# Introducción Shell-UNIX y Programación en BASH

### Semana Internacional de la Ciencia



Facultad de Ciencias Escuela de Física

19 de septiembre de 2018





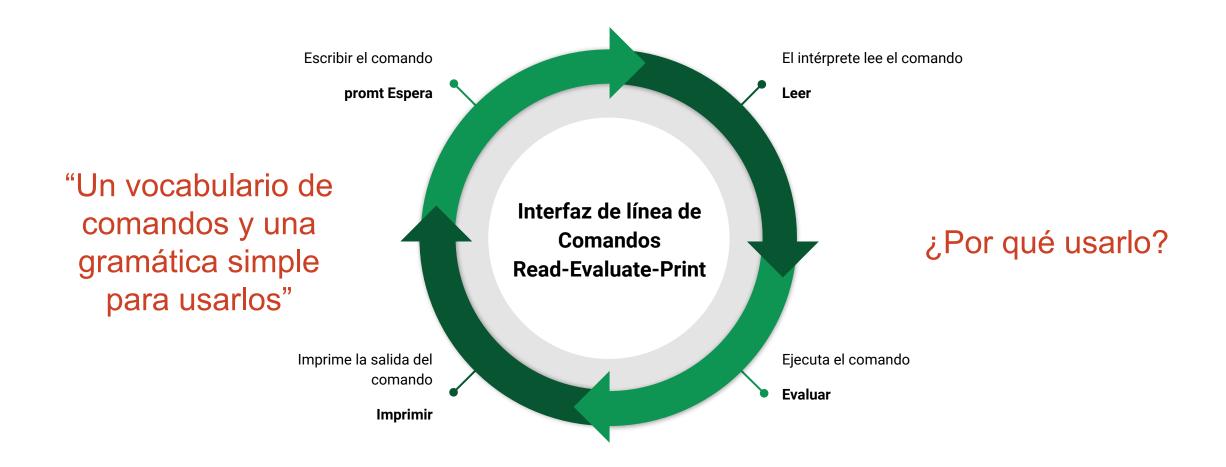




## **Estructura del Shell**

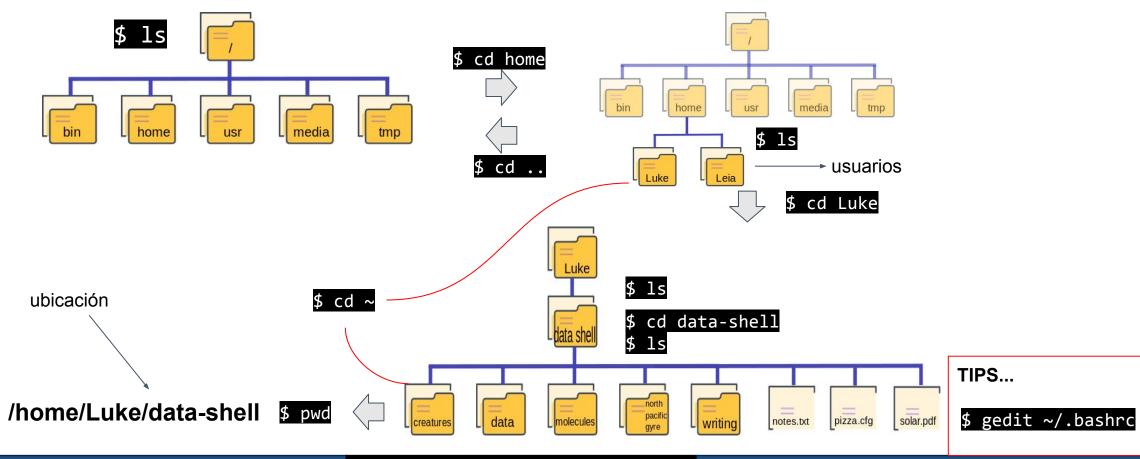


### Presentando el Shell



### Sistemas de archivos

Comandos: pwd, ls, cd, cd..



### Sistemas de archivos

A partir de /Users/amanda/data/, ¿cuál de los siguientes comandos podría utilizar Amanda para navegar a su directorio personal, /Users/amanda?

- 1. cd.
- 2. cd/
- cd /home/amanda
- 4. cd ../..
- 5. cd ~
- 6. cd home
- 7. cd ~/data/..
- 8. cd
- 9. cd..

### Sistemas de archivos

A partir de /Users/amanda/data/, ¿cuál de los siguientes comandos podría utilizar Amanda para navegar a su directorio personal, /Users/amanda?

- 1. cd.
- 2. cd/
- 3. cd /home/amanda
- 4. cd ../..
- 5. cd ~
- 6. cd home
- 7. cd ~/data/..
- 8. cd
- 9. cd..

- 1. No: representa el directorio actual.
- 2. No: / representa el directorio raíz.
- 3. No: el directorio principal de Amanda es /Users/amanda.
- **4. No:** esto sube dos niveles, es decir, termina en /Users.
- 5. Sí: ~ representa el directorio de inicio del usuario, en este caso /Users/amanda.
- **6.** No: esto navegaría a un directorio llamado home, si existe.
- 7. **Sí:** innecesariamente complicado, pero correcto.
- 8. Sí: acceso directo para volver al directorio de inicio del usuario.
- 9. Sí: sube un nivel

## **Archivos y Directorios**

Con el comando cd ingrese al directorio de trabajo data-shell

Estando en data-shell cree un nuevo directorio llamado tesis \$ mkdir tesis

Verifique que ha creado el directorio tesis dentro de el directorio data-shell \$ 1s -F

Ingrese al directorio tesis \$ cd tesis, cree y edite un archivo de texto llamado borrador.txt \$ gedit borrador.txt

Podemos eliminar el archivo creado \$ rm borrador.txt

\$ rm tesis ??

## **Archivos y Directorios**

Con el comando cd ingrese al directorio de trabajo data-shell

Estando en data-shell cree un nuevo directorio llamado tesis \$ mkdir tesis

Verifique que ha creado el directorio tesis dentro de el directorio data-shell \$ 1s -F

Ingrese al directorio tesis \$ cd tesis, cree y edite un archivo de texto llamado borrador.txt \$ gedit borrador.txt

Podemos eliminar el archivo creado \$ rm borrador.txt

\$ rm tesis ??

\$ rm −r tesis 🗸

rm: cannot remove tesis: Is a directory

TIPS... \$ rm -i

## **Archivos y Directorios**

Estando en el directorio data-shell creemos de nuevo el directorio tesis \$ mkdir tesis

Cree y edite los archivos de texto llamados final.txt y final\_final.txt

```
$ gedit final.txt
```

\$ gedit final\_final.txt

Ingrese al directorio tesis \$ cd tesis y verifiquemos qué hay dentro de este este directorio con \$ 1s

### No están los archivos!

Se han creado erróneamente en el directorio data-shell, si queremos moverlos a el directorio tesis podemos utilizar comando mv

TIPS...

\$ cp copiar

### **Usando "Wildcards"**

- \* es un wildcard. Coincide con cero o más caracteres.
- ? también es un wildcard, pero solo coincide con un solo caracter.

Estando en el directorio molecules cuál comando con \$ 1s retorna los siguientes archivos ? ethane.pdb methane.pdb

```
1. $ 1s *t*ane.pdb
```

- 2. \$ ls \*t?ne.\*
- 3. \$ 1s \*t??ne.pdb
- 4. \$ ls ethane.\*

#### **RESUMEN: DIRECTORIOS Y WILDCARDS**

- cd ruta cambia el directorio de trabajo actual.
- 1s ruta imprime una lista de un archivo o directorio específico; 1s una lista del directorio de trabajo actual.
- pwd imprime el directorio de trabajo actual
- es el directorio raíz de todo el sistema de archivos.
- significa 'el directorio sobre el actual'; le por sí solo significa 'el directorio actual'.
- cp copia un archivo.
- mkdir crea un nuevo directorio.
- mv mueve (renombra) un archivo o directorio.
- ❖ rm elimina un archivo; rm -1 elimina un archivo.
- \* coincide con cero o más caracteres en un nombre de archivo, por lo que \*.txt coincide con todos los archivos que terminan en .txt.
- con any.txt.

"World Count": wc este cuenta el número de líneas, palabra y caracteres de un archivo.

Podemos redireccionar la salida de un comando a un archivo, en lugar de verlo en la pantalla de Shell, usando >

```
$ cd molecules
$ wc -1 *.pdb > lineas.txt
$ cat lineas.txt
```

\$ cat muestra el contenido en el archivo.

```
$ wc -1 número de líneas
$ wc -w número de palabras
```

\$ wc -c número de caracteres

TIPS...

Cual es la diferencia? entre > y >>

```
$ echo hello > prueba01.txt
```

```
$ echo hello >> prueba02.txt
```

En el primer ejemplo con >, la cadena "hello" se escribe en archivo de prueba01.txt, **pero el archivo** se sobrescribe cada vez que ejecutamos el comando.

Vemos en el segundo ejemplo que el operador >> también escribe "hola" en un archivo (en este caso, archivo de prueba02.txt), pero agrega la cadena al archivo

```
$ head -n 3 animals.txt > animalsUpd.txt
$ tail -n 2 animals.txt >> animalsUpd.txt
```

Cuáles lineas quedarán en el archivo animalsUpd.txt??

```
TIPS...
```

```
$ head primeras líneas
```

\$ tail últimas líneas

La barra vertical, , entre los dos comandos se llama tubería ("**pipe**"). Le dice al shell que queremos usar la salida del comando de la izquierda como entrada al comando de la derecha.

Queremos enviar la salida de wc directamente a sort, y luego la salida resultante a head.

```
$ wc -1 *.pdb | sort -n
$ wc -1 *.pdb | sort -n | head -n 1
```

```
$ wc -1 *.pdb |
                                        $ wc -1 *.pdb
                sort -n
$ wc -1 *.pdb |
                sort -n
                            head -n 1
                                           wc -1 *.pdb out
                                                                 Output in Shell
                                        $ wc -1 *.pdb > lengths
                                           wc -1 *.pdb оит
                                                                  Output in File
                                        $ wc -1 *.pdb | sort -n | head -n 1
                                           wc -1 *.pdb OUT IN
                                                                  sort -n out IN head -n 1 out
                                                                                                      Output in Shell
```

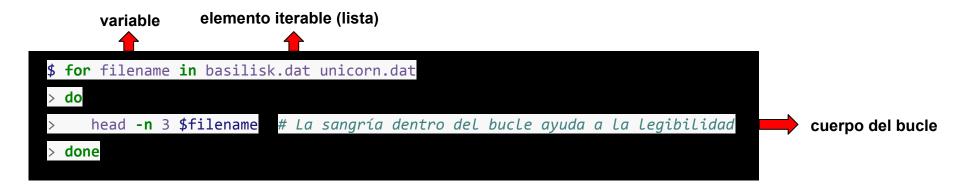
#### **RESUMEN: TUBERIAS Y FILTROS**

- cat muestra el contenido de sus entradas.
- head muestra las primeras 10 líneas de su entrada.
- tail muestra las últimas 10 líneas de su entrada.
- sort ordena sus entradas.
- wc cuenta líneas, palabras y caracteres en sus entradas.
- Comando > archivo redirige la salida de un comando a un archivo.
- Primero | Segundo es una tubería: la salida del primer comando se usa como entrada para el segundo.

### **Bucles**

Los bucles (loops en inglés) son fundamentales para mejorar la productividad a través de la automatización, debido a que nos permiten ejecutar comandos de forma repetitiva.

Sintaxis de un bucle for

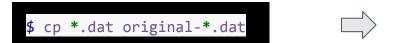


#### Nota:

- ❖ \$ se refiere al prompt, pero también se utiliza para pedir que la terminal obtenga el valor de una variable.
- se refiere al prompt dentro del bucle, pero también se utiliza para redirigir la salida de un comando.
- ; se utiliza para separar dos comando escritos en una sola línea.

En este ejemplo, usaremos el directorio creatures que sólo tiene dos archivos de ejemplo, pero los principios se pueden aplicar a muchos más archivos a la vez. Nos gustaría modificar estos archivos, pero también guardar una versión de los archivos originales, nombrando las copias original-basilisk.dat y original-unicorn.dat.

No se puede usar:



Puesto que se expandiría a:

```
$ cp basilisk.dat unicorn.dat original-*.dat
```

Esto no respalda nuestros archivos, en su lugar obtenemos un error:

```
cp: target `original-*.dat' is not a directory
```

En cambio, podemos usar un **bucle** para ejecutar una operación a la vez sobre cada cosa en una lista:

```
$ for filename in *.dat
> do
> cp $filename original-$filename
> done
```

Este bucle ejecuta el comando cp una vez para cada nombre de archivo:

- cp basilisk.dat original-basilisk.dat
- cp unicorn.dat original-unicorn.dat

Aquí un ejemplo sencillo que muestra las tres primeras líneas de cada archivo de una sola vez:



Teniendo en cuenta el bucle anterior, hemos llamado a la variable en este bucle **filename** con el fin de hacer su propósito más claro para los lectores humanos. A la terminal no le importa el nombre de la variable; si escribimos este bucle como:

```
$ for x in basilisk.dat unicorn.dat
> do
> head -n 3 $x
> done
$ for temperature in basilisk.dat unicorn.dat
> do
> head -n 3 $temperature
> done
```

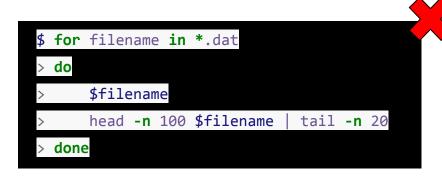
#### Nota:

Evitar el uso nombres sin sentido x o nombres engañosos temperature.

He aquí un bucle un poco más complicado:

```
$ for filename in *.dat
> do
> echo $filename
> head -n 100 $filename | tail -n 20
> done
```

En este caso, ya que la terminal expande **\$filename** para que sea el nombre de un archivo, **echo \$filename** sólo imprime el nombre del archivo hello there. Ten en cuenta que no podemos escribir esto como:

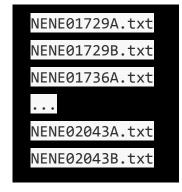


La primera vez a través del bucle, cuando **\$filename** se expande a **basilisk.dat**, la terminal intentará ejecutar **basilisk.dat** como un programa.

#### PIPELINE DE NELLE: PROCESANDO ARCHIVOS

Nelle ahora está lista para procesar sus archivos de datos. Dado que todavía está aprendiendo cómo utilizar la terminal, decide construir los comandos requeridos en etapas. Su primer paso es asegurarse de que puede seleccionar los archivos correctos (recuerda, aquellos cuyos nombres terminan en 'A' o 'B', en lugar de 'Z'). Posicionada en su directorio home, Nelle teclea:

```
$ cd north-pacific-gyre/2012-07-03
$ for datafile in NENE*[AB].txt
> do
> echo $datafile
> done
```



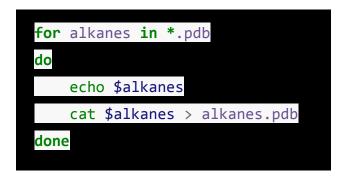
Su siguiente paso es decidir cómo llamar a los archivos que creará el programa de análisis goostats. Prefijar el nombre de cada archivo de entrada con "stats" parece simple, así que modifica su bucle para hacer eso:

```
$ for datafile in NENE*[AB].txt
> do
> echo $datafile stats-$datafile
> done
```



NENE01729A.txt stats-NENE01729A.txt
NENE01729B.txt stats-NENE01729B.txt
NENE01736A.txt stats-NENE01736A.txt
...
NENE02043A.txt stats-NENE02043A.txt
NENE02043B.txt stats-NENE02043B.txt

#### REDIRECCIONAMIENTO DE LA SALIDA



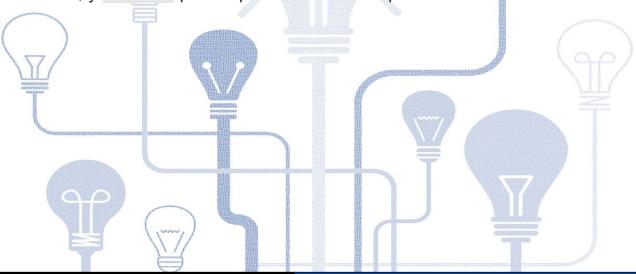
El texto de cada archivo se escribe (uno a la vez) en alkanes.pdb. Sin embargo, el archivo se sobrescribe en cada iteración del bucle.



>> concatena en un archivo, en lugar de sobrescribirlo con la salida del comando. Dado que la salida del comando cat ha sido redirigida, nada se imprime en pantalla.

#### **RESUMEN: BUCLES**

- Un bucle for repite comandos una vez para cada elemento de una lista.
- Cada bucle for necesita una variable para referirse al elemento en el que está trabajando actualmente.
- ❖ Uso de \$name para expandir una variable (es decir, obtener su valor). También se puede usar \${name}.
- No utilizar espacios, comillas o caracteres especiales como "" o "?" en nombres de directorios, ya que complica la expansión de variables.
- Proporcionar a los archivos nombres coherentes que sean fáciles de combinar con los caracteres especiales para facilitar la selección de los bucles.
- Utilizar la tecla de flecha hacia arriba para desplazarse por los comandos anteriores para editarlos y repetirlos.
- Usar Ctrl-R para buscar a través de los comandos previamente introducidos.
- Usar history para mostrar comandos recientes, y !number para repetir un comando por número.



## Scripts de Shell

Vamos a tomar los comandos que repetimos con frecuencia y los vamos a guardar en archivos, de modo que, podemos volver a ejecutar todas esas operaciones escribiendo un sólo comando. Por razones históricas, a un conjunto de comandos guardados en un archivo se le llama un **script** de la terminal.

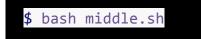
Comencemos por volver a molecules/ creando un nuevo archivo, middle.sh, que se convertirá en nuestro script de la terminal:

```
$ cd molecules
$ gedit middle.sh
```

Podemos usar el editor de texto para editar directamente el archivo - simplemente insertaremos la siguiente línea:



Una vez que hayamos guardado el archivo, podemos pedirle a la terminal que ejecute los comandos que contiene. Nuestra terminal se llama bash, por lo que ejecutamos el siguiente comando:





ATOM	9	Н	1	-4.502	0.681	0.785	1.00	0.00
ATOM	10	Н	1	-5.254	-0.243	-0.537	1.00	0.00
ATOM	11	Н	1	-4.357	1.252	-0.895	1.00	0.00
ATOM	12	Н	1	-3.009	-0.741	-1.467	1.00	0.00
ATOM	13	Н	1	-3.172	-1.337	0.206	1.00	0.00

#### **GENERALIZACIÓN DE SCRIPTS**

¿Qué pasa si queremos seleccionar líneas de un archivo arbitrario? Podríamos editar middle.sh cada vez para cambiar el nombre de archivo, pero eso probablemente llevaría más tiempo que simplemente volver a escribir el comando. En cambio, editemos middle.sh y hagamos que sea más versátil:

\$ gedit middle.sh

Ahora, dentro de "gedit", reemplaza el texto octane.pdb con la variable especial denominada \$1:

head **-n** 15 **"\$1"** | tail **-n** 5

Dentro de un **script** de la terminal, \$1 significa "el primer nombre de archivo (u otro parámetro) en la línea de comandos". Ahora podemos ejecutar nuestro **script** de esta manera:

\$ bash middle.sh octane.pdb



ATOM	9	Н	1	-4.502	0.681	0.785	1.00	0.00
ATOM	10	Н	1	-5.254	-0.243	-0.537	1.00	0.00
ATOM	11	Н	1	-4.357	1.252	-0.895	1.00	0.00
ATOM ATOM ATOM ATOM	12	Н	1	-3.009	-0.741	-1.467	1.00	0.00
ATOM	13	Н	1	-3.172	-1.337	0.206	1.00	0.00

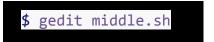
\$ bash middle.sh pentane.pdb



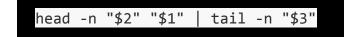
ATOM	9	Н	1	1.324	0.350	-1.332	1.00	0.00
ATOM	10	Н	1	1.271	1.378	0.122	1.00	0.00
ATOM ATOM ATOM ATOM	11	Н	1	-0.074	-0.384	1.288	1.00	0.00
ATOM	12	Н	1	-0.048	-1.362	-0.205	1.00	0.00
ATOM	13	Н	1	-1.183	0.500	-1.412	1.00	0.00

### **GENERALIZACIÓN DE SCRIPTS**

Aún así necesitamos editar middle.sh cada vez que queramos ajustar el rango de líneas. Vamos a arreglar esto usando las variables especiales \$2 y \$3 para el número de líneas que se pasarán respectivamente a head y tail:







Ahora podemos ejecutar:





ATOM	9	Н	1	1.324	0.350	-1.332	1.00	0.00
ATOM	10	Н	1	1.271	1.378	0.122	1.00	0.00
ATOM	11	Н	1	-0.074	-0.384	1.288	1.00	0.00
ATOM	12	Н	1	-0.048	-1.362	-0.205	1.00	0.00
ATOM	13	Н	1	-1.183	0.500	-1.412	1.00	0.00

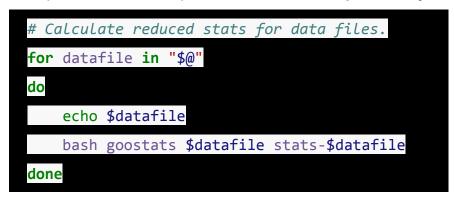
\$	bash	middle.sh	pentane.pdb	20	5
----	------	-----------	-------------	----	---



ATOM	14	Н	1	-1.259	1.420	0.112	1.00	0.00
ATOM	15	Н	1	-2.608	-0.407	1.130	1.00	0.00
ATOM	16	Н	1	-2.540	-1.303	-0.404	1.00	0.00
ATOM	17	Н	1	-3.393	0.254	-0.321	1.00	0.00
TER	18		1					

#### PIPELINE DE NELLE: CREANDO UN SCRIPT

El supervisor de Nelle insistió en que todos sus análisis deben ser reproducibles. Nelle se da cuenta que debería haber proporcionado un par de parámetros adicionales a goostats cuando procesó sus archivos. La forma más fácil de capturar todos los pasos es en un **script**. Ella ejecuta el editor y escribe lo siguiente:



Guarda esto en un archivo llamado do-stats.sh para que ahora pueda volver a hacer la primera etapa de su análisis escribiendo:

Por otra parte, el script podría haberlo escrito así:

\$ bash do-stats.sh NENE\*[AB].txt



# Calculate stats for Site A and Site B data
files.
for datafile in NENE\*[AB].txt
do
 echo \$datafile
 bash goostats \$datafile stats-\$datafile
done

### **DEPURACIÓN (DEBUGGING) DE SCRIPTS**

Supongamos que se ha guardado el siguiente **script** en un archivo denominado do-errors.sh en el directorio north-pacific-gyre/2012-07-03 de Nelle:

```
# Calcular las estadísticas de los archivos de datos.
for datafile in "$@"
do
    echo $datafile; bash goostats $datafile stats-$datafile
done
```

Ahora, ejecuta el **script** utilizando la opción -x

```
$ bash -x do-errors.sh NENE*[AB].txt
```

#### Nota:

El indicador -x hace que bash se ejecute en modo de depuración.

#### **RESUMEN: SCRIPTS**

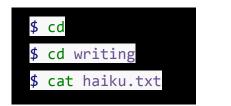
- Guardar comandos en archivos (normalmente llamados scripts de la terminal) para su reutilización.
- bash filename ejecuta los comandos guardados en un archivo.
- \$ \$@se refiere a todos los parámetros de la línea de comandos de un **script** de la terminal.
- \* \$1, \$2, etc., se refieren al primer parámetro de la línea de comandos, al segundo parámetro de la línea de comandos, etc.
- Coloque las variables entre comillas si los valores tienen espacios en ellas.

Dejar que los usuarios decidan qué archivos procesar es más flexible y más consistente con los comandos de Unix.

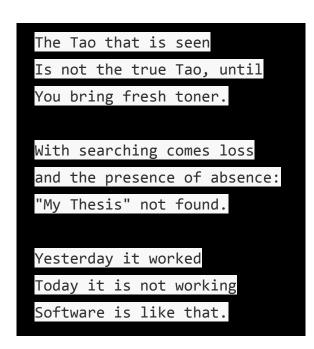


## **Encontrando archivos: grep**

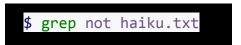
El comando grep encuentra e imprime líneas en archivos que coinciden con un patrón. Para nuestros ejemplos, usaremos un archivo que contenga tres haikus publicados en 1998 en la revista *Salon*. Para este conjunto de ejemplos, Vamos a estar trabajando en el subdirectorio "writing":







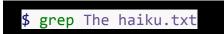
Busquemos líneas que contengan la palabra "not":





Is not the true Tao, until
"My Thesis" not found
Today it is not working

Vamos a probar un patrón distinto: "The".



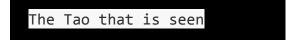


The Tao that is seen "My Thesis" not found.

Para restringir los aciertos a las líneas que contienen la palabra "The" por sí sola, podemos usar grep con el indicador -w. Esto limitará los coincidencias a palabras.







A veces, queremos buscar una frase, en vez de una sola palabra. Esto puede hacerse fácilmente con grep usando la frase entre comillas.

\$ grep -w "is not" haiku.txt



Today it is not working

Otra opción útil es -n, que numera las líneas que coinciden:



Podemos combinar la opción –w para encontrar las líneas que contienen la palabra "the" con –n para numerar las líneas que coinciden:



Ahora queremos usar la opción -i para hacer que nuestra búsqueda sea insensible a mayúsculas y minúsculas:

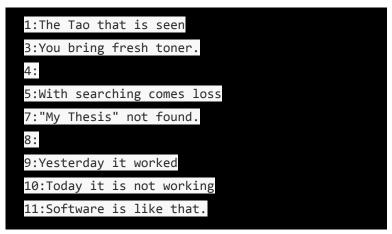
```
$ grep -n -w -i "the" haiku.txt

2:Is not the true Tao, until
6:and the presence of absence:
```

Ahora, queremos usar la opción -v para invertir nuestra búsqueda, es decir, queremos que obtener las líneas que NO contienen la palabra "the".



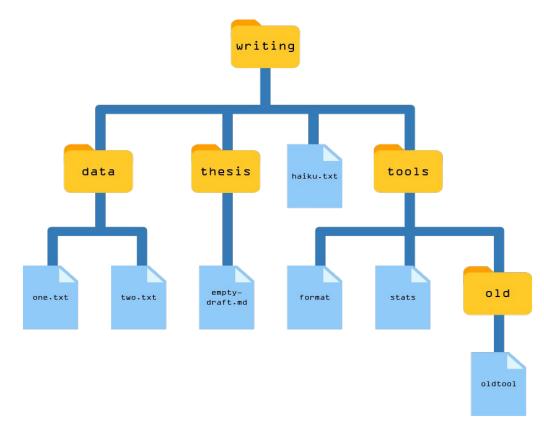




grep tiene muchas otras opciones. Para averiguar cuáles son, podemos escribir:

## **Encontrando archivos: find**

Mientras grep encuentra líneas en los archivos, El comando find busca los archivos. Una vez más, tienes muchas opciones. Para mostrar cómo funcionan las más simples, utilizaremos el árbol de directorios que se muestra a continuación.

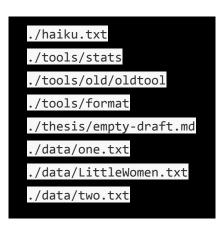


La primera opción en nuestra lista es -type d que significa "encontrar directorios".



Si cambiamos -type d por -type f, recibimos una lista de todos los archivos:





Ahora tratemos de buscar por nombre:

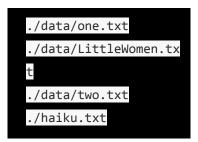


El problema es que el shell amplía los caracteres comodín como \* antes de ejecutar los comandos. El comando que ejecutamos era:

```
$ find . -name haiku.txt
```

Para conseguir lo que queremos, vamos a hacer lo que hicimos con grep: escribe \* txt entre comillas simples para evitar que el shell expanda el comodín \*





#### LISTAR Y BUSCAR

ls y find se pueden usar para hacer cosas similares dadas las opciones correctas, pero en circunstancias normales, ls enumera todo lo que puede, mientras que find busca cosas con ciertas propiedades y las muestra. ¿Cómo podemos combinar eso con wc -l para contar las líneas en todos esos archivos?

La forma más sencilla es poner el comando find dentro de \$():



```
11 ./haiku.txt
300 ./data/two.txt
21022
./data/LittleWomen.txt
70 ./data/one.txt
21403 total
```

#### **RESUMEN: ENCONTRANDO ARCHIVOS**

- find encuentra archivos con propiedades específicas que coinciden con los patrones especificados.
- \* grep selecciona líneas en archivos que coinciden con los patrones especificados.
- --help es un indicador usado por muchos comandos bash y programas que se pueden ejecutar desde dentro de Bash, se usa para mostrar más información sobre cómo usar estos comandos o programas.
- \* man command muestra la página del manual de un comando.

