

正規表示式超集合

正規表示式在編輯器vi/vim和emacs中，以及過濾器grep, egrep, sed, awk與perl中都會使用到。不幸地，它們所使用的都僅是正規表示式中的某一部分集合，且大部分人無法得知一個中介字元是否所有指令都支援。在這個附錄裡整理了所有指令支援／不支援的中介字元，這是經過實際測試後所得的結果。後面也會附上例子方便瞭解。

基本正規表示式

| 符號 | vi | emacs | grep | sed | egrep | awk | perl | Linux | | | | 比對 |
|-----------|----|-------|------|-----|-------|-----|------|-------|------|-----|------|-------------------|
| | | | | | | | | vim | grep | sed | gawk | |
| * | . | . | . | . | . | . | . | . | . | . | . | 前一個字出現零次以上 |
| g* | . | . | . | . | . | . | . | . | . | . | . | 沒有字, g, gg, ggg等等 |
| gg* | . | . | . | . | . | . | . | . | . | . | . | g, gg, ggg等等 |
| . | . | . | . | . | . | . | . | . | . | . | . | 單一字元 |
| .* | . | . | . | . | . | . | . | . | . | . | . | 沒有字或任意個字元 |
| [abc] | . | . | . | . | . | . | . | . | . | . | . | a或b或c |
| [1-3] | . | . | . | . | . | . | . | . | . | . | . | 位數字，在1和3之間 |
| [^Z] | . | . | . | . | . | . | . | . | . | . | . | 除了Z之外的任何字元 |
| [^a-zA-Z] | . | . | . | . | . | . | . | . | . | . | . | 非字母字元 |
| ^DM | . | . | . | . | . | . | . | . | . | . | . | 行首為DM的行 |
| bash\$ | . | . | . | . | . | . | . | . | . | . | . | 行末為bash的行 |

C-2 UNIX終極指南

延伸的正規表示式

| 符號 | vi | emacs | grep | sed | egrep | awk | perl | vim | grep | sed | gawk | 比對 |
|----------------------|----|-------|------|-----|-------|-----|------|-----|------|-----|------|--------------------|
| + | | • | | | • | • | • | | | | • | 前一字元出現一次以上 |
| g+ | | • | | | • | • | • | | | | • | 至少出現一個g |
| g\+ | | | | | | | | • | • | • | | 同上 |
| ? | | • | | | • | • | • | | | | • | 前一個字元出現零或一次 |
| g? | | • | | | • | • | • | | | | • | 沒有字或一個g |
| g\? | | | | | | | | | • | • | | 同上 |
| GIF JPEG | | | | | • | • | • | | | | • | GIF或JPEG |
| GIF\ JPEG | | • | | | | | | • | • | • | | 同上 |
| wood(cock house) | | | | • | • | • | | | | | • | woodcock或woodhouse |
| wood\ (cock\ house\) | | | • | | | | | | • | • | • | 同上 |
| \<pat> | • | • | | | | | | • | • | • | • | 樣式pat出現在字的開頭 |
| pat\> | • | • | | | | | | • | • | • | • | 樣式pat出現在字的結尾 |

區段或標籤式的正規表示式

這些更高深的正規表示式並沒有被egrep和awk支援。gawk和perl支援區段正規表示式，但不需要大括號前的反斜線 \。gawk另外還需要加上選項 **--posix** 或 **-W re-interval** 才行。perl也不需要加上小括號前的反斜線 \。

| 符號 | vi | emacs | grep | sed | egrep | awk | perl | vim | grep | sed | gawk | 比對 |
|--------------|----|-------|------|-----|-------|-----|------|-----|------|-----|------|--------------------------|
| \{m\} | • | | • | • | | | | • | • | • | | 前一個字元出現m次 |
| {m} | | | | | | | • | | | | • | 同上 |
| ^\{9\}nobody | • | | • | • | | | | • | • | • | | 行首跳過9個字元後出現的nobody |
| ^\{9\}nobody | | | | | | | • | | | | • | 同上 |
| \{m.\} | • | | • | • | | | | • | • | • | | 前一個字元至少出現m次 |
| {m,} | | | | | | | • | | | | • | 同上 |
| \{m,n\} | • | | • | • | | | | • | • | • | | 前一個字元出現m次到n次之間 |
| {m,n} | | | | | | | • | | | | • | 同上 |
| \(exp\) | • | • | • | • | | | | • | • | • | | exp表示式，支援標籤\1, \2等來代表exp |
| (exp) | | | | | | | • | | | | | 同上，亦可用 \$1, \$2等 |

跳脫序列

| 符號 | Linux | | | | | | | | | | | 比對 |
|--------|-------|-------|------|-----|-------|-----|------|-----|------|-----|------|-------------------------|
| | vi | emacs | grep | sed | egrep | awk | perl | vim | grep | sed | gawk | |
| \b | | • | | | | | • | | • | • | | 在字組邊界上 |
| wood\b | | • | | | | | • | | • | • | | wood但不是woodcock |
| \B | | • | | | | | • | | • | • | • | 在非字組邊界上 |
| wood\B | | • | | | | | • | | • | • | • | woodcock但非wood |
| \w | | • | | | | | • | • | • | • | • | 一字組字元（如同 [a-zA-Z0-9_]） |
| \W | | • | | | | | • | • | • | • | • | 非字組字元（如同 [^a-zA-Z0-9_]） |
| \d | | | | | | | • | • | | | | 一位元數字（如同 [0-9]） |
| \D | | | | | | | • | • | | | | 非數字（如同 [^0-9]） |
| \s | | | | | | | • | • | | | | 空白字元（包含空格與定位鍵） |
| \S | | | | | | | • | • | | | | 非空白字元（不包含空白及定位鍵） |
| \t | | • | | | | | • | • | | | • | 定位鍵字元 |
| \n | | | | | | | • | | | | • | 新行字元 ([Ctrl-j]) |
| \r | | | | | | | • | | | | • | 回車字元 ([Ctrl-m]) |
| \f | | | | | | | • | | | | • | 進紙字元 ([Ctrl-l]) |
| \nnn | | | | | | | • | | | | • | ASCII八進位值nnn |
| \014 | | | | | | | • | | | | • | ASCII八進位值14 |
| \xnn | | | | | | | • | | | | • | ASCII十六進位值xnn |

POSIX字元類別

| 符號 | Linux | | | | | | | | | | | 比對 |
|-----------|-------|-------|------|-----|-------|-----|------|-----|------|-----|------|------------------|
| | vi | emacs | grep | sed | egrep | awk | perl | vim | grep | sed | gawk | |
| [:alpha:] | | | | | | | | • | • | • | • | 字母字元 |
| [:lower:] | | | | | | | | • | • | • | • | 一個小寫字母字元 |
| [:upper:] | | | | | | | | • | • | • | • | 一個大寫字母字元 |
| [:digit:] | | | | | | | | • | • | • | • | 一個數字字元 |
| [:alnum:] | | | | | | | | • | • | • | • | 一個數字字母字元 |
| [:space:] | | | | | | | | • | • | • | • | 一個空白字元包含進 紙字元 |
| [:cntrl:] | | | | | | | | | • | • | • | 一個控制字元 |
| [:blank:] | | | | | | | | • | • | • | • | 一個空格或定位鍵 |
| [:print:] | | | | | | | | • | • | • | • | 一個可列印字元 |

C-4 UNIX終極指南

| 符號 | Linux | | | | | | | | | | 比對 | |
|--------------|-------|-------|------|-----|-------|-----|------|-----|------|-----|----|--------------------------|
| | vi | emacs | grep | sed | egrep | awk | perl | vim | grep | sed | | gawk |
| [[:punct:]] | | | | | | | | . | . | . | . | 一個標點字元（不含空格、字母、數字以及控制字元） |
| [[:xdigit:]] | | | | | | | | . | . | . | . | 一個十六進位字元 |