

Uso de comandos básicos en Linux y su aplicación en biocómputo

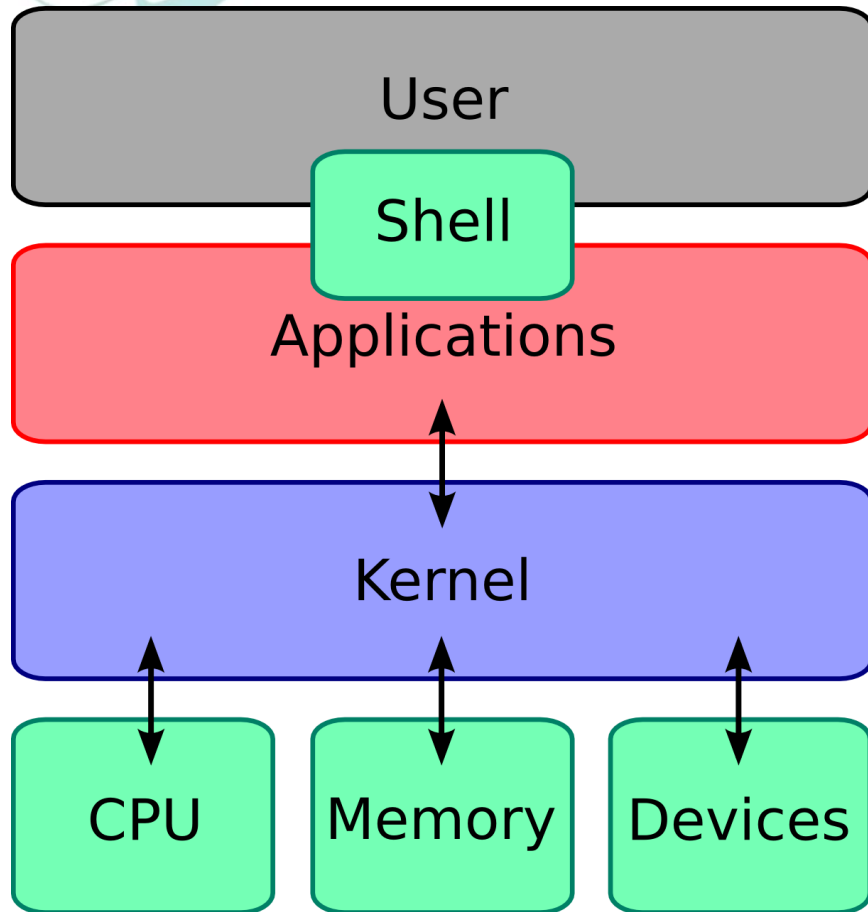
Victor Flores

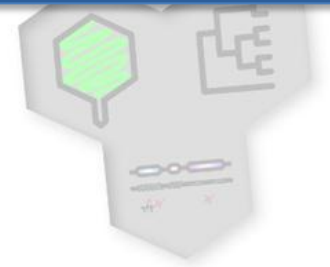
genos
médica

Centro especializado en Genética

Febrero 2017

Kernel, shell, PATH, aplicaciones?



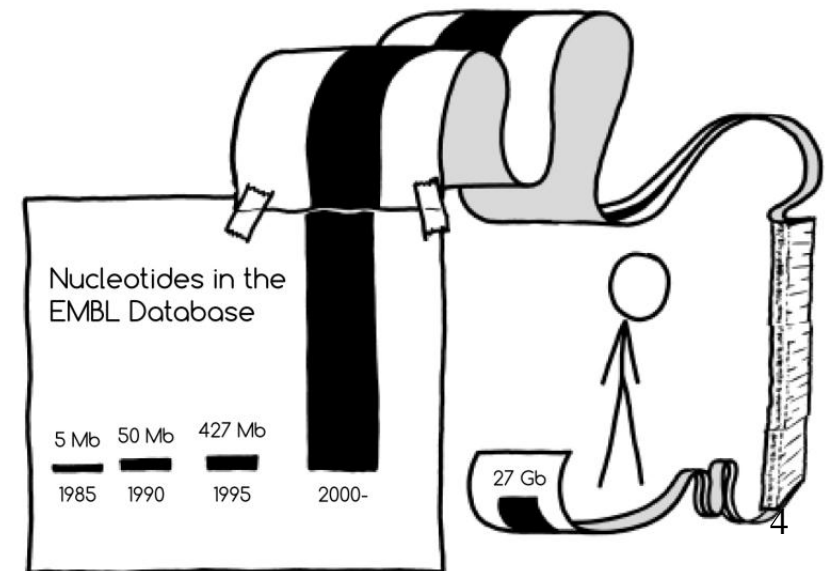


Interfaces de usuario

- El usuario interactúa con la computadora a través de periféricos, los cuales envían una señal al shell.
- Normalmente lo hacemos a través de una interfaz gráfica
 - Usamos un puntero, un teclado, una pantalla táctil y otros botones para decirle a nuestra computadora qué hacer.
- Existen otras formas de acceder al shell, quizá no tan sencillas pero con mas control: la línea de comandos

Cuando manejamos datos biológicos, usualmente requerimos obtener la información de un elemento o de un conjunto de ellos.

- Si manejamos tablas pequeñas. esto puede hacerse manualmente
- Si nuestras listas y tablas contienen cientos o miles de elementos no puede hacerse manualmente
- Provee de un ambiente controlado
- Nos indicará de errores que ocurran durante la ejecución de un programa



Command line

The command line



What I think
it was



What I think
it is

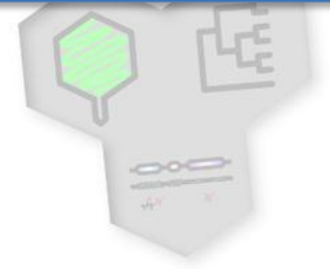


```
File Edit View Search Terminal Help
[vflorele]@[-]on[Mon Feb 17]@[19:24:00]$ 
echo Hello world!\n
Hello world!n
[vflorele]@[-]on[Mon Feb 17]@[19:24:10]$ 
echo Hello world!; echo
Hello world!
[vflorele]@[-]on[Mon Feb 17]@[19:24:24]$
```

What it really
is



Cómo acceder la terminal?



- Windows
 - Menú inicio -> run -> cmd
- Linux
 - GNOME
 - Accessories -> system tools -> Terminal
 - KDE
 - Programs -> System -> Konsole
- Mac OSX
 - Finder -> Applications -> Utilities -> Terminal

- Es sensible a mayúsculas/minúsculas

En nuestro sistema windows en un mismo directorio no pueden existir los siguientes archivos:

“A.pdf” y “a.pdf”

En nuestro sistema Linux podemos tener en un mismo directorio

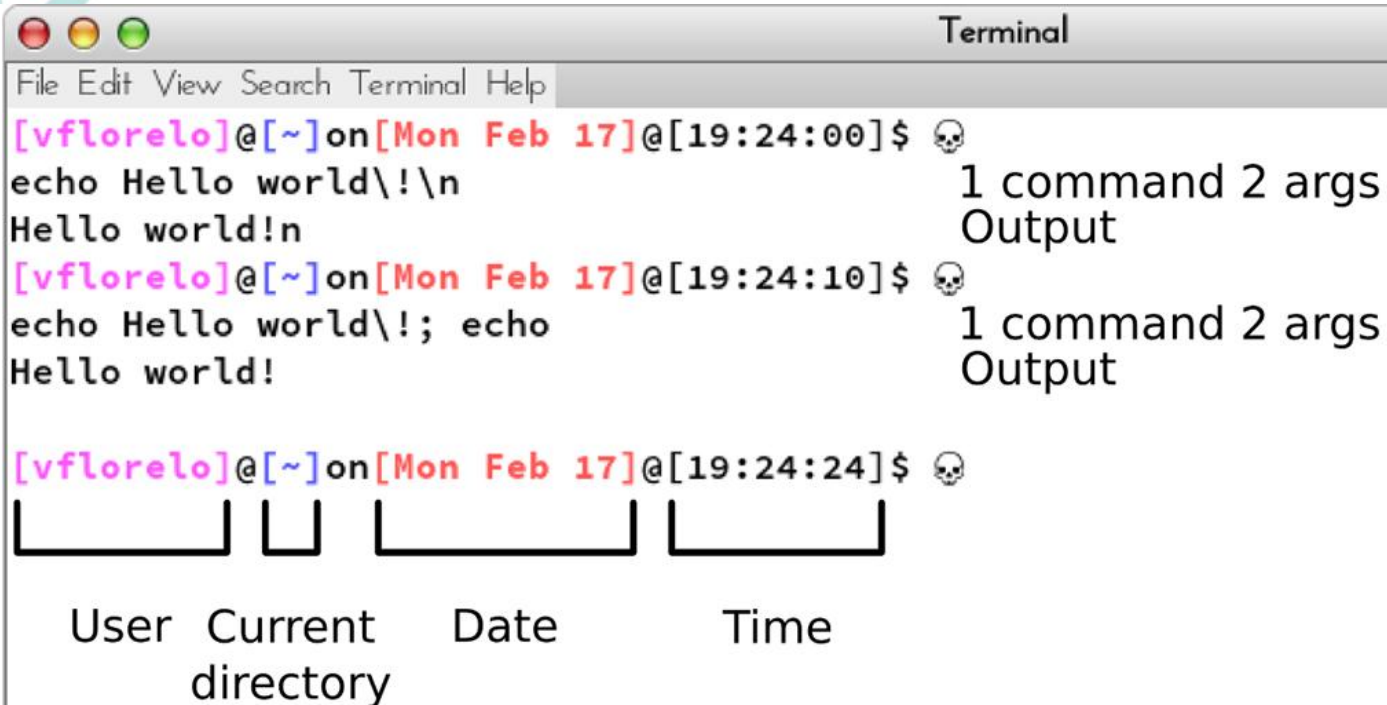
“abcD.pdf” “Abcd.pdf” “aBcd.pdf” “abCd.pdf”
“ABCd.pdf”

- Muchos caracteres estan reservados para el sistema o tienen un significado especial:

! @ # \$ % ^ & * - + = ? \ | /

() { } [] < > ; : ' " , .

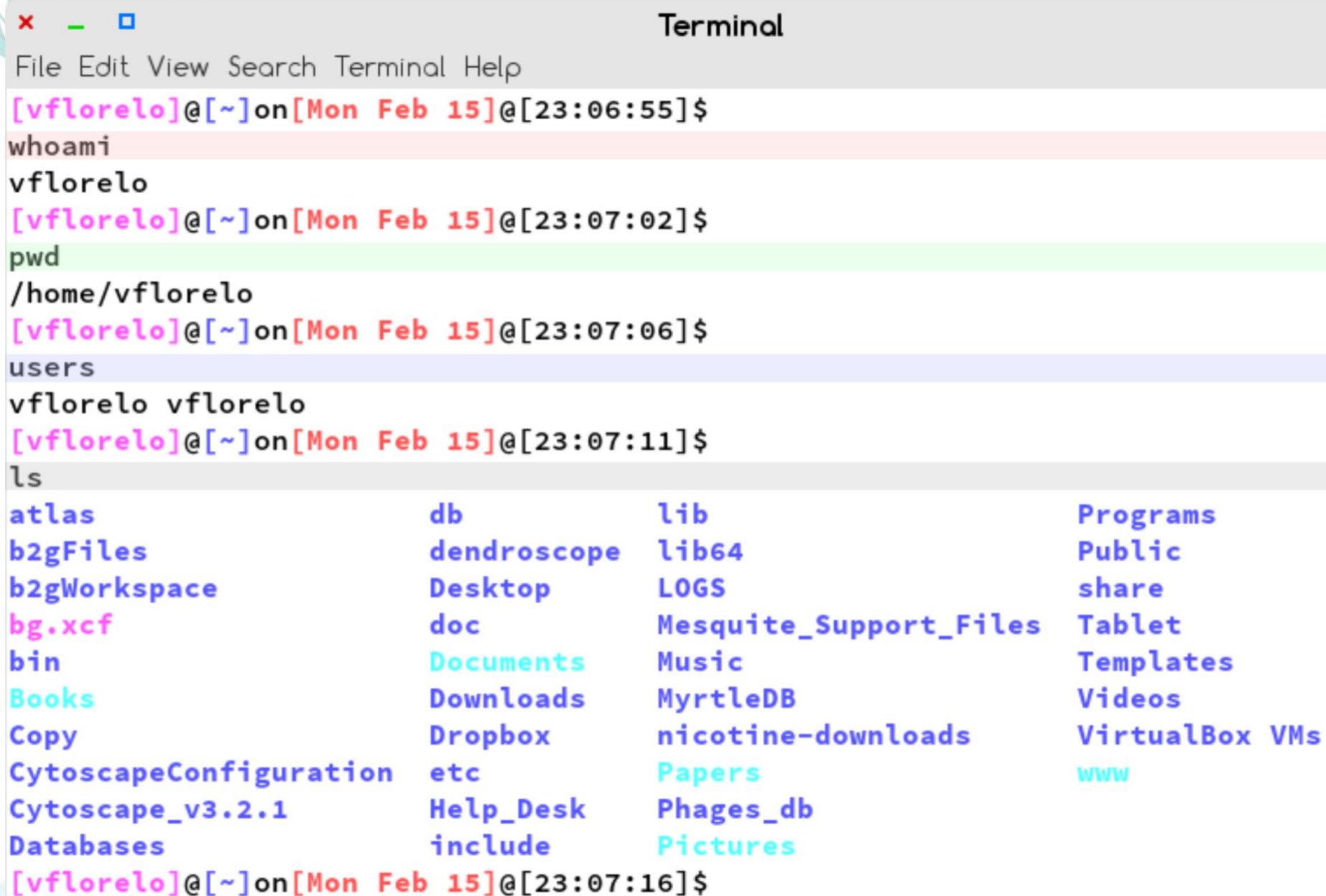
La terminal



```
File Edit View Search Terminal Help
[vflorello]@[~]on[Mon Feb 17]@[19:24:00]$ 
echo Hello world!\n
Hello world!\n
[vflorello]@[~]on[Mon Feb 17]@[19:24:10]$ 
echo Hello world\!; echo
Hello world!
[vflorello]@[~]on[Mon Feb 17]@[19:24:24]$
```

User Current directory Date Time

Estructura básica: un sólo comando



A terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help). The window shows a series of commands and their outputs, each on a new line. The prompt is [vflorelo]@[~]on[Mon Feb 15]@[23:06:55]\$. The commands and outputs are: whoami (vflorelo), pwd (/home/vflorelo), users (vflorelo vflorelo), and ls (a list of files and directories). The ls output is formatted in a grid-like structure with four columns. The prompt at the end is [vflorelo]@[~]on[Mon Feb 15]@[23:07:16]\$.

```
Terminal
File Edit View Search Terminal Help
[vflorelo]@[~]on[Mon Feb 15]@[23:06:55]$
whoami
vflorelo
[vflorelo]@[~]on[Mon Feb 15]@[23:07:02]$
pwd
/home/vflorelo
[vflorelo]@[~]on[Mon Feb 15]@[23:07:06]$
users
vflorelo vflorelo
[vflorelo]@[~]on[Mon Feb 15]@[23:07:11]$
ls
atlas          db          lib          Programs
b2gFiles       dendroscope lib64         Public
b2gWorkspace  Desktop    LOGS         share
bg.xcf         doc        Mesquite_Support_Files Tablet
bin           Documents  Music        Templates
Books         Downloads  MyrtleDB     Videos
Copy          Dropbox    nicotine-downloads VirtualBox VMs
CytoscapeConfiguration etc        Papers       www
Cytoscape_v3.2.1 Help_Desk  Phages_db
Databases     include   Pictures
[vflorelo]@[~]on[Mon Feb 15]@[23:07:16]$
```

Estructura básica: 1 comando, 1 argumento

Terminal

File Edit View Search Terminal Help

[vflorelo]@[~]on[Mon Feb 15]@[23:13:27]\$

touch test_file

[vflorelo]@[~]on[Mon Feb 15]@[23:13:42]\$

cat test_file

Esto es un archivo de texto

[vflorelo]@[~]on[Mon Feb 15]@[23:14:10]\$

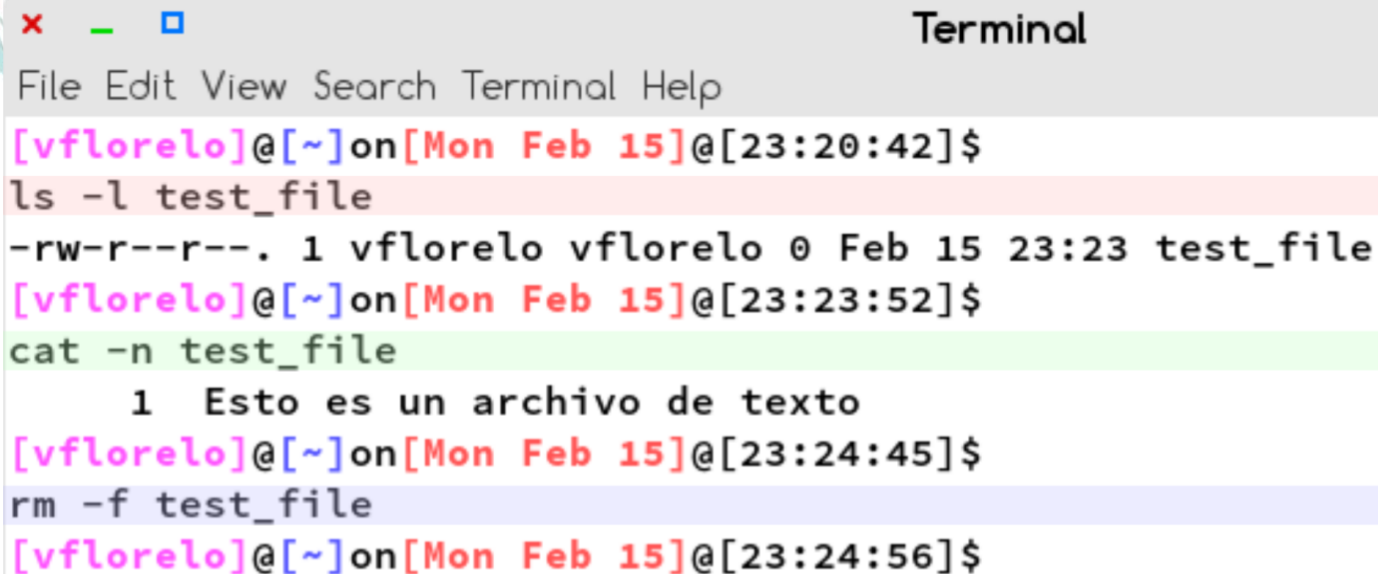
rm test_file

[vflorelo]@[~]on[Mon Feb 15]@[23:14:24]\$

Estructura básica: 1 comando, 1 opción

```
Terminal
File Edit View Search Terminal Help
[vflorelo]@[~]on[Mon Feb 15]@[23:20:31]$
uname -o
GNU/Linux
[vflorelo]@[~]on[Mon Feb 15]@[23:20:40]$
df -h
Filesystem      Size  Used Avail Use% Mounted on
devtmpfs        1.9G   0    1.9G   0% /dev
tmpfs           1.9G  1.3M   1.9G   1% /dev/shm
tmpfs           1.9G  972K   1.9G   1% /run
tmpfs           1.9G   0    1.9G   0% /sys/fs/cgroup
/dev/sda3       56G   18G   37G   33% /
tmpfs           1.9G  156K   1.9G   1% /tmp
/dev/sda1       477M  153M  295M   35% /boot
/dev/sda2      856G  322G  491G   40% /home
tmpfs           376M   28K  376M   1% /run/user/1000
[vflorelo]@[~]on[Mon Feb 15]@[23:20:42]$
```

Estructura básica: 1 comando, 1 opción, 1 argumento

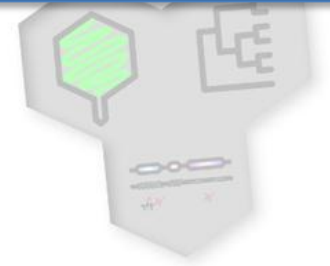


A terminal window titled "Terminal" with a menu bar (File, Edit, View, Search, Terminal, Help). The prompt is [vflorelo]@[~]on[Mon Feb 15]@[23:20:42]\$. The user enters 'ls -l test_file', showing file details for 'test_file'. Then they enter 'cat -n test_file', showing line 1: 'Esto es un archivo de texto'. Finally, they enter 'rm -f test_file' to delete the file. The prompt is [vflorelo]@[~]on[Mon Feb 15]@[23:24:56]\$. The terminal has a light gray background with colored highlights for each command and its output.

```
[vflorelo]@[~]on[Mon Feb 15]@[23:20:42]$  
ls -l test_file  
-rw-r--r--. 1 vflorelo vflorelo 0 Feb 15 23:23 test_file  
[vflorelo]@[~]on[Mon Feb 15]@[23:23:52]$  
cat -n test_file  
1 Esto es un archivo de texto  
[vflorelo]@[~]on[Mon Feb 15]@[23:24:45]$  
rm -f test_file  
[vflorelo]@[~]on[Mon Feb 15]@[23:24:56]$
```



Para no perderse



whoami

Who am I? Nos indica con que nombre de usuario estamos trabajando

pwd

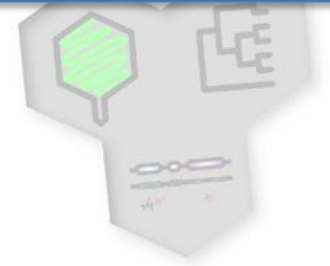
Print working directory, Nos indica en que directorio estamos trabajando

ls

List, nos indica que elementos hay en el directorio actual



Para trabajar archivos y directorios



cd

Change **D**irectory, nos cambia al directorio que le indiquemos

mkdir

Make **D**irectory, crea un directorio con el nombre que le indiquemos

cp

Copy, copia un archivo a un directorio

mv

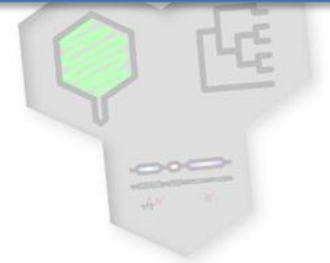
Move, mueve un archivo a un directorio

rm

Remove, elimina el archivo indicado

rmdir

Remove **D**irectory, elimina el directorio indicado



Para examinar archivos

cat

Concatenate, nos da la totalidad de uno o más archivos. También sirve para unir múltiples archivos

head

Head, nos da las primeras N líneas de un archivo

tail

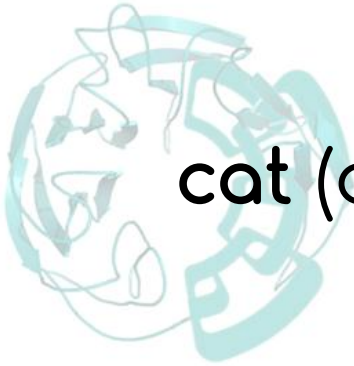
Tail, nos da las últimas N líneas de un archivo

paste

Paste nos une horizontalmente 2 o más archivos, útil para crear tablas

wc

Word count, nos da el contenido de un archivo (o string), líneas, palabras, caracteres



cat (concatenate) → nos sirve para mostrar el contenido de uno o varios archivos

```
$ cat archivo_1
```

Esto es un archivo de texto

```
$ cat archivo_2
```

Esto también

```
$ cat archivo_1 archivo_2
```

Esto es un archivo de texto

Esto también

```
$ cat archivo_1 archivo_2... ..archivo_n
```

```
$ cat archivo_*
```

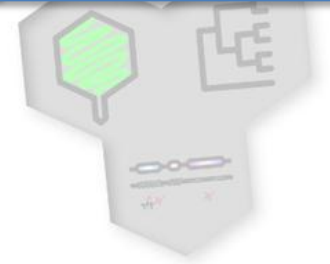
grep (Globally search Regular Expression and Print) nos muestra líneas de un archivo donde ocurra una expresión regular.

```
$ cat lista_super
Cereal      $34.50      400g
Leche $15.50      1L
Huevos      $24.00      1Kg
Carne $95.00      1Kg
```

```
$ grep Leche lista_super
Leche $15.50      1L
```

```
$ grep C lista_super
Cereal      $34.50      400g
Carne $95.00      1Kg
```

```
$ grep Harina lista_super
```



-n

Muestra el número de líneas

-C

Cuenta el número de ocurrencias

-w

Limita la búsqueda a la palabra completa

-m

Entrega un máximo de m líneas por término de búsqueda

-i

Insensible a mayúsculas/minúsculas

-v

Solo muestra las líneas donde no ocurre nuestro término de búsqueda

cut nos sirve para extraer columnas de un archivo tipo tabla (la versión vertical de grep)

Requiere que el archivo tenga cierto formato para que podamos definir el delimitador de campo (-d) y el número de campo a extraer (-f)

```
$ cat id_list
>gi|589892942|ref|YP_009004271.1| [Clavibacter phage CN1A]
>gi|589891419|ref|YP_009007658.1| [Acinetobacter phage Presley]
>gi|548650442|gb|AGR47742.2| [Sinorhizobium phage phiM12]
```

^^ ^^^^^^^^^^^ ^^ ^^^^^^^^^^^^^ ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

C1 C2 C3 C4 C5

En el ejemplo, el caracter "|" nos sirve como delimitador de campo

```
$ cut -d\| -f2 id_list
589892942
589891419
548650442
```

sort nos sirve para ordenar una serie de datos
ya sea numéricamente o alfanuméricamente

```
$ cat lista_super
Cereal      $34.50      400g
Leche       $15.50      1L
Huevos      $24.00      1Kg
Carne       $110.00     1Kg
```

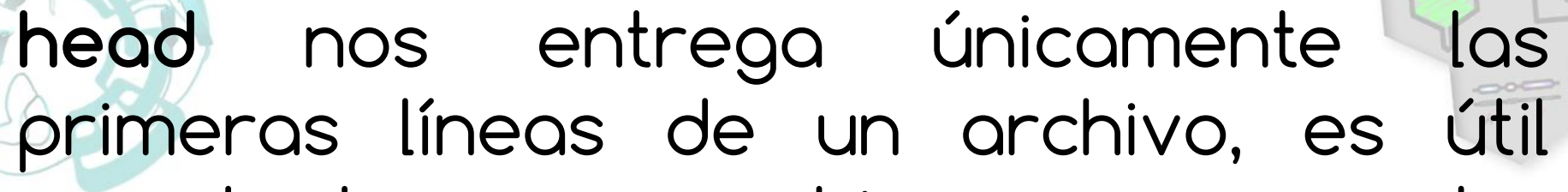
```
$ sort lista_super
Carne       $110.00     1Kg
Cereal      $34.50      400g
Huevos      $24.00      1Kg
Leche       $15.50      1L
```

```
$ cat lista_ocurrencias
23
120
3
98
```

```
$ sort lista_ocurrencias
120
23
3
98
```

```
$ cat lista_ocurrencias
23
120
3
98
```

```
$ sort -n lista_ocurrencias
3
23
98
120
```



head nos entrega únicamente las primeras líneas de un archivo, es útil cuando tenemos archivos muy grandes (como los de secuenciación masiva) y únicamente queremos ver qué es lo que hay en ese archivo, sin verlo todo

tail es la contraparte de head, únicamente entrega las últimas líneas del archivo, es útil cuando transferimos archivos y queremos verificar si estos están completos

- Una de las utilidades de usar línea de comandos, es que podemos conectar varios comandos en una sola línea y obtener un resultado en menos pasos.
 - Es decir, alimentar un segundo comando con la salida del primer comando

Solución 1:

```
$ sort lista_super > lista_ordenada
$ grep -n Carne lista_ordenada > linea_carne
$ cat linea_carne
1:Carne    $95.00    1Kg
```

Solución 2:

```
$ sort lista_super | grep -n Carne > linea_carne
$ cat linea_carne
1:Carne    $95.00    1Kg
```

sed (**s**tream **e**ditor) nos sirve para manejo general de datos, en particular para sustituir palabras o términos

```
$ sed 's/Carne/Frijoles/' lista_super  
Cereal  
Leche  
Huevos  
Frijoles
```

```
$ sed 's/Escherichia/E./' org_list  
E. coli  
Rhodopseudomonas palustris  
Salmonella typhi  
Pseudomonas aeruginosa
```

```
$ sed 's/Pseudomonas/P./i' org_list  
E. coli  
RhodoP. palustris  
Salmonella typhi  
P. aeruginosa
```


awk (**A**ho, **W**einberger & **K**ernighan) es un lenguaje de programación, y en su forma de comando nos sirve para manejo general de datos, extracción de campos (tablas) y operaciones algebraicas simples

La forma más sencilla del uso de awk es a través de print

```
$ awk '{print $1}' org_list
```

Escherichia

Rhodopseudomonas

Salmonella

Pseudomonas

```
$ awk '{print $2}' org_list
```

coli

palustris

typhi

aeruginosa

De forma natural, awk asume el espacio en blanco (y tabuladores) como delimitador de campo, y el número de campos lo entiende como $\$1, \$2, \$3 \dots \n

```
$ awk '{print $1,$2}' org_list
Escherichia coli
Rhodopseudomonas palustris
Salmonella typhi
Pseudomonas aeruginosa
```

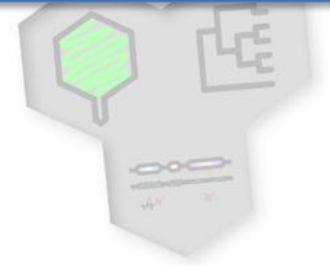
```
$ head BLAST_table.txt
T4_AAF23757 AF203974_1 100.00 100.00 607 0 0 1 607 1 607 0.0 1271
T4_AAF23757 AAQ15394_1 99.84 99.84 607 1 0 1 607 1 607 0.0 1271
T4_AAF23757 ABR24664_1 99.51 99.51 607 3 0 1 607 1 607 0.0 1265
T4_AAF23757 ACL78106_1 98.68 99.01 607 8 0 1 607 1 607 0.0 1253
```

```
$ awk '{print $3,$12,$13}' BLAST_table.txt
100.00 0.0 1271
99.84 0.0 1271
99.51 0.0 1265
98.68 0.0 1253
```

awk puede realizar operaciones simples (+ - × ÷) y dependiendo de la sintaxis que usemos, podemos construir operaciones complejas en un lenguaje relativamente sencillo.

```
$ echo 9 18
9 18
$ echo 9 18 | awk '{print $1+2,$2-2}'
11 16
$ echo 9 18 | awk '{print $1+$2}'
27
$ echo 9 18 | awk '{print $1-$2}'
-9
$ echo 9 18 | awk '{print $1*$2}'
162
$ echo 9 18 | awk '{print $1/$2}'
0.5
$ echo 9 18 | awk '{print ($2-$1)/$1}'
1
```

Por si te pierdes



man program

Manual, nos da la información de como correr un programa

program -h --h -help --help

Help, nos da un menú de ayuda para distintos programas, no todos lo traen

tfm program

The function manual (solo EMBOSS), nos da la ayuda de los programas que vienen en la paqueteria EMBOSS

wosname term

- ? Nos entrega el nombre de programas cuya descripción incluye el termino term (solo EMBOSS)