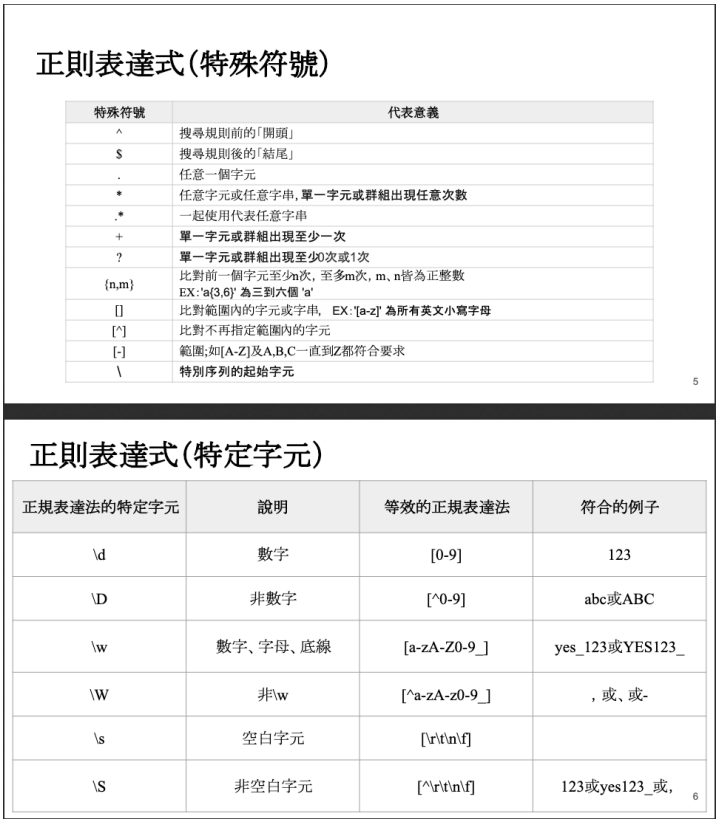
正則表達式

- 正則表達式是透過一些特殊符號來比對字串的方法, 並可對符合比對條件的字串進行搜尋、截取、替代、轉換等等



正則表達式(特殊符號)

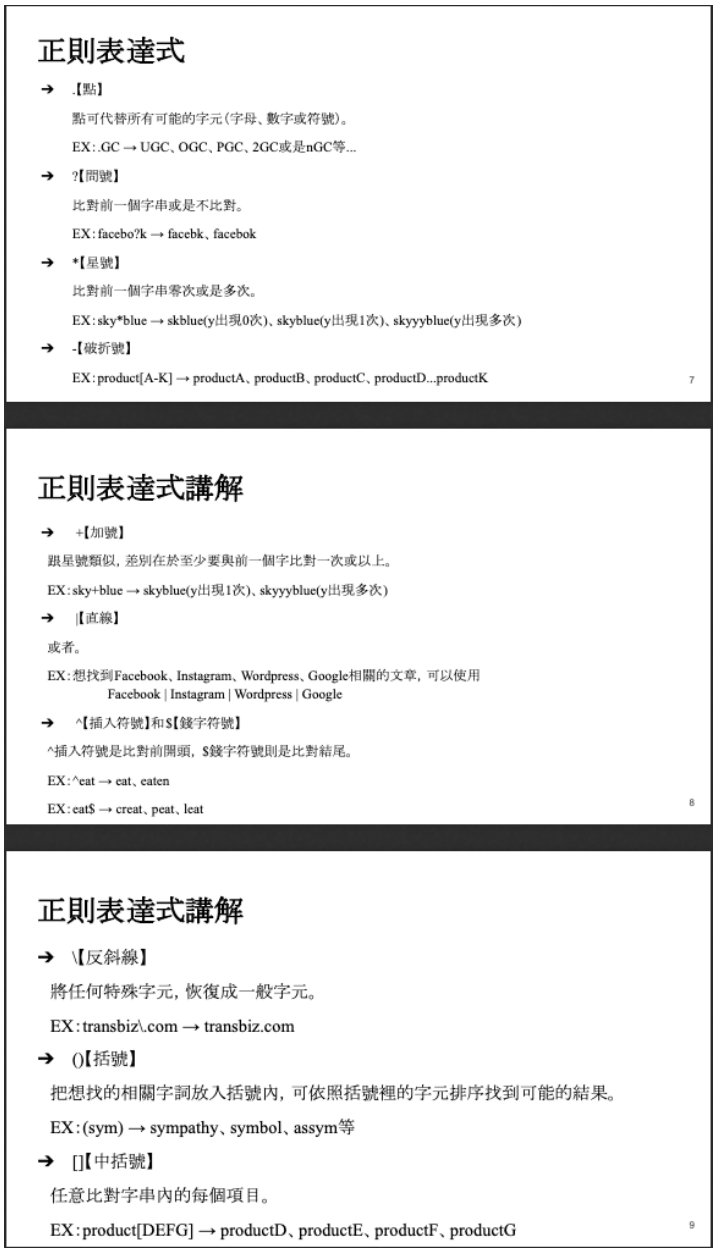
特殊符號	代表意義
^	搜尋規則前的「開頭」
$	搜尋規則後的「結尾」
.	任意一個字元
*	任意字元或任意字串, **單一字元或群組出現任意次數**
.*	一起使用代表任意字串
+	**單一字元或群組出現至少一次**
?	**單一字元或群組出現至少0次或1次**
{n,m}	比對前一個字元至少n次, 至多m次, m、n皆為正整數 EX:'a{3,6}' **為三到六個 'a'**
[]	比對範圍內的字元或字串, EX:'[a-z]' **為所有英文小寫字母**
[^]	比對不再指定範圍內的字元
[-]	範圍;如[A-Z]及A,B,C一直到Z都符合要求
\	**特別序列的起始字元**

5

正則表達式(特定字元)

正規表達法的特定字元	說明	等效的正規表達法	符合的例子
\d	數字	[0-9]	123
\D	非數字	[^0-9]	abc或ABC
\w	數字、字母、底線	[a-zA-Z0-9_]	yes_123或YES123_
\W	非\w	[^a-zA-z0-9_]	，或、或-
\s	空白字元	[\r\t\n\f]	
\S	非空白字元	[^\r\t\n\f]	123或yes123_或,

6



Linux文字處理工具

grep

- 可從資料或檔案中，使用關鍵字或正規表達法(Regex)找出想要的內容
  grep[option]filename

wc文字處理工具

- 計算指定檔案內容的換行數、字數與位元組數wc[option]filename
  - -l 只計算換行數
  - -w 只計算字數
  - -c 只計算位元組數
  - -m 只計算字元數
  - -L 計算最常行的長度

cut文字處理工具

- 逐行擷取部分字元或欄位資料
    - -b:輸出的指定範圍以bytes作為單位
    - -c:輸出的指定範圍以字元數量作為單位
    - -d:指定分隔字元, default為tab做為分隔
    - -f:輸出的指定範圍(每筆data的第幾column作為區分)
    - -s:若該行無分隔字元則不顯示

paste文字處理工具

- 將每個文件以列隊列的方式進行合作paste[option]filename

diff文字處理工具

- 比較文件的弄戎, 特別是兩版本不同的同份文件diff[option]filename1 filename2
    - -y 以並列方式顯示文件的意痛之處
    - -W 使用-y參數時, 指定欄寬
    - -C 前後輸出格式
    - -u 統一格式輸出

文字處理工具

- >: 覆蓋原有檔案
- >>: 追加內容，不覆蓋繼續寫

sort文字處理工具

- 處理文字的排序問題sort[option]filename
    - -f: 忽略大小寫
    - -u: 去除重複資料
    - -r: 反向排序
    - -t: 指定欄位的分隔字元
    - -k: 指定欄位的編號
    - -n: 依照實際數值的大小排序
    - -h: 隊友單位的數值排序
    - -M 依照月份排序

uniq文字處理工具

- 將連續重複文字刪除uniq[option]filename
    - -c: 計算文字行重複次數
    - -s: 將重複行刪除
    - -u: 只輸出沒有重複的文字行
    - -f: 指定要跳過的欄位
    - -s: 跳過每一行開頭幾個字元
    - -i: 忽略大小寫

tr文字處理工具

- 字串替換或刪除tr[option]set1 set2
    - -c 用set1中字符集的補集替換此字符集
    - -d 刪除檔案中所有在set1中出現的字元
    - -s 輸儲檔案中重複且set1中出現的字元, 只保留一個

join文字處理工具

- 將兩個文件中, 指定欄位內容相同的行連結起來join[option] filname1 filename2
    - -1: 連結filename1指定的欄位
    - -2: 連接filename2指定的欄位
    - -t: 使用欄位的分隔符號
    - -i: 忽略大小寫
    - -o: 按指定的格式顯示結果
    - -a: 除顯示結果, 原檔案的其他行也顯示