

PRÁCTICA 4: *Comandos para filtrar ficheros en bash*

ALUMNO:

1. INTRODUCCIÓN

Después de ver los comandos para paginar, visualizar y editar ficheros, así como de ver aquellos que realizan búsquedas de ficheros y patrones, vamos a ver cómo filtrar ficheros en *bash*. Los principales comandos para este propósito son *sort*, *uniq*, *wc*, *cut*, *sed* y *awk*.

1.1. *Sort*

El comando ***sort*** ordena las líneas de un fichero mostrándolas por la salida estándar. De no especificarse un fichero toma la entrada estándar.

sort [opciones] [ficheros]

Algunas opciones son las siguientes:

Opción	Descripción
-r	Ordena al revés
-f	Ingresa las diferencias entre mayúsculas y minúsculas
-g	Ordena de forma numérica

1.2. *Uniq*

El comando *uniq* elimina las líneas repetidas de un fichero ordenado, imprimiéndolo por la salida estándar o en otro fichero argumento. De no especificarse un fichero toma la entrada estándar.

uniq [opciones] [ficheros] [salida]

Algunas opciones son las siguientes:

Opción	Descripción
-c	Utiliza como prefijo en cada línea su número de ocurrencias
-d	Solo imprime las líneas duplicadas

1.3. *wc*

El comando `wc` imprime el número de líneas, palabras y bytes de uno o varios ficheros. Si son varios ficheros hace también un resumen de los totales. De no especificarse un fichero toma la entrada estándar.

wc [opciones] [ficheros]

Algunas opciones son las siguientes:

Opción	Descripción
-l	Solo cuenta las líneas
-c	Solo cuenta los <i>bytes</i>
-w	Solo cuenta las palabras

1.4. *cut*

El comando `cut` nos permite cortar una línea de texto, para obtener un subconjunto en lugar de la línea completa. Podemos cortar por número de caracteres, por campos, etc.

cut [opciones] [ficheros]

Algunas opciones son las siguientes:

Opción	Descripción
-c N-M	Corta desde el carácter número N hasta el número M
-c N-	Corta desde el carácter número N hasta el final
-c -N	Corta desde el principio del archivo hasta el carácter número N
-c N,M	Corta el carácter número N y el carácter número M
-d ":" -f 1	Separa la línea en campos divididos por el carácter : y nos muestra solo el primer campo
-d "-" -f 3	Separa la línea en campos divididos por el carácter – y nos muestra solo el tercer campo

Por ejemplo, si queremos extraer la primera palabra, siendo el espacio en blanco el delimitador de las palabras usamos los modificadores `-d " "` para indicar cuál es el delimitador y `-f 1` para indicar que queremos el primer campo:

```
$ echo "Salió un ratón barbicano, colilargo, hociquirromo" | cut -d " " -f 1
Salió
```

1.5. *sed*

Sed realmente es un mini lenguaje de programación capaz de hacer sustituciones realmente complejas de cadenas de texto.

Sed -n -e "comandos" -f archivo [opciones] [ficheros]

Algunas opciones son las siguientes:

Opción	Descripción
-n	No produce la salida estándar
-e "comandos"	Indica qué comandos va a ejecutar <i>sed</i>
-f	Si queremos que los comandos se tomen de un archivo

1.6. *awk*

Dejaremos este comando para la siguiente práctica por su complejidad.

2. OBJETIVO

El objetivo de la práctica es que os familiaricéis con los comandos que sirven para filtrar archivos en *bash*.

3. EJERCICIOS

3.1. Cuenta el número total de líneas del archivo *fichero.txt*.

```
joseangel@joseangel-virtualbox:~/Desktop$ cd Desktop/
joseangel@joseangel-virtualbox:~/Desktop$ wc -l fichero.txt
70 fichero.txt
joseangel@joseangel-virtualbox:~/Desktop$
```

3.2. Cuenta el número total de palabras en *allstar.txt*.

```
joseangel@joseangel-virtualbox:~/Desktop$ wc -w allstar.txt
286 allstar.txt
```

3.3. Cuenta el número de veces que aparece la palabra “de” en el archivo *allstar.txt* con el siguiente comando:

```
joseangel@joseangel-virtualbox:~/Desktop$ grep -o -i "de" allstar.txt | wc -l  
33
```

`grep -o -i "de" allstar.txt | wc -l`

3.4. Ordena las líneas del archivo *fichero.txt* de manera inversa.

```
joseangel@joseangel-virtualbox:~/Desktop$ sort -r fichero.txt  
5  
9  
8  
7  
6  
5  
40      camaleón  
4  
39      cayó  
38      país  
37      soldó  
36      vagón  
35      anís  
34      desolación  
33      organización  
32      Abrió  
31      Ángel  
30      camión  
3  
29  
28  
27  
26  
25  
24
```


3.8. Muestra las líneas únicas de *fichero.txt* y comprueba que las 10 líneas que has añadido ya no aparecen. Utiliza los comandos:

`sort ficheros.txt | uniq -c`

```
joseangel@joseangel-virtualbox:~/Desktop$ sort fichero.txt | uniq -c
15 1
1 10
1 11
1 12
1 13
1 14
1 15
1 16
1 17
1 18
1 19
1 2
1 20
1 21
1 22
1 23
1 24
1 25
1 26
1 27
1 28
1 29
1 3
```

3.9. Extrae solo la primera columna del archivo *allstar.txt* con el comando *cut*, como en el ejemplo realizado en esta práctica en el apartado 1.

```
joseangel@joseangel-virtualbox:~/Desktop$ cut -f1 allstar.txt
```

```
la cancha de los Pacers, acoge esta madrugada el  
'All Star Game' (2.00 h, Mov+),  
la gran fiesta anual del baloncesto de la NBA que reúne a  
los mejores jugadores del mundo.  
Una constelación de estrellas juntas en un partido en el  
que brillará con luz propia LeBron James,  
que a sus 39 años se convertirá en el jugador con  
más presencias en este evento (20) superando la marca  
del legendario Kareem Abdul-Jabbar (19).
```

3.10. Extrae ahora la segunda palabra de cada línea haciendo lo mismo.

```
de 1993 con un tal Michael Jordan.  
joseangel@joseangel-virtualbox:~/Desktop$ cut -f2 -d' ' allstar.txt  
Gainbridge  
la  
Star  
la  
mejores  
Una  
brillará  
que  
presencias  
del  
  
detrás  
detrás  
Por  
Tim  
todos  
único  
incombustible  
esta
```

3.11. Ahora vamos a ver cómo reemplazar dos palabras en un archivo. Si queremos reemplazar la palabra *baloncesto* por *basket* debemos introducir el siguiente comando con *sed*:

sed 's/palabra1/palabra2/g' allstar.txt

```
joseangel@joseangel-virtualbox:~/Desktop$ sed s/baloncesto/basket/g allstar.t  
xt
```

3.12. En las anteriores prácticas hemos tenido problemas para eliminar líneas en blanco de un documento. Con *sed* podemos solucionarlo con el comando:

sed '/^\$/d' fichero.txt

```
joseangel@joseangel-virtualbox:~/Desktop$ sed '/^$/d' allstar.txt  
El Gainbridge Fieldhouse de Indianápolis,  
la cancha de los Pacers, acoge esta madrugada el  
'All Star Game' (2.00 h, Mov+),  
la gran fiesta anual del baloncesto de la NBA que reúne a  
los mejores jugadores del mundo.  
Una constelación de estrellas juntas en un partido en el  
que brillará con luz propia LeBron James,  
que a sus 39 años se convertirá en el jugador con  
más presencias en este evento (20) superando la marca  
del legendario Kareem Abdul-Jabbar (19).  
Por detrás quedan ya el malogrado Kobe Bryant (18),  
Por detrás quedan ya el malogrado Kobe Bryant (18),  
Por detrás quedan ya el malogrado Kobe Bryant (18),  
Tim Duncan, Kevin Garnett y Shaquille O'Neal,  
todos ellos con 15 presencias, y Kevin Durant (14),  
único jugador en activo junto al  
incombustible LeBron James entre los 24 primeros  
de esta prestigiosa nómina de estrellas.  
Será un escaparate planetario para el lucimiento  
de los grandes candidatos al MVP de la  
temporada. No está entre ellos 'King' James,  
al menos de momento salvo imprevista machada  
de sus Lakers, pero sí otros como Giannis Antetokounmpo
```


3.13. Muestra las l neas del fichero *allstar.txt* que tienen al menos 20 caracteres con los comandos

`sed -n '/.\{50\}/p' allstar.txt`

```
joseangel@joseangel-virtualbox:~/Desktop$ sed -n '/.\{50\}/p' allstar.txt
la gran fiesta anual del baloncesto de la NBA que re ne a
Una constelaci3n de estrellas juntas en un partido en el
que a sus 39 a os se convertir  en el jugador con
m s presencias en este evento (20) superando la marca
    Por detr s quedan ya el malogrado Kobe Bryant (18),
    Por detr s quedan ya el malogrado Kobe Bryant (18),
    Por detr s quedan ya el malogrado Kobe Bryant (18),
todos ellos con 15 presencias, y Kevin Durant (14),
incombustible LeBron James entre los 24 primeros
    Ser  un escaparate planetario para el lucimiento
de sus Lakers, pero s  otros como Giannis Antetokounmpo
(Bucks), Jayson Tatum (Celtics), Nikola Jokic (Nuggets),
Luka Doncic (Mavericks) o Shai Gilgeous-Alexander (Thunder),
aspirantes al cetro de mejor jugador de la temporada en
ausencia de Joel Embiid (Sixers), m ximo candidato hasta que
Recupera el 'All Star Game' el tradicional formato de Este
vs.Oeste, que no se viv a desde 2017, tras el fracaso de
audiencias de los  ltimos a os con la elecci3n de los equipos
por los capitanes sin distinci3n de Conferencias. El pasado
a o s lo 4,6 millones de espectadores siguieron el partido
en la edici3n de Utah. Muy lejos, a a os luz, quedan los 23
millones de audiencia r cord que se registraron en la edici3n
joseangel@joseangel-virtualbox:~/Desktop$
```

3.14. Muestra l neas de *fichero.txt* que tengan al menos 3 caracteres y explica lo que pasa.

```
joseangel@joseangel-virtualbox:~/Desktop$ grep -E '^\{3,\}' fichero.txt
joseangel@joseangel-virtualbox:~/Desktop$
```

No sale nada porque no existen palabras de 3 caracteres

3.15. Pon el siguiente comando para *fichero.txt* y di lo que pasa:

`sed '/[02468]$/d' fichero.txt`

```
joseangel@joseangel-virtualbox:~/Desktop$ sed '/[02468]$/d' fichero.txt
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
1
3
5
7
9
11
```

Se eliminan las líneas pares

3.16. Haz lo mismo con los impares.

```
joseangel@joseangel-virtualbox:~/Desktop$ sed '/[13579]$/d' fichero.txt
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
40
```


3.19. Con el comando *sort* también podemos contar el número de veces que aparece un determinado número en *fichero.txt*. Utiliza el siguiente comando para contar:

sort fichero.txt | uniq -c

```
joseangel@joseangel-virtualbox:~/Desktop$ sort fichero.txt | uniq -c
15 1
18 10
1 11
1 12
1 13
1 14
1 15
1 16
1 17
1 18
1 19
1 2
1 20
1 21
1 22
1 23
1 24
1 25
1 26
1 27
1 28
1 29
1 3
```

3.20. Añade diez líneas en *allstar.txt* que comiencen por un espacio en blanco.

```
joseangel@joseangel-virtualbox:~/Desktop$ cat allstar.txt
    El Gainbridge Fieldhouse de Indianápolis,
    la cancha de los Pacers, acoge esta madrugada el
    'All Star Game' (2.00 h, Mov+),
    la gran fiesta anual del baloncesto de la NBA que reúne a
    los mejores jugadores del mundo.
    Una constelación de estrellas juntas en un partido en el
    que brillará con luz propia LeBron James,
    que a sus 39 años se convertirá en el jugador con
    más presencias en este evento (20) superando la marca
    del legendario Kareem Abdul-Jabbar (19).
```

3.21. Elimina las líneas anteriores con el comando

`sed '/^[:space:]]/d' allstar.txt`

```
joseangel@joseangel-virtualbox:~/Desktop$ sed '/^[:space:]]/d' allstar.txt

de los grandes candidatos al MVP de la
de sus Lakers, pero sí otros como Giannis Antetokounmpo

Recupera el 'All Star Game' el tradicional formato de Este
audiencias de los últimos años con la elección de los equipos
```

3.22. De forma parecida al ejercicio anterior se pueden eliminar las líneas que comienzan por mayúscula. Simplemente cambia del comando anterior *space* por *upper*.

```
joseangel@joseangel-virtualbox:~/Desktop$ sed '/^[[:upper:]]/d' allstar.txt
El Gainbridge Fieldhouse de Indianápolis,
la cancha de los Pacers, acoge esta madrugada el
'All Star Game' (2.00 h, Mov+),
la gran fiesta anual del baloncesto de la NBA que reúne a
los mejores jugadores del mundo.
Una constelación de estrellas juntas en un partido en el
que brillará con luz propia LeBron James,
que a sus 39 años se convertirá en el jugador con
más presencias en este evento (20) superando la marca
del legendario Kareem Abdul-Jabbar (19).

    Por detrás quedan ya el malogrado Kobe Bryant (18),
    Por detrás quedan ya el malogrado Kobe Bryant (18),
    Por detrás quedan ya el malogrado Kobe Bryant (18),
Tim Duncan, Kevin Garnett y Shaquille O'Neal,
todos ellos con 15 presencias, y Kevin Durant (14),
único jugador en activo junto al
```