

Práctica 5: Expresiones regulares

Sistemas Operativos

Grado en Ingeniería Informática
Departamento de Ingeniería Informática

Universidad de Cádiz

Práctica 5:
Expresiones
regulares

Sistemas
Operativos

Expresiones
regulares

La familia grep

- 1 Expresiones regulares
- 2 La familia grep

Definición

Una expresión regular es un conjunto de caracteres que definen un patrón de texto, con el que debe concordar.

- Las expresiones regulares son utilizadas por utilidades del sistema, tales como los editores **ed**, **vi**, **ex**, y **emacs**, y los filtros **grep**, **egrep**, **sed** y **awk**, entre otros.
- La mayor parte de estos programas buscan, en un fichero o en la entrada estándar, una cadena o secuencia de caracteres que se ajuste a un patrón. Cuando la encuentran decimos que la expresión regular concuerda, equivale a, o coincide con la cadena.

Expresiones regulares II

Práctica 5: Expresiones regulares

Sistemas Operativos

Expresiones regulares

La familia grep

- Una expresión regular es una cadena en la que distinguimos dos tipos de caracteres:
 - **Caracteres ordinarios** Concuerdan consigo mismos.
 - **Caracteres especiales** Tienen un significado especial para el programa que los interpreta.
- Hay caracteres especiales que tienen un significado especial independientemente del lugar de la cadena donde aparezcan: `.` `*` `\` `[`
- Hay otros que sólo tienen un significado especial cuando se sitúan en una posición determinada: el carácter `^` sólo tiene un significado especial al comienzo de una cadena y el carácter `$` sólo lo tiene cuando va al final
- Si queremos usar un carácter especial dentro de una cadena como carácter ordinario deberemos protegerlo con `\`

Expresiones regulares básicas

Práctica 5: Expresiones regulares

Sistemas Operativos

Expresiones regulares

La familia grep

Hacen uso de los siguientes caracteres especiales:

Carácter	Significado
.	Concuerda con cualquier carácter simple.
*	Concuerda con 0 o más ocurrencias de la expresión regular precedente.
\$	Cuando es el último carácter de la expresión regular, concuerda con el final de línea.
^	Cuando es el primer carácter de la expresión regular, concuerda con el principio de línea.
[conjunto]	Concuerda con un carácter simple del <i>conjunto</i> .
[c1-c2]	Concuerda con el conjunto de caracteres ASCII comprendidos en el rango de c1 a c2.
[^conjunto]	Concuerda con cualquier carácter simple que no esté en el <i>conjunto</i> .

Expr. regular	Texto
<code>.ora</code>	ahora hora horas oración morada
<code>10*</code>	1 10 11 100 1000 210 401 3331
<code>^[A-Z]</code>	Concuerda con cualquier línea que empiece por mayúsculas
<code>[^0-9]</code>	Concuerda con cualquier carácter que no sea un dígito
<code>\.\$</code>	Concuerda con cualquier línea que termine en punto

Pueden hacer uso de los caracteres que forman parte de las expresiones regulares básicas más los que aparecen en la tabla siguiente:

Carácter	Significado
+	Concuerta con una o más ocurrencias de la expresión regular precedente
?	Concuerta con cero o una ocurrencia de la expresión regular precedente
<i>expr1 expr2</i>	Concuerta con cualquiera de las dos expresiones regulares
<i>(expr)</i>	Se utiliza para agrupar una expresión regular

Expr. regular	Texto
35+	3 35 355 1355 21735 315
ab?c	abc ac aab b cc aabcc
juan Juan	juan Juan
a(bc)?d	ad abcd abc abd acd

- Está formada por 3 filtros: grep, egrep y fgrep
- `grep [opciones] expr-regular [fichero, ...]`
- `egrep [opciones] expr-regular [fichero, ...]`
- `fgrep [opciones] cadenas [fichero, ...]`
- Diferencias:
 - **fgrep** sólo admite cadenas literales
 - **grep** admite expresiones regulares básicas
 - **egrep** admite expresiones regulares básicas y extendidas
- Todos ellos seleccionan aquellas líneas de la entrada que concuerdan con la expresión regular o la cadena de caracteres especificada y las muestran en la salida estándar.

La familia grep II

Práctica 5:
Expresiones
regulares

Sistemas
Operativos

Expresiones
regulares

La familia grep

Todas las órdenes de la familia grep admiten las siguientes opciones:

- **-n** Precede las líneas seleccionadas del nº de línea que le corresponde dentro del fichero
- **-v** Muestra las líneas que **no** concuerdan
- **-c** Sólo visualiza el nº de líneas que concuerdan y no su contenido
- **-w** Exige que la expresión regular concuerde con una palabra y no con parte de ella