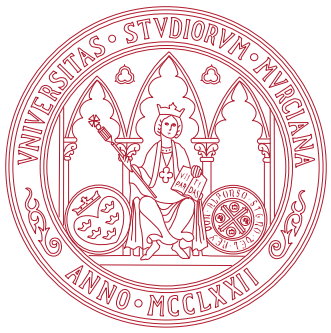

Universidad de Murcia

Facultad de Informática

Departamento de Ingeniería y Tecnología de Computadores

Área de Arquitectura y Tecnología de Computadores

PRÁCTICAS DE I.S.O.

2º de Grado en Ingeniería Informática

Boletín de prácticas 2 – Manejo avanzado del shell de Linux

Curso 2017/2018

Índice

1. Introducción

Partiendo de las destrezas en el manejo de órdenes básicas, adquiridas en la asignatura Fundamentos de Computadores de primer curso y repasadas en el boletín 1, en este boletín veremos primero cómo podemos modificar el modo en el que interactuamos con una orden redirigiendo su entrada y/o salidas de datos a ficheros y/o otras órdenes. Tras ello, describiremos algunas órdenes que son de gran utilidad para filtrar y procesar la información generada por una orden o almacenada en ficheros.

1.1. Objetivos

Al terminar el boletín el alumno debe ser capaz de:

- Manejar correctamente los redireccionamientos, tanto de la entrada como de las salidas de una orden.
- Integrar diferentes órdenes mediante la interconexión de sus entradas y salidas estándares a través de tuberías.
- Usar apropiadamente las principales ordenes de filtrado y procesamiento de datos, tanto individualmente como combinadas a través de tuberías.

1.2. Órdenes utilizadas

Las órdenes que usaremos para resolver estos boletines son:

- `grep`
- `tr`
- `uniq`
- `tail`
- `tee`
- `sort`
- `cut`
- `head`
- `wc`
- `column`

En las páginas de manual de cada una de estas órdenes encontrarás información detallada de cómo usarlas.

2. Entrada/salida estándar y redirección

La filosofía de UNIX/Linux es en extremo modular. Se prefieren las herramientas pequeñas que realizan tareas concretas a las macro-herramientas que realizan de todo. Para completar el modelo, es necesario proporcionar mecanismos que permitan ensamblar estas herramientas pequeñas de tal forma que sea posible realizar un procesamiento complejo de la información. Estos mecanismos son el redireccionamiento de las entradas y salidas estándares de los procesos y las tuberías. A continuación describimos ambos.

Habitualmente, los procesos disponen de tres descriptores de fichero[1] a través de los que se comunican con otros procesos y con el usuario. Estos tres descriptores son:

- Descriptor 0, conocido como *entrada estándar*: los procesos leen de este descriptor para recibir datos de entrada. Normalmente, el descriptor 0 está asociado a la entrada del terminal en la que se está ejecutando el proceso, es decir, al teclado.
- Descriptor 1, conocido como *salida estándar*: los procesos escriben en este descriptor para mostrar sus resultados o datos de salida. Normalmente, este descriptor está asociado a la salida del terminal en el que se está ejecutando el procesos, es decir, la pantalla.
- Descriptor 2, conocido como *salida estándar de error*: los procesos escriben en este descriptor para mostrar mensajes de error que indiquen la causa de algún fallo. Al igual que el descriptor anterior, normalmente, este descriptor está asociado a la salida del terminal en el que se está ejecutando el procesos, es decir, la pantalla.

[1] En UNIX/Linux, cuando se realiza una llamada al sistema `open` para abrir un fichero, el núcleo del sistema operativo devuelve un número para operar con ese fichero. A este número se le llama *descriptor de fichero*.

Aunque, como hemos dicho, estos tres descriptores están asociados a la terminal en la que se ejecuta un proceso, es posible *redireccionarlos* para:

- Almacenar los datos de salida en un fichero determinado.
- Recibir los datos de entrada de un fichero concreto.
- Comunicar unos procesos con otros, de forma que trabajen como una unidad, haciendo cada uno una tarea especializada.

Cuando redireccionamos, lo que hacemos es cambiar el fichero al que representa un descriptor. Por ejemplo, cuando redireccionamos la entrada estándar, lo que hacemos es que el descriptor 0 deje de estar asociado al teclado para estar asociado a, por ejemplo, un fichero. Así, cuando un proceso lea del descriptor 0 para recibir nuevos datos de entrada, lo que ocurrirá es que recibirá dichos datos del fichero y no del teclado. Es más, el proceso ni se dará cuenta del cambio pues para él el descriptor 0 sirve para recibir datos de entrada, sin importarle de dónde vienen esos datos.

Es importante remarcar que toda la labor de redireccionamiento de la entrada/salida de un proceso la realiza el shell que estemos utilizando, siendo totalmente transparente y ajena a dicho proceso. Como acabamos de explicar, el proceso recibirá datos de su entrada estándar y escribirá sus resultados y mensajes de error en su salida estándar y salida estándar de error, respectivamente, sin importarle qué sea en cada momento esa entrada o esas salidas estándares.

Hay varios operadores para redireccionar la entrada y las salidas de un proceso:

- >: redirecciona la salida estándar a un fichero; si el fichero existe, lo sobrescribe:

```
$ who > usuarios.txt
# Escribe en usuarios.txt el listado de usuarios conectados
```

- >>: redirecciona la salida estándar a un fichero; si el fichero existe, añade los datos al final del mismo.
- 2 >: redirecciona la salida estándar de error a un fichero; si el fichero existe, lo sobrescribe:

```
$ find / -type d  2>errores.txt
# Busca todos los directorios existente a partir del directorio '/'
# Los mensajes de error los escribe en el fichero errores.txt
```

- 2 >> : similar a >> pero para la salida estándar de error.
- `n>&m`: redirecciona el descriptor de fichero `n` al descriptor de fichero `m`; en caso de que `n` se omita, se sobrentiende un 1 (es decir, la salida estándar):

```
$ cat file directorio > salida.txt 2>&1
# Redirecciona la salida estándar al fichero salida.txt y la salida
# estándar de error a la salida estándar. El resultado es que toda
# la salida genera por la orden va al mismo fichero.
```

- <: redirige la entrada estándar a un fichero:

```
$ grep cadena < fichero.txt
# Busca "cadena" dentro de fichero.txt
```

- | (tubería): redirecciona la salida estándar de una orden a la entrada estándar de la orden que le sigue:

```
$ who | grep pilar
# Busca "pilar" entre el listado de usuarios conectados
```

- `|&`: tubería en la que se redirecciona la salida estándar de error de una orden a la entrada estándar de la orden que le sigue usando:

```
$ find /etc |& grep Permission
# Busca aquellas entradas a partir de /etc para las que no se
# tiene permiso de acceso
```

La tubería es quizás el tipo de redirección más importante, puesto que se usa para integrar diferentes órdenes y programas, mediante la interconexión de sus entradas y salidas estándares. Más concretamente, con una *tubería* o *pipe* (símbolo |) hay varias órdenes que se ejecutan de manera interrelacionada, de forma que la salida estándar de la primera orden se envía a la entrada estándar de la segunda, la salida estándar de la segunda orden se envía a la entrada estándar de la tercera, y así sucesivamente.y así hasta que se termine de ejecutar la última orden:

```
$ orden 1 | orden 2 | ... | orden n
```

Las siguientes secciones de este boletín describen la funcionalidad de diversas órdenes para el procesamiento y el filtrado de información. A estas órdenes se les llama *filtros* porque son capaces de «filtrar» la información que reciben para seleccionar y/o procesar sólo aquella información que nos interesa. Es más, puesto que estas órdenes pueden recibir la información por su entrada estándar y mostrar el resultado por su salida estándar, pueden colocarse en cualquier lugar de una tubería para filtrar la información que reciben de la orden que queda a su izquierda en la tubería antes de pasar dicha información a la orden que queda a su derecha.

3. Orden `grep`

La orden `grep` constituye una útil herramienta para buscar cadenas de texto en ficheros. Su sintaxis es:

```
grep [opciones] patrón [fichero...]
```

Por defecto, esta orden realiza una búsqueda del patrón de texto especificado en uno o más ficheros, mostrando todas las líneas de texto de estos ficheros que contienen dicho patrón. Si se especifica más de un fichero, esta orden muestra justo antes de cada línea encontrada el nombre del fichero al que pertenece.

Si no se especifica ningún fichero, esta orden toma la entrada estándar del sistema como entrada de datos. Ejemplos:

```
$ grep calle datos/clientes
calle mayor, 12, murcia
calle europa, 22, alicante
calle verde, 33, almeria

$ grep calle datos/*
datos/clientes: calle mayor, 12, murcia
datos/clientes: calle europa, 22, alicante
datos/clientes: calle verde, 33, almeria
datos/proveedores: calle roja, 11, madrid
datos/proveedores: calle asia, 9, barcelona
```

Las opciones más comunes para esta orden son:

- `-i`: Ignora las diferencias entre mayúsculas y minúsculas, las considera equivalentes.
- `-n`: Muestra las líneas y el número de cada línea.

- `-c`: Muestra la cuenta de líneas coincidentes, pero no las líneas en sí.
- `-l`: Muestra los nombres de los ficheros con líneas coincidentes, pero no las líneas en sí.
- `-h`: Muestra las líneas coincidentes, pero no los nombres de ficheros.
- `-v`: Muestra las líneas que no coinciden.
- `-w`: Muestra las líneas que contienen el patrón como una palabra completa.
- `-x`: Muestras las líneas que coinciden completamente con el patrón buscado.
- `-o`: Muestra sólo las partes coincidentes de una línea (no líneas coincidentes enteras) mostrando cada parte coincidente en una línea separada.

En los patrones de búsqueda se pueden utilizar expresiones regulares[2], construida mediante ciertos caracteres especiales. Los más habituales son:

- `.`: Un único carácter cualesquiera.
- `c*`: Cero o más ocurrencias del carácter indicado, `c`.
- `c\+`: Una o más ocurrencias del carácter indicado, `c`.
- `^`: Comienzo de línea.
- `$`: Final de línea.
- `[...]`: Uno de entre un conjunto de caracteres. En el interior de esta expresión, el carácter «`^`» tiene el significado de negación cuando aparece al principio, justo detrás de «[».
- `c\{n,m\}`: Entre n y m repeticiones del carácter indicado, `c`.
- `\`: Para deshabilitar el significado especial del carácter que se indique a continuación.

Puesto que algunos de estos caracteres especiales tienen también un significado especial para Bash, cuando los usemos en un patrón, dicho patrón deberíamos encerrarlo entre comillas simples para evitar que Bash los interprete y así lleguen intactos a la orden `grep`.

3.1. Ejercicios

A lo largo de los siguientes ejercicios se utilizarán algunos ficheros de texto cuyo contenido, a modo de ejemplo, podría ser el que se muestra a continuación. Puedes crear tú mismo estos ficheros con este contenido o con el que estimes oportuno.

```
$ cat compras/compras_nacionales

manzanas Lleida 200 Kilogramos
peras Zaragoza 150 kilogramos
manzanas Girona 10 kilogramos
naranjas Valencia 500 Kilogramos
uvas Alicante 1000 Kilogramos
uvas Logroño 300 Kilogramos

$ cat compras/compras_europeas
```

[2]Una expresión regular es una serie de caracteres que forman un patrón, normalmente representativo de otro grupo de caracteres mayor. Por ejemplo, el grupo formado por las cadenas Handel, Händel y Haendel se describe mediante el patrón `"H[aäae]ndel"`.

```
manzanas Italia 1000 kilogramos
uvas Francia 500 Kilogramos

$ cat compras/compras_extracomunitarias

manzanas Chile 2000 Kilogramos
peras Argentina 1500 Kilogramos
manzanas Argentina 100 kilogramos
naranjas Chile  500 kilogramos
uvas Marruecos 1000 Kilogramos
uvas Argelia 3000 Kilogramos

$ cat compras/proveedores

Chile Frutosa 34
Argentina Agricolasa
Francia BFF 54
Italia FDI
```

1. Busca la cadena «manzanas» en el fichero `compras/compras_nacionales`, mostrando cada línea que contenga esta cadena.
2. Busca la cadena «manzanas» en todos los ficheros del directorio `compras`.
3. Lista los nombres de todos los ficheros del directorio `compras` que contienen la cadena «naranjas». Se debe listar los nombres de los ficheros, pero no las líneas de éstos que contienen la cadena buscada.
4. Busca la cadena «peras» en todos los ficheros que empiezan por `compras` del directorio `compras`, ignorando la distinción entre mayúsculas y minúsculas.
5. Busca la cadena «peras» en todos los ficheros que empiezan por `compras` del directorio `compras`, ignorando la distinción entre mayúsculas y minúsculas. Muestra las líneas que contengan esta cadena, pero no los nombres de los ficheros.
6. Busca la cadena «1500 Kilogramos» en todos los ficheros que empiezan por `compras` del directorio `compras`. Muestra las líneas que contengan esta cadena y sus respectivos números de línea dentro del fichero.
7. Muestra todas las líneas del fichero `compras/compras_extracomunitarias` que no contienen la cadena «uvas».
8. Cuenta el número de veces que aparece la cadena «manzanas» en cada uno de los ficheros del directorio `compras`.
9. Muestra los datos de todas las compras de frutas que se hayan realizado en cantidades de 10 ó 100 ó 1000, ... kilogramos.

   ```
   $ grep "100* [Kk]ilogramos" compras/*
   ```

10. Busca todas las líneas del fichero de proveedores que contengan un dígito.
11. Muestra los datos de todas las compras de frutas que se hayan realizado en cantidades múltiplo de 1000 kilogramos, mayores de 1000 kilogramos y menores de 10000 kilogramos.

    ```
    $ grep "[^1]000 [Kk]ilogramos" compras/*
    ```

12. Muestra todas las líneas del fichero `compras/compras_nacionales` que empiecen con una letra «m».

13. Encuentra todos los ficheros del directorio `compras` que acaben en `les`.

```
$ ls -l compras | grep 'les$'
```

14. Muestra todos los subdirectorios del directorio actual que tengan permiso de ejecución para otros usuarios.

```
$ ls -l | grep 'd........x'
```

4. Orden `sort`

Esta orden sirve para ordenar las líneas de un fichero alfabéticamente o numéricamente. Puede considerar como clave de ordenación cada línea del fichero al completo, o bien, uno o más campos de cada línea. Por defecto, el delimitador entre campos es el espacio en blanco. Su sintaxis es:

```
sort [opciones] [fichero]...
```

4.1. Ordenaciones sencillas

Por defecto, `sort` ordena alfabéticamente las líneas de los ficheros. Si no se especifica ningún fichero, entonces ordena la líneas que recibe por su entrada estándar.

Opciones de ordenación más comunes:

- `-n`: Ordena numéricamente, incluyendo números negativos y con decimales. En el caso de los números con decimales, hay que tener en cuenta el idioma usado en el terminal, ya que en inglés el separador de decimales es el punto «.» y en español la coma «,». También hay que tener en cuenta que `sort` considera que un número termina cuando aparece el primer carácter no numérico, incluyendo los espacios en blanco.
- `-r`: Orden inverso.

Por ejemplo, en orden alfabético:

```
$ cat f1
.this line begins with a period
a line that begins with lowercase a.
This is a line.
abracadabra
1234
Where will this line sort?
A line that begins with uppercase a.

$ sort f1
1234
abracadabra
a line that begins with lowercase a.
A line that begins with uppercase a.
This is a line.
.this line begins with a period
Where will this line sort?
```

En orden numérico:

```
$ cat n1
-18
18
0
-1,4
0,54
0,0
3
0,1
$ sort -n n1
-18
-1,4
0
0,0
0,1
0,54
3
18
```

4.2. Ordenaciones por columnas

Con `sort` también se puede ordenar el contenido de un fichero de acuerdo a las columnas (campos) de sus líneas, usando la opción `-k`. Por ejemplo:

```
$ cat fic
Susan Jones
Jill Zane
John Smith
Andrew Carter

$ cat fic_comas
Susan,Jones
Jill,Zane
John,Smith
Andrew,Carter

$ cat fic_largo
Susan Jones 123 Plaza Mayor
Jill Zane 34 Calle Real
John Smith 455 Calle Libertad
Andrew Carter 344 Plaza Agosto
```

La siguiente ejecución de `sort` ordenará el fichero `fic_largo` a partir del cuarto campo inclusive:

```
$ sort -k4 fic_largo
John Smith 455 Calle Libertad
Jill Zane 34 Calle Real
Andrew Carter 344 Plaza Agosto
Susan Jones 123 Plaza Mayor
```

mientras que la siguiente ordenará sólo por el cuarto campo (observa que el resultado no coincide con el de la orden anterior):

```
$ sort -k4,4 fic_largo
Jill Zane 34 Calle Real
```

```
John Smith 455 Calle Libertad
Andrew Carter 344 Plaza Agosto
Susan Jones 123 Plaza Mayor
```

Esta opción se puede combinar con las opciones de ordenación general anteriormente descritas, como `-n` y `-r`. En ese caso, estas opciones serán las usadas por defecto para la ordenación de los campos, salvo que se indique otra cosa para cada campo, como veremos ahora. De igual forma, se puede especificar el carácter que marca la separación entre los campos, con la opción `-t`, así como indicar la ordenación por varios campos en un orden de prioridad. Por ejemplo:

- Ordena el fichero `fic` numéricamente empezando la clave de ordenación por el primer carácter del primer campo y continuando hasta el final de cada línea. O sea, es equivalente a una ordenación general numérica:

  ```
  $ sort -k 1n fic
  ```

- Ordena el fichero `fic_comas` únicamente por el primer campo, tomando el carácter «,» como el delimitador entre los campos:

  ```
  $ sort -k 1,1 -t, fic_comas
  ```

- Ordena el fichero `fic_largo` usando como clave los campos del 3 al 5 y a continuación el campo 2:

  ```
  $ sort -k 3,5 -k 2,2 fic_largo
  ```

Uso de la opción `-k` en `sort`

Observa que `-k3,5` no es lo mismo que `-k3,3 -k4,4 -k5,5`. En el primer caso, hay una única clave de ordenación formada por los campos del 3 al 5 de cada línea, *incluyendo* todos los caracteres que separen dichos campos. Es decir, los caracteres de separación (espacios, tabuladores, etc.) forman parte de la clave. En el segundo caso, se ordena por el campo tres de cada línea; *sólo si hay empate*, se usa también el campo cuatro para ordenar; y, *sólo si hay empate de nuevo*, se usa el campo cinco. Por lo tanto, hay tres claves de ordenación, y no sólo una, y los caracteres de separación no se tienen en cuenta al ordenar. A continuación tiene un ejemplo que muestra las diferencias.

Para comprobar que no es lo mismo usar varios campos conjuntamente que usarlos individualmente, supongamos el siguiente fichero, donde los campos vienen separados por el carácter «:»:

```
$ cat pruebaopcionk
ll:jj:cc:b:ee
mm:xx:cc:a:gg
mm:xx:cc:aa:pp
```

La ordenación utilizando los campos 3, 4 y 5 conjuntamente producirá la siguiente ordenación:

```
$ sort -t: -k3,5 pruebaopcionk
mm:xx:cc:aa:pp
mm:xx:cc:a:gg
ll:jj:cc:b:ee
```

ya que la cadena «cc:aa:pp» va lexicográficamente antes que la cadena «cc:a:gg» y ésta antes que «cc:b:ee». Sin embargo, si se usan los campos individualmente, el resultado será este otro:

```
$ sort -t: -k3,3 -k4,4 -k5,5 pruebaopcionk
mm:xx:cc:a:gg
mm:xx:cc:aa:pp
ll:jj:cc:b:ee
```

porque, al haber empate en el tercer campo (cadena «cc»), se consulta el cuarto campo, y aquí no hay duda de que la cadena «a» va antes que «aa» y ésta antes que «b». Al no haber empate en el cuarto campo, no es necesario consultar el quinto.

4.3. Ejercicios

A lo largo de los siguientes ejercicios se utilizarán algunos ficheros de texto cuyo contenido, a modo de ejemplo, podría ser el que se muestra a continuación. Puedes crear tú mismo estos ficheros con este contenido o con el que estimes oportuno.

```
$ cat empleados
pepe lopez 123 calle mayor
juan gutierrez 22 calle quevedo
pepe lopez 23 calle quijote
juan gutierrez 22 calle lope
juan gutierrez 3 calle cervantes

$ cat empleados2puntos
pepe:lopez:123:calle:mayor
juan:gutierrez:22:calle:quevedo
pepe:lopez:23:calle:quijote
juan:gutierrez:22:calle:lope
juan:gutierrez:3:calle:cervantes
```

1. Ordena el fichero `empleados` por el primer campo, si este coincide entonces por el tercer campo y, finalmente, si este también coincide, por el quinto campo.
2. Ordena el fichero `empleados` por el primer campo usando orden alfabético, si este coincide entonces por el tercer campo usando orden numérico.
3. Ordena el fichero `empleados2puntos` por el tercer campo usando orden alfabético, teniendo en cuenta que la separación entre campos está marcada por el carácter «:».

5. Orden `tr`

La sintaxis de la orden `tr` es:

```
tr [opción]... conjunto1 [conjunto2]
```

Esta orden copia el texto desde su entrada estándar, reemplazando los caracteres indicados en `conjunto1` por los caracteres correspondientes de `conjunto2`, reemplazando múltiples ocurrencias de caracteres de `conjunto1` por un único carácter, o eliminando los caracteres de `conjunto1`, dependiendo de las opciones usadas en casa caso. El resultado final lo muestra en su salida estándar. Por ejemplo:

```
$ tr abc xyz < infile > outfile
```

sustituye el carácter «a» por «x», el «b» por «y» y «c» por «z» en el fichero `infile`, guardando los resultados en `outfile`. No requiere que los caracteres «abc» aparezcan juntos como un patrón en `infile`. Esto se ve claramente en el siguiente ejemplo:

```
$ echo La ballena estaba cansada | tr abc xyz
Lx yxllenx estxyx zxnsxdx
```

La orden `echo` simplemente escribe a su salida estándar los parámetros que recibe y, gracias a la tubería, dichos parámetros llegan a la orden `tr` que hace las sustituciones indicadas.

Las opciones de uso más frecuentes son:

- `-s`: Elimina repeticiones contiguas de los caracteres especificados en `conjunto1`.
- `-d`: Elimina los caracteres especificados en `conjunto1`.
- `-c`: Utiliza el conjunto complementario de caracteres de `conjunto1`, es decir, todos los caracteres excepto los indicados en `conjunto1`.

Para indicar `conjunto1` y `conjunto2` se pueden utilizar también ciertas especificaciones especiales:

- `x-y`: Para especificar un rango de caracteres desde el carácter «x» hasta el «y».
- `\c`: Para especificar una secuencia de escape, donde «c» puede ser el carácter «n» para indicar un salto de línea (`\n`), el carácter «t» para indicar un tabulador (`\t`), ...
- `[:class:]`: Para especificar una clase de caracteres: los caracteres alfanuméricos (`:alnum:`), los dígitos (`:digit:`), los caracteres de espaciado horizontal (`:blank:`), ...
- `[c*n]`: Para indicar `n` repeticiones del carácter `c`.

Recuerda que algunos de los caracteres usados en estas especificaciones, como «\», «[» y «*», son especiales para Bash, por lo que tendrás que encerrarlos entre comillas simples para evitar que Bash los interprete y así puedan llegar a la orden `tr`.

5.1. Ejercicios

A lo largo de los siguientes ejercicios se utilizarán algunos ficheros de texto cuyo contenido, a modo de ejemplo, podría ser el que se muestra a continuación. Puedes crear tú mismo estos ficheros con este contenido o con el que estimes oportuno.

```
$ cat empleadostabulados
pepe lopez     123 calle mayor
juan gutierrez 22  calle quevedo
pepe lopez     23  calle quijote
juan gutierrez 22  calle lope
juan gutierrez 3   calle cervantes
```

1. Muestra en pantalla el contenido del fichero `empleadostabulados` comprimiendo todos los espacios en blanco consecutivos en uno único.

   ```
   $ cat empleadostabulados | tr -s " "
   ```

2. Muestra en pantalla el contenido del fichero `empleadostabulados` cambiando cualquier conjunto de espacios en blanco consecutivos en un único tabulador.

3. Muestra en pantalla el contenido del fichero `empleadostabulados` cambiando todas las letras minúsculas por mayúsculas.

   ```
   $ cat empleadostabulados | tr "[a-z]" "[A-Z]"
   ```

   o bien así:

```
$ cat empleadostabulados | tr "[:lower:]" "[:upper:]"
```

4. Muestra en pantalla el contenido del fichero `empleadostabulados` cambiando todos los caracteres que no pertenezcan a la clase «espacios en blanco» por la letra «x».

5. Completa la tubería

   ```
   echo Secreto a voces es un oxímoron | tr ...
   ```

   para que cada palabra aparezca en una línea distinta.

6. Orden `cut`

La orden `cut` sirve para seleccionar columnas de un fichero de texto y mostrarlas en la salida estándar. Si no se especifica ningún fichero, toma su entrada estándar. Las columnas se pueden especificar en bytes, caracteres o campos, indicando qué delimitadores concretos queremos manejar para diferenciarlas. Su sintaxis es:

```
cut opción... [fichero]...
```

Las opciones más comunes son:

- `-c`: las columnas son especificadas por posiciones de caracteres.
- `-f`: las columnas son especificadas por campos. Por defecto, el delimitador es el tabulador.
- `-d`: usada junto a la opción `-f` para indicar un delimitador específico distinto del tabulador para los campos.
- `-s`: usada junto a la opción `-f` para indicar que, si una línea no contiene ningún delimitador, no debe mostrarse en la salida.

A la hora de indicar los caracteres o campos a mostrar de cada línea, se pueden indicar varios caracteres o campos, separados por comas, por ejemplo, `-c1,2,3` o `-f3,5`. También se pueden indicar rangos, por ejemplo, `-c1-8` o `-f1-3,5`.

Hay dos consideraciones importantes respecto a la orden `cut`. La primera es que, como hemos dicho, el caracter separador de campos por defecto es el tabulador. Sin embargo, es raro encontrar dicho caracter separador en ficheros y en salidas de otras órdenes. Por eso, es importante no olvidar el usar la opción `-d` cuando se use la opción `-f`. La otra consideración es que `cut` no considera un mismo carácter separador repetido como uno sólo, sino que considera cada ocurrencia individualmente. Observa el resultado de la siguiente orden (hemos usado comillas simples para evitar que Bash interprete los espacios en blanco):

```
$ echo 'campo1    campo2' | cut -f2 -d' '
```

Como vemos, no se la seleccionado el segundo campo, como cabría esperar, sino el segundo espacio en blanco tras «campo1». Si queremos obtener el resultado esperado, debemos usar la orden `tr` para eliminar repeticiones del carácter separador. Observa el resultado de las siguientes órdenes:

```
$ echo 'campo1    campo2' | tr -s ' '
campo1 campo2
$ echo 'campo1    campo2' | tr -s ' ' | cut -f2 -d' '
campo2
```

Por eso, es importante no olvidar el uso de la orden `tr` antes de una orden `cut` cuando se usa la opción `-f` de ésta última.

6.1. Ejercicios

A lo largo de los siguientes ejercicios se utilizarán algunos ficheros de texto cuyo contenido, a modo de ejemplo, podría ser el que se muestra a continuación. Puedes crear tú mismo estos ficheros con este contenido o con el que estimes oportuno.

```
$ cat dataset1
Pine 906 26 1.0 211
Beech 933 26 2.3 160
Fur 1246 27 2.44 162
Palm 671 25 3.8 888

$ cat dataset3
Trees of the Forest
Pine,906,26,020079,130.0,80.3,17.1,211
Beech,933,26,030079,48.0,85.2,22.7,160
Fur,1246,27,070079,31.0,86.5,6.9,162
Palm,671,25,100077,41.0,87.3,15.0,888
```

1. Corta los caracteres del 13 al 17 del fichero `dataset1`.

   ```
   $ cut -c 13-17 dataset1
   ```

2. Corta los caracteres del 4, 5 y del 11 al 15 del fichero `dataset1`.

3. Corta los campos 1, 3, 4, 6 y 8 del fichero `dataset3`, teniendo como delimitador de campos el carácter «,». En la solución final no debe aparacer la línea «Trees of the Forest» al no contar con ningún carácter delimitador.

   ```
   $ cut -d',' -f 1,3,4,6,8 -s dataset3
   ```

4. Lista los tamaños de todos los ficheros del directorio actual, junto a su nombre.

   ```
   $ ls -l
   # primero lista los ficheros

   total 8
   -rw-r--r-- 1 javiercm profesor  13 sep 11 13:56 f1
   -rw-r--r-- 1 javiercm profesor 160 sep 11 13:57 f2
   -rw-r--r-- 1 javiercm profesor   0 sep 11 13:56 f3
   -rw-r--r-- 1 javiercm profesor   0 sep 11 13:55 fic1
   -rw-r--r-- 1 javiercm profesor   0 sep 11 13:55 fic2
   -rw-r--r-- 1 javiercm profesor   0 sep 11 13:55 fic3

   $ ls -l | tr -s " "
   #con "tr -s" se comprimen espacios en blanco consecutivos

   total 8
   -rw-r--r-- 1 javiercm profesor 13 sep 11 13:56 f1
   -rw-r--r-- 1 javiercm profesor 160 sep 11 13:57 f2
   -rw-r--r-- 1 javiercm profesor 0 sep 11 13:56 f3
   -rw-r--r-- 1 javiercm profesor 0 sep 11 13:55 fic1
   -rw-r--r-- 1 javiercm profesor 0 sep 11 13:55 fic2
   ```

```
-rw-r--r-- 1 javiercm profesor 0 sep 11 13:55 fic3


$ ls -l | tr -s " " | cut -d " " -f 5,9
#con "cut" selecciona los campos delimitados por 1 unico espacio en blanco

13 f1
160 f2
0 f3
0 fic1
0 fic2
0 fic3
```

5. Lista los 8 primeros caracteres de los nombres de todos los ficheros del directorio actual.
6. Lista la fecha de modificación de todos los ficheros del directorio actual, junto a su nombre.

7. Orden `uniq`

La orden `uniq` sirve para eliminar líneas duplicadas de un fichero, siempre que éstas aparezcan juntas. Por eso, para poder llevar a cabo su labor, el fichero se suele ordenar previamente. Su sintaxis es:

```
uniq [opción]... [fichero_entrada [fichero_salida]]
```

Si no se indica fichero de entrada o de salida, se utilizará la entrada estándar o la salida estándar, respectivamente. Observa que, según la sintaxis, no es posible indicar un fichero de salida si no se indica un fichero de entrada.

Las opciones más importantes de `uniq` son:

- `-c`: escribe a la salida el número de veces que aparece cada línea en el fichero de entrada justo antes de escribir la propia línea.
- `-d`: cada línea duplicada es escrita a la salida solamente una vez. No escribe ninguna línea que no estuviera duplicada en la entrada.
- `-D`: igual que `-d`, pero sin sustituir las líneas duplicadas por una sola.
- `-u`: escribe a la salida únicamente las líneas no duplicadas en la entrada.
- `-i`: no distingue entre mayúsculas y minúsculas a la hora de comparar.
- `-s N`: no tiene en cuenta los `N` primeros caracteres de cada línea a la hora de comparar.
- `-w N`: solamente tiene en cuenta los `N` primeros caracteres de cada línea a la hora de comparar.

Por ejemplo, si partimos del fichero `errlog` con todos los errores que se han ido produciendo en un ordenador durante la última sesión:

```
$ cat errlog
error 11: /tmp directory not found
error 22: out of memory
error 11: /tmp directory not found
error 17: low disk space
error 11: /tmp directory not found
error 22: out of memory
error 04: connection failure
error 11: /tmp directory not found
```

Podemos obtener, con la ayuda de la orden `sort`, un listado con los tipos de errores de esta manera:

```
$ sort errlog > serrlog

$ cat serrlog
error 04: connection failure
error 11: /tmp directory not found
error 11: /tmp directory not found
error 11: /tmp directory not found
error 11: /tmp directory not found
error 17: low disk space
error 22: out of memory
error 22: out of memory

$ uniq serrlog > uerrlog

$ cat uerrlog
error 04: connection failure
error 11: /tmp directory not found
error 17: low disk space
error 22: out of memory
```

Evidentemente, también podríamos haber escrito simplemente lo siguiente para obtener el mismo resultado, en el caso de no necesitar los ficheros intermedios:

```
$ sort errlog | uniq
```

Combinando esta orden con las ordenes `cut` y `sort` podemos obtener la contabilidad resumida de los códigos de error producidos:

```
$ cut -f1 -d':' serrlog
error 04
error 11
error 11
error 11
error 11
error 17
error 22
error 22
# Con "cut" cortamos el primer campo,
# tomando como delimitador entre campos el caracter ":"


$ cut -f1 -d':' serrlog | sort | uniq -c
1 error 04
4 error 11
1 error 17
2 error 22
# En este caso, el "sort" no habria sido necesario
# porque las filas iguales están ya consecutivas
```

7.1. Ejercicios

Partimos del fichero `purchases` que contiene una lista de compras realizadas. Por cada línea contiene el nombre de un cliente, la fecha de la compra y el código del artículo comprado («Unit 12», «Unit 05», ...).

```
$ cat purchases
Jimmy James Jan 2 Unit 12
Jane Doe Jan 4 Unit 17
Jimmy James Jan 10 Unit 12
John Huber Jan 15 Unit 17
Sue Butler Jan 2 Unit 05
Jane Doe Jan 10 Unit 12
Liz Tyler Feb 2 Unit 04
Jimmy James Feb 4 Unit 03
```

1. Genera una lista de cuántas unidades se han vendido de cada artículo.

```
$ cut -d' ' -f5,6 purchases
    Unit 12
    Unit 17
    Unit 12
    Unit 17
    Unit 05
    Unit 12
    Unit 04
    Unit 03

$ cut -d' ' -f5,6 purchases | sort
    Unit 03
    Unit 04
    Unit 05
    Unit 12
    Unit 12
    Unit 12
    Unit 17
    Unit 17

$ cut -d' ' -f5,6 purchases | sort | uniq -c
    1 Unit 03
    1 Unit 04
    1 Unit 05
    3 Unit 12
    2 Unit 17
```

2. Genera una lista de clientes, guardándola en el fichero «customers».

```
$ cut -d' ' -f1,2 purchases | sort | uniq > customers
# con "cut" cortamos el campo 1º y 2º del fichero "purchases"
# usando el " " como delimitador entre campos
# con "sort" ordenamos el resultado
# con "uniq" eliminamos líneas repetidas
# finalmente la salida se escribe en el fichero "customers"

$ cat customer
    Jane Doe
    Jimmy James
    John Huber
    Liz Tyler
```

```
Sue Butler
```

3. Genera una lista de clientes que han realizado más de una compra.
4. Genera una lista del número de artículos vendidos por cada día.
5. Genera una lista de aquellos artículos de los que se han vendido más de 1 unidad.

8. Miscelánea

A continuación mostramos de forma resumida algunas oras órdenes que pueden ser útiles como filtros:

- `head`: muestra las primeras líneas de la información que le llega.
- `tail`: muestra las últimas líneas de la información que le llega.
- `column`: muestra, organizada por columnas, la información que le llega.
- `wc`: muestra un recuento del número de líneas, palabras y bytes de la información que le llega.
- `tee`: envía la información que le llega por su entrada estandar tanto a su salida estandar como a los ficheros indicados como parámetros.

Todas estas órdenes disponen de diversas opciones que modifican su comportamiento. Consulta sus páginas de manual para verlas.

Veamos algunos ejemplos de uso de estas órdenes:

- Lista los nombres de las 10 primeras entradas del directorio actual:

  ```
  ls | head -10
  ```

- Cuenta el número de subdirectorios que hay a partir del directorio actual:

  ```
  find . -type d | wc -l
  ```

  Es decir, primero listamos todos los directorios y luego contamos el número de líneas de ese listado

- Muestra en pantalla y también guardar en el fichero `listado` los nombres de todos los ficheros terminados en `.odt` que hay a partir del directorio actual:

  ```
  find . -name "*.odt" | tee listado
  ```

- Lista el nombre completo y el tamaño de cada ficheros acabado en `.sh` que hay a partir del directorio actual. El listado se muestra alineando las diferentes columnas de datos:

  ```
  find . -name "*.sh" -printf "Fichero: %p Tamaño: %s Bytes\n" | column -t
  ```

9. Bibliografía

- Páginas de manual de las diferentes órdenes descritas y del intérprete de órdenes Bash.
- *Shell & Utilities: Detailed Toc.* The Open Group Base Specifications. http://pubs.opengroup.org/onlinepubs/9699919799/utilities/contents.html.
- *Linux: Domine la administración del sistema*, 2ª edición. Sébastien Rohaut. ISBN 9782746073425. Eni, 2012.

10. Ejercicios propuestos

Los siguientes ejercicios deben resolverse mediante una única línea de órdenes separadas por tuberías.

1. A partir de la salida de la orden `ps aux`, que muestra información de todos los procesos existentes en el sistema en el momento de su ejecución, selecciona todos aquellos procesos que pertenecen al usuario `alumno` (columna USER).

2. Modifica la solución anterior para mostrar cuántos procesos hay de `alumno`, pero sin mostrar la información de dichos procesos.

3. Modifica la solución anterior para mostrar cuántos procesos hay de todos los usuarios, menos del usuario `alumno`.

4. Usando las órdenes `find` y `grep`, busca todos los ficheros regulares que hay a partir de `/etc` que contienen la palabra «alumno». Sólo hay que mostrar los nombres de los ficheros y la búsqueda se debe realizar sin distinguir entre mayúsculas y minúsculas.

5. Muestra la línea del fichero `/etc/passwd` que pertenece al usuario `alumno`.

6. Obtén todas las palabras (deben contener sólo letras, sin distinguir entre mayúsculas y minúsculas) que aparecen en la salida de la orden `ps aux`. Muestra cada palabra en una línea distinta. Usa la orden `tail -n +2` en la posición adecuada de la tubería para eliminar la cabecera de la orden `ps aux` (es decir, la primera línea).

7. Modifica la solución del ejercicio anterior para quedarse sólo con las palabras de 4 letras. La lista debe aparecer ordenada alfabéticamente en orden creciente.

8. Modifica la solución del ejercicio anterior para que, además, las palabras se muestren en columnas.

9. Obtén todos los números enteros positivos, incluyendo el 0, que aparecen en la salida de la orden `ps aux`. Muestra cada número en una línea distinta.

10. Modifica la solución del ejercicio anterior para que los números aparezcan ordenados de mayor a menor.

11. Ordena la salida de la orden `ps aux` por consumo de memoria RAM de cada proceso (columna RSS; el número que aparece es el consumo en KiB).

12. Obtén la columna RSS de la orden `ps aux` (sólo debe aparecer por pantalla el contenido de esta columna).

13. Usando la orden `find`, obtén una lista de todos los ficheros regulares que se encuentran en `/etc` y sus subdirectorios ordenada por tamaño de los ficheros. La salida obtenida debe seguir el siguiente formato:

    ```
    fichero:tamaño
    ```

14. A partir de la salida de la orden `ps aux`, obtén una lista de todos los usuarios propietarios de los procesos existentes.

15. Modifica la solución anterior para que sólo aparezcan los usuarios que son propietarios de más de un proceso.

16. Modifica la solución anterior para que los nombres de los usuarios aparezcan en mayúsculas.

17. Modifica la solución anterior para mostrar el total de usuarios, pero no los nombres de los mismos.