



# 字串處理

朱克剛

# 字串長度 – len()

```
str = "abcdefg"  
print (len(str)) # 7
```

# 子字符串

a	b	c	d	e	f	g
0	1	2	3	4	5	6

```
str = "abcdefg"
```

```
print (str[1:3])    #bc  
print (str[2:])     #cdefg  
print (str[:1])     #a
```

# 字串搜尋 – index()

## ■ 語法

- `idx = str.index(str, beg = 0, end = len(string))`
- 找不到時產生exception

```
str = "abcdefg"

try:
    idx = str.index("hi")
    print(idx)
except ValueError:
    print("Not found")
```

```
str = "abcdefg"

try:
    idx = str.index("de")
    print(idx) # 3
except ValueError:
    print("Not found")
```

# 字串搜尋 – find()

## ■ 語法

- `idx = str.find(str, beg = 0, end = len(string))`
- 找不到時傳回 -1

```
str = "abcdefg"
```

```
idx = str.find("hi")  
print(idx) # -1
```

```
idx = str.find("de")  
print(idx) # 3
```

# 頭尾去空白 – trim

```
str = "    abc \r\n"

print(str.strip()) # "abc"
print(str.lstrip()) # "abc \r\n"
print(str.rstrip()) # "    abc"
```

# 字符串取代 – replace()

```
str = "abcdefg"
```

```
substr = str.replace("fg", "hhh")  
print(substr)    # abcdehhh
```

# 分割字符串 – split()

```
str = "ab,cd,ef"
```

```
print(str.split(","))      # ['ab', 'cd', 'ef']  
print(str.split(",", 1))   # ['ab', 'cd,ef']
```



# 大小寫

- `"abcd".upper()`：大寫
- `"ABCD".lower()`：小寫

# 排序

- 動到陣列本身

```
a = [1, 4, 5, 3, 2]  
a.sort()  
print(a)
```

- 不動到陣列本身

```
a = [1, 4, 5, 3, 2]  
b = sorted(a)  
print(b)
```

反向排序：  
加上 `reverse=True` 參數

# Regular Expression

- 正規表示法
- 用多到「嚇死人」「看了就討厭」「厲害」的符號來描述某特定字串
- 參考資料
  - 官方：<https://docs.python.org/3.7/howto/regex.html>
  - 值得看：[https://www.tutorialspoint.com/python/python\\_reg\\_expressions.htm](https://www.tutorialspoint.com/python/python_reg_expressions.htm)

# 範例1 - 貪心符合

```
import re

str = '<div id="a01" style="color: red">hello</div>'

match = re.search('<.*>', str)
print (match.group())
```

-----

執行結果

<div id="a01" style="color: red">hello</div>

## 範例2 – 知足符合

```
import re

str = '<div id="a01" style="color: red">hello</div>'

match = re.search('<.*?>', str)
print (match.group())
```

---

執行結果

```
<div id="a01" style="color: red">
```

# 範例3

```
import re

str = '<div id="a01" style="color: red">hello</div>'

match = re.search('(co.*)"', str)
print (match.group())
print (match.group(1))
```

---

執行結果

```
color: red"
color: red
```

# 範例4

```
import re

str = '<div id="a01" style="color: red">hello</div>'

match = re.search('>(.*?)<', str)
print (match.group())
print (match.group(1))
```

---

執行結果  
>hello<  
hello

# 範例5

```
import re

str = '<div id="a01" style="color: red">hello</div>'

match = re.search('id="(.*?)"', str)
print (match.group())
print (match.group(1))
```

---

執行結果

```
id="a01"
a01
```