally`



try...except...finally 語法結構

- try:
 - 可能引發錯誤的程式區塊
- except 例外情形一 [as 參數]:
 - 處理錯誤的程式區塊
- except 例外情形二 [as 參數]:
 - 處理錯誤的程式區塊

except Exception [as 參數]:

處理所有其他可能發生的錯誤

finally:

一定會執行的程式區塊

try...except...finally 使用方式

```
try:  
    print(n)  
except:  
    print("變數 n 不存在")
```

```
1  try:  
2      print(n)  
3  except Exception as e:  
4      print(e) # name 'n' is not defined  
5  
name 'n' is not defined
```

```
1  try:  
2      print(n)  
3  except:  
4      print('變數 n  is not defined')  
5  finally:  
6      print("一定會執行的程式區塊")  
  
變數 n  is not defined  
一定會執行的程式區塊
```

try...except 範例

- 輸入 a 、 b 後求和並顯示，若輸入非數值的字元將引起錯誤
- 發生錯誤的處理，顯示「發生輸入非數值的錯誤」

```
try:  
    a=int(input("請輸入第一個整數"))  
    b=int(input("請輸入第二個整數"))  
    r = a + b  
    print("r=", r)  
except:  
    print("發生輸入非數值的錯誤")
```

```
C:\PythonPracitce>python TryE  
請輸入第一個整數 3  
請輸入第二個整數 5  
r= 8  
  
C:\PythonPracitce>python TryE  
請輸入第一個整數 3  
請輸入第二個整數 a  
發生輸入非數值的錯誤
```

try...except 常用錯誤表

錯誤名稱	說明
IOError	輸入/輸出錯誤
NameError	變數名稱未宣告的錯誤
ValueError	數值錯誤
ZeroDivisionError	除數為 0 的錯誤

```
try:  
    a=int(input("請輸入第一個整數"))  
    b=int(input("請輸入第二個整數"))  
    r = a % b  
    print("r=", r)  
except ValueError:  
    print("發生輸入非數值的錯誤")  
except Exception as e:  
    print("發生 ",e," 的錯誤 包括分母為 0 的錯誤")  
finally:  
    print("一定會執行的程式區塊")
```

請輸入第一個整數 3
請輸入第二個整數 x
發生輸入非數值的錯誤
一定會執行的程式區塊

C:\PythonPracitce>python TryException03.py
請輸入第一個整數 3
請輸入第二個整數 0
發生 integer division or modulo by zero 的錯誤 包括分母為 0 的錯誤
一定會執行的程式區塊

教師資訊



- 姓名:吳政瑋 (小吳老師)
- 現職:宜大資工專案助理教授
- 學歷:成功大學資工博士
- 研究興趣:資料探勘、人工智慧、AIoT應用
- 通訊方式
 - 電子信箱:wucw@niu.edu.tw
 - 校內電話: (03)9317331
 - Line: silvemoonfox
 - Office: 格致大樓E405室
 - 數位學習園區
- 實驗室:AI與資料科學實驗室
 - <https://sites.google.com/view/cwwuadslab/>

意見交流

歡迎提供意見與指導!!

您的寶貴意見將使本系更進步!!