

07 常用資料處理函式

黃彬華編撰

- ❖ 常用數學函式
- ❖ 常用亂數函式
- ❖ 常用文字函式
- ❖ 日期時間相關名詞解釋
- ❖ 日期時間功能
- ❖ 正規表示式

- ❖ Python提供許多實用的數學函式
 - 有些屬於 builtins 模組 (不用 import)
 - 有些屬於 math 模組 (需要 import)

範例

黃彬華編撰

❖ MathDemo

常用亂數函式

黃彬華編撰

- ❖ 產生整數亂數，包含 start 與 stop 的值
 - `random.randint(start, stop)`
- ❖ 產生間距亂數，包含 start，但不包含 stop 的值
 - `random.randrange(start, stop, step)`
- ❖ 產生浮點亂數，包含 start 與 stop 的值
 - `random.uniform(start, stop)`
- ❖ 集合類型亂數
 - 隨意挑出一個元素
 - ✦ `random.choice(list)`
 - 洗牌 (隨機重排)
 - ✦ `random.shuffle(list)`

範例

黃彬華編撰

❖ RandomDemo

練習 7-1

黃彬華編撰

使用亂數產生幾個 1~100 (不含) 整數？ 5

18 不是質數

60 不是質數

21 不是質數

83 是質數

14 不是質數

- ❖ 輸入要產生幾個 1~100 (不含) 整數後按確定，會使用亂數產生對應個數的整數，並檢查是否為質數

常用文字函式

黃彬華編撰

- ❖ String 功能可以分成 3 類：取值、比值、改值
- ❖ String 物件呼叫方法改值不會改變自己的實例內容，而是將改變結果存入新的物件實例，稱為 Immutable

範例

黃彬華編撰

❖ StringDemo

日期時間相關名詞解釋

黃彬華編撰

❖ 本初子午線

- 即 0 度經線，亦稱格林威治子午線或本初經線，是經過英國格林威治天文台的一條經線。本初子午線的東西兩邊分別定為東經和西經，於 180 度相遇

❖ UTC (Coordinated Universal Time，世界標準時)

- 0 度經線的標準時

❖ 時區 (Time Zone)

- 0 度經線為起點，每 15 度為一個時區 (加或減一小時)，全球分成 24 個時區，台灣為 "+8" 時區

日期時間功能 - 1

黃彬華編撰

❖ datetime 可處理日期時間

- datetime 物件內儲存著從參考時間點起至指定時間點所經過的微秒 (microsecond, 10^{-6} 秒) 數，所以 Python 處理日期時間，可以精準到微秒
 - ✦ 上述參考時間點稱為 epoch (紀元)，是 UTC 時間 1970-1-1 00:00:00。在此時間點以後為正值，之前為負值
- datetime.timestamp(dt) 可以將指定日期時間轉成時間戳記 ("秒.微秒" 型式)
- datetime 可以取得現在或指定日期時間
- 也可以比較 2 個 datetime 大小

❖ date 僅處理日期，其餘與 datetime 功能相同

日期時間功能 - 2

黃彬華編撰

❖ 日期時間的加減

- timedelta 可以加減指定單位時間，但不包含年、月
- relativedelta 可以加減年、月，但必須額外安裝 python-dateutil 套件
 - ✦ `conda install python-dateutil`

日期時間功能 - 3

黃彬華編撰

❖ 格式化與解析

- 格式化是將原始資料轉成特定樣式的資料，且格式化完畢的結果為文字
- 格式化目的
 - ✦ 適合使用者閱讀
 - ✦ 資料交換
- 資料交換時會約定好要交換的樣式
 - ✦ 傳送資料前要將資料格式化成約定好的樣式
 - ✦ 收到資料時還要解析回原始資料型式

❖ 格式化日期

- `datetime.strftime(fmt)`
- `str.format()`
- 日期時間格式符號

❖ 解析日期

- `datetime.strptime(datetimeStr, fmt)`

❖ 時區

- 建立 datetime 物件沒有指定時區，就是建立當地日期時間
 - ✦ `datetime.astimezone()` 可以取得帶有時區的日期時間
- 建立 UTC 時區的 datetime 物件
 - ✦ 時區參數套用 `timezone.utc`
- 建立指定時區的 datetime 物件
 - ✦ 時區參數套用 `timezone(offset=timedelta(hours=時差))`
 - * 時差可以為正或負整數

範例

黃彬華編撰

❖ DateTimeDemo

練習 7-2

黃彬華編撰

借書日期 (年-月-日): 2023-4-28

可借幾日: 7

2023-05-05 前歸還

今天是 2023-04-29

您必須在 6 天內歸還

借書日期 (年-月-日): 2023-4-20

可借幾日: 7

2023-04-27 前歸還

今天是 2023-04-29

逾期天數: 2

❖ 使用者輸入

- 借書日期：格式為「年-月-日」
- 可借幾日：例如 7 日

❖ 顯示

- 還書日期
- 顯示今天是否逾期
 - ✦ 未逾期，顯示還有幾天才需還書
 - ✦ 若逾期，顯示逾期天數

正規表示式 - 1

黃彬華編撰

- ❖ 正規表示式 (regular expression) 可用來檢查格式是否符合規定，例如要求使用者輸入正確的手機號碼格式
 - 需要「import re」
 - 編譯正規表示式並回傳 Pattern 物件
 - ✦ `pattern = re.compile(r"09[0-9]{8}")`
 - ✦ 使用「r'''」raw string，可免除使用跳脫符號麻煩
 - 呼叫 `fullmatch()` 比對，而非 `match()`
 - ✦ `fullmatch()` 會比對整個字串；`match()` 只會比對開始位置
- ❖ 正規表示式常用的格式符號如下頁所示
- ❖ 正規表示式完整的格式符號可參看 [Regular Expression Metacharacters](#)

❖ 正規表示式格式符號列表

正規表示式的格式符號	
符號	說明
[ABC]	A、B、C任一個字元都符合要求 範例：[ABC]ook代表可以是Aook、Book、Cook
[^ABC]	不可以含有A、B、C任一個字元 範例：[^ABC]ook代表不可以是Aook、Book、Cook
[A-C]	可以是A到C連續字元中的任何一個 範例：[A-C]ook代表可以是Aook、Book、Cook
[^A-C]	不可以含有A到C連續字元中的任何一個 範例：[^A-C]ook代表不可以是Aook、Book、Cook
{n,m}	代表指定字元出現次數最少n次，最多m次（注意：逗號與m之間不可以有任何空白），n與m都是大於等於0的整數，但是n <= m 範例：Book{1,2}代表k最少要出現1次，最多只能出現2次，所以可以是Book、Bookk
{n}	代表指定字元正好出現n次 範例：Book{1}代表K要出現正好1次，所以只能為Book
{n,}	代表指定字元至少出現n次 範例：Book{1,}代表k要出現1次以上，所以可以是Book、Bookk...

❖ 正規表示式格式符號列表

\d	可以是0到9任何一個數字；相當於 [0-9]
\D	不可以是0到9任何一個數字；相當於 [^0-9]
\s	可以是任何空白的字元，相當於 [\t\n\r\f\v]
\S	不可以是空白的字元，相當於 [^ \t\n\r\f\v]
\w	可以是一個英文字母或數字，相當於 [a-zA-Z0-9_]
\W	不可以是任何英文字母或數字，相當於 [^a-zA-Z0-9_]
?	指定字元最多出現1次，也可以不出現。相當於{0,1} 範例：S?PP代表可以是PP或SPP。「?»在S後面代表S出現次數受到「?»所限制
+	指定字元至少出現1次以上。相當於{1,}
*	指定字元出現0次以上。相當於{0,}
	OR功能，A B代表A或B皆可
.	任一字元
()	括號內，代表同一個群組 範例：(SPP){2}代表SPP是一體，必須同時出現2次：SPPSPP
\	跳脫符號：取消原運算符號的功能，使其變成單純文字 需要跳脫方能成為單純文字的符號：* ? + [() { } ^ \$ \.

範例

黃彬華編撰

❖ RegexDemo

練習 7-3

黃彬華編撰

生日 (例如1981-1-1): 2000-2-30

生日格式錯誤，請再輸入一次！

生日 (例如1981-1-1): 2000-2-28

生日格式正確: 2000-2-28

年齡 (可以為小數): 23.

年齡格式錯誤，請再輸入一次！

年齡 (可以為小數): 23.12

年齡格式正確: 23.12

- ❖ 檢查使用者輸入的生日與年齡是否合於要求格式
 - 生日格式必須為yyyy-mm-dd，其中yyyy只可以為 1900 到 2020，例如: 1981 或 2001。mm 或 dd 都可以為 1 到 2 位數字。如果格式正確，顯示「格式正確」，否則顯示「格式錯誤」
 - 年齡格式可以為整數或帶有小數的數字，介於 0（含）~ 120（不含），且不可為負數