

正規表示法

2024/12/24

上午10:51

.，所有名稱&所有副檔名=全部檔案，

萬用資源、正則符號(正則表示法) -> 字串、內容比對

RE(Regular Expression):

. -> 後面接著任意一個字元(數字或其它)

[] -> [abc] 三個任意一個或是 [a-c]

[^] -> [^abc] 三個非任意一個

^ -> 從^後開始匹配

三大圖片格式: gif jpg(手機相機)

png(arpg): alpha值 -> 透明度，可去背

shorthand: 捷徑

\d: 數字 -> 用在id、class

\s: 字元

\w: 字元數字含"_"，用在密碼比對

() : (a(bc)(d))，先比abcd都符合再bc再d最後印內容

數量定義詞:

?: <=1個字元，可選的，bool

*: 任意數目的字元，前面字符可出現0或多次

+: >=1個字元

{}: N個數量字元，{N,M}: N~M

例子:

12.*，可匹配12加上任意字符

12.不成立，因為12是兩個字元

12.+不成立，"."至少一個，12要接至少一個字元

1不匹配1.但匹配1*

例子:

^09\d{8} 是從^0{1}9{1}\d{8}簡化

^[A-Z][12]\d{8}

re.findall(pattern, 目標)

例子: 段(\d+-*\d*)，*指沒有也沒關係

例子: str = "a123b456b"

re.findall(r"a(+)b", str) # ['123b456'] -> 貪婪

re.findall(r"a(.*?)b", str) # ['123'] -> 非貪婪

例子: str = "a23b\na34b"

re.findall(r"a(\d+)b.+a(\d+)b", str, re.S) # 不會換行

re.findall(r"a(\d+)b", str, re.M) # ['23', '34']

re.M: ^\$會匹配每一行，不然預設指匹配第一行

網頁換行符號:

(?P<name>pattern): ?跟上面不一樣。將捕獲字元命名

r"^\w\.-]+@[\w\.-]+\.[a-z]{2,6}\$":

\$判斷+後面，因為兩個是同級