



東吳大學 X 理律學堂

Python_資料正則化_Regex

陳靖翔助教



使用時機

文本分析

資料探勘

=>簡單的說就是從一篇很長的文章中找出你要的東西



對象

有一定規律

你講得出他是什麼跟什麼組成的

身分證字號: 英文字母X1 + 數字X9

生日: 數字X2 + “月” + 數字X2 + ”日”

方法---普通字元

通用式	說明	比對成立之字串
/a/	含字母“a”的字串。	"ab", "bac", "cba"
/a./	含字母“a” 以及其後任一個字元的字串。	"ab", "bac"
/^xy/	以“xy”開始的字串。	"xyz", "xyab"
/xy\$/	以“xy”結尾的字串。	"axy", "abxy"
[13579]	包含 "1" 或 "3" 或 "5" 或 "7" 或 "9" 的字串,	"a3b", "1xy"
b[aeiou]t	"bat", "bet", "bit", "bot", "but"	"bat", "bit"
[0-9]	含數字之字串	-
[a-z0-9]	含數字或小寫字母之字串	-
[a-zA-Z0-9]	含數字或字母之字串	-
[^0-9]	不含數字之字串	-

方法---特殊字元

通用式	說明	等效正規表達法
^	符合輸入字串的開始位置。	-
\$	符合輸入字串的結束位置。	-
d	匹配一個數字字元。[0-9]。	[0-9]
D	匹配一個非數字字元。	[^0-9]
w	匹配包括底線的任何單詞字元	[A-Za-z0-9_]
W	匹配任何非單詞字元	[^A-Za-z0-9_]
s	匹配任何空白字元, 包括空格、制表符、換頁符	[\r\t\n\f]
S	匹配任何非空白字元	"bxt", "bzt"

方法---數量

通用式	說明	比對成立之字串
*	符合前面的子表達式零次或多次。	zo* 符合「z」、「zo」以及「zoo」。
+	符合前面的子表達式一次或多次。	「zo+」符合「zo」以及「zoo」, 但不能符合「z」。
?	符合前面的子表達式零次或一次。	「do(es)?」符合「does」中的「do」和「does」。?等價於{0,1}。
{n}	n是一個非負整數。符合確定的n次。	「o{2}」不能符合「Bob」中的「o」, 但是符合「food」中的兩個o。
{n,}	n是一個非負整數。至少符合n次。	「o{2,}」不能符合「Bob」中的「o」, 但能符合「fooooood」中的所有o。「o{1,}」等價於「o+」。「o{0,}」則等價於「o*」。
{n,m}	m和n均為非負整數, 其中n<=m。最少符合n次且最多符合m次。	「o{1,3}」符合「foooooood」中的前三個o。「o{0,1}」等價於「o?」。請注意在逗號和兩個數之間不能有空格。



相關function

group(?P<id名稱>要找的,字串...) - 返回群組, 可賦予id名稱尋找

search(要找的, 字串) - 第一個符合的, 回傳位置、那個東西

match(要找的, 字串) - 第一個符合的, 回傳那個東西, 只找首行

findall(要找的, 字串) - 尋找所有符合的, 回傳那個東西

sub(原本的,改成新的,字串) - 替換

spilt(切割的區隔,字串) - 按照我們設定的把資料拆開

compile - 包裝

有空練習

<https://regexone.com/>



RegexOne

Learn Regular Expressions with simple, interactive exercises.

Interact

Lesson 1: An Introduction, and the ABCs

Regular expressions are extremely useful in extracting information from text such as code, log files, spreadsheets, or even documents. And while there is a lot of theory behind formal languages, the following lessons and examples will explore the more practical uses of regular expressions so that you can use them as quickly as possible.

The first thing to recognize when using regular expressions is that **everything is essentially a character**, and we are writing patterns to match a specific sequence of characters (also known as a string). Most patterns use normal ASCII, which includes letters, digits, punctuation and other symbols on your keyboard like %#\$@!., but unicode characters can also be used to match any type of international text.

Below are a couple lines of text, notice how the text changes to highlight the matching characters on each line as you type in the input field below. To continue to the next lesson, you will need to use the new syntax and concept introduced in each lesson to write a pattern that matches all the lines provided.

Go ahead and try writing a pattern that matches all three rows, **it may be as simple as the common letters on each line.**

Exercise 1: Matching Characters

Task	Text
------	------

Match	abcdefg
-------	---------

Match	abcde
-------	-------

Match	abc
-------	-----

Continue >

Solve the above task to continue on to the next problem, or read the [Solution](#).

小組競賽

臺灣南投地方法院民事小額判決105年度投小字第319號

臺灣南投地方法院民事小額判決105年度投小字第319號

原告：泰安產物保險股份有限公司
被告：陳慶隆、陳致安、陳建興、蔡松麟、劉道明、李松季

案由：損害賠償事件

本院於民國105年11月3日言詞辯論終結，茲判決如下：

一、原告起訴主張略以：原告承保訴外人陳頤瑄所有並由其駕駛之車號000-0000號自用小客車（下稱系爭A車），於民國105年12月7日9時許自西向東行駛至南投縣○○鄉○○巷○○號前時（下稱系爭時地），被告於同一時、地駕駛車號000-00號自用小客車（下稱系爭B車）欲由北向西右轉，因轉彎車未禮讓直行車及未盡車前注意義務，致撞擊系爭A車，並致系爭A車受有損害。原告已依保險契約賠付陳頤瑄系爭A車維修費用新臺幣（下同）109,218元（零件56,337元、工資39,316元、塗裝13,565元）。本件車禍發生時，原告駕駛系爭A車雖有未依規定減速、未盡車前注意義務之過失，惟被告駕駛系爭B車亦有右轉未依規定、未盡車前注意義務之過失存在，且系爭B車之過失比例應屬較高，因此原告與被告間之過失比例應為3成、7成等語。爰依保險法第53條第1項，民法第184條第1項前段、第191條之2規定提起本件訴訟等語。並聲明：被告應給付原告76,453元，及自起訴狀繕本送達翌日起至清償日止按週年利率5%計算之利息。

二、被告則以：被告不爭執就本件車禍事故有過失責任存在，惟依本件車禍事故現場照片所顯示系爭A、B車之相對位置，系爭A車應有未依規定減速、未緊靠右側行駛及未盡車前注意義務之疏失，因此雙方過失比例各應各為5成等詞。並聲明：原告之訴駁回。

三、本院之判斷：被告應就本件車禍事故負損害賠償責任：1.按因故意或過失，不法侵害他人之權利者，負損害賠償責任；汽車、機車或其他非依軌道行駛之動力車輛，在使用中加損害於他人者，駕駛人應賠償因此所生之損害。民法第184條第1項前段、第191條之2前段分別規定甚明。次按被保險人因保險人應負保險責任之損失發生，而對於



目標

把裡面3000、n清掉

找出錢錢放到一個list

把幾成都變成幾%

把趴數抓出來放到一個list



小小小提示

\u3000: 我是空格喔、\n: 我是換行符號喔

找出錢錢: findall還記得嗎?

把幾成都變成幾%: sub、compile

把趴數抓出來: findall再來一遍

接起來: for 跟 if 基本功