

Herramientas de Programación en Bioinformática y Biología Computacional

Resumen

Esta es una asignatura introductoria de los conceptos más importantes que se irán extendiendo a lo largo de todo el máster. Se divide en tres módulos: Linux, Python y Bases de Datos.

Índice general

I	Linux - Bash	2
I.1	Tratamiento de ficheros y directorios	2
I.2	Mostrar contenidos en pantalla	4
I.3	Buscar contenidos de ficheros	4
I.4	Redireccionamientos y pipes	5
I.5	Wildcards y convenciones en nombres de archivo	6
I.6	Seguridad del sistema de archivos (permisos de acceso	6
I.7	Procesos y trabajos	7
I.8	Otros comandos útiles de UNIX	8
I.9	Variables	9
I.10	Scripts	10
I.10.1	Variables especiales en scripts	10
I.11	Branching	11
I.11.1	Expresiones condicionales	11
I.11.2	Bucles	13
II	Programación en Python	14
II.1	Variables	15
II.2	Primeras nociones	15
II.3	Expresiones condicionales	16
II.3.1	Operadores condicionales	16
II.3.2	Operadores lógicos	17
II.4	Bucles o loops	17
II.4.1	For loops	17
II.4.2	While loops	18
II.5	Cadenas o strings	18
II.5.1	F-strings	19
II.5.2	Indexación y slicing	19
II.5.3	Métodos de cadenas	20
II.6	Estructuras de datos	20
II.6.1	Listas	20
II.6.2	Diccionarios	21
II.7	Leer ficheros	22

Capítulo I

Linux - Bash

Todos los comandos se lanzan desde la terminal de Linux. Se pueden explorar los ficheros del árbol de directorios con un explorador que es equivalente al explorador de Windows. En cuanto se maneje con ficheros grandes, o se requieran acciones mínimas, las herramientas de la interfaz gráfica no van a ser suficientes - de ahí que tengamos que trabajar con la terminal.

man

Con el comando `man comando`, se accede al manual para buscar el comando con su nombre, descripción y posibles parámetros tanto en su versión larga como corta.

I.1. Tratamiento de ficheros y directorios

ls

El comando `ls` muestra la lista de la información del directorio en el que nos encontremos. Muestra tanto los ficheros como los directorios. En Ubuntu, distingue los directorios en azul oscuro y los ficheros en verde, pero el código de colores puede cambiar en la configuración y puede no ser igual siempre, por lo que no nos deberíamos fijar en eso. El comando se puede completar con una serie de **parámetros detrás de un guion que modifican el comportamiento del comando inicial**. `ls -l` lista el contenido del directorio con información adicional que incluye la localización del fichero, el tipo de fichero, el tamaño en bytes, los permisos, el inode y cuándo se modificó por última vez. Algunos tamaños pueden ser difíciles de leer a simple vista. Para ello, se emplea `ls -lh` que lista el contenido con el tamaño largo, de forma que indica las unidades de forma más comprensible a simple vista. De hecho, la `h` viene de "human". El parámetro `-t` ordena la lista por orden de creación o modificación de más reciente a más antiguo, y `-p` añade una barra invertida a los directorios para poder distinguirlos de los ficheros. Si se escribe `-lp`, la primera columna tiene una `d` al inicio de todos los directorios, mientras que los ficheros tienen un guion como primer carácter. Así, todos los parámetros se pueden concatenar en `ls -lhtp`. El parámetro `ls -a` lista el contenido del directorio en el que nos encontremos, incluyendo los ficheros ocultos cuyo nombre empieza por un punto (por ejemplo, `".config"`). Estos ficheros no aparecen a no ser que se añada el parámetro `a` para protegerlos y que no se modifiquen sin querer. "Si no quieres que alguien meta la pata, no le enseñes dónde puede meter la pata". En función de los permisos que se tenga, luego se pueden modificar esos ficheros o no. El parámetro `-i` añade un inode que será importante cuando seamos administradores,

ya que es un índice que único para cada fichero que permite modificar un fichero sin modificar otro aunque se llamen igual. Aunque no se puedan tener dos ficheros en el mismo directorio que se llamen igual porque el path sería el mismo, sí puede haber dos ficheros con el mismo nombre en directorios distintos, ya que el path sería diferente, al igual que el inode.

mkdir Para crear nuevas carpetas o directorios, en el explorador es click derecho y crear nueva carpeta. En la terminal, esto se traduce en el comando `mkdir nombre-carpeta-nueva`. El tamaño de un directorio no es el tamaño de lo que tenga dentro el directorio; al crear una carpeta nueva, ya tiene un tamaño de 4K, y si se tuvieran muchos ficheros muy pesados, no tiene por qué ser la suma de todos los tamaños que contenga ese directorio.

cd Para cambiar de directorio, se emplea el comando `cd nuevo-directorio`. Así, el siguiente prompt incluye el directorio en el que se encuentra. Al entrar en un directorio que acabamos de crear, con el comando `ls -lht` no aparece nada, pero con `ls -lhta` aparecen dos directorios ocultos llamados "." y "..". Los dos puntos son un **enlace al directorio anterior o padre**, de forma que con `cd ..` se va al directorio anterior. El punto simple es el **directorio actual**.

pwd El comando `pwd` muestra dónde se encuentra el **directorio actual**, es decir, el **directorio de trabajo o working directory**. El primer símbolo es una barra (/) que indica el **directorio raíz**. La virgulilla (~) lleva a la "carpeta personal", que es una traducción de "**home directory**". Se puede mover entre directorios ya sea por path directo mediante el uso del directorio raíz o mediante el árbol del directorio con los dobles puntos o la virgulilla.

cp Para copiar un archivo, se emplea el comando `cp archivo-original archivo-nuevo`. Hay comandos que no necesitan **argumentos**, y hay comandos (como este) que sí necesitan. En este caso, requiere de dos argumentos: el archivo que se quiere copiar y el nombre que se le quiere dar a la copia. El orden es siempre **comando -parámetro argumento**. Si se quiere copiar un directorio con todo su contenido a otro directorio, es necesario añadir el parámetro `cp -r directorio-original directorio-nuevo` (r de recursivo, es decir, todo lo que haya dentro). La fecha de los ficheros copiados es de cuando se crea la copia, no la del fichero original. Solo se podrán copiar archivos en aquellos directorios en los que tengamos los permisos apropiados por el administrador. Antes de realizar el comando para copiar, la terminal realiza una serie de comprobaciones, de forma que si el nombre del archivo a copiar tiene alguna errata, salta un error al no poder realizarse el comando `stat`, que forma parte de las comprobaciones previas. Si se copia un directorio en otro preexistente, se copia el primer directorio en el segundo. Esto resulta en que el segundo directorio incluye los ficheros que ya estaban y un nuevo directorio nuevo dentro - no se sobrescribe todo el directorio ni se borra lo que hubiese ahí. Aun así, sigue siendo recomendable proteger los directorios sensibles añadiendo un punto antes del nombre o directamente cambiando los permisos.

mv Para mover un archivo, se utiliza el comando `mv`. Se utiliza de forma similar a `cp`, pudiendo mover un fichero a otro directorio o al mismo con el mismo u otro nombre siempre que el nuevo directorio ya exista previamente. Si se mueve un directorio entero, se mueve con todo el contenido que tenga en el nuevo directorio, y también se le puede cambiar el nombre.

rm
rmdir

A la hora de querer borrar, se utiliza **rm** para los archivos y **rmdir** para los directorios. Para esto último, es necesario que el directorio esté vacío. En caso de que no, se puede utilizar **rm -r directorio/** o **rm -r ./directorio/** para borrar tanto el directorio como su contenido.

ln

En lugar de copiar un fichero y tenerlo doble, se puede crear un link simbólico. Esto quiere decir que se crea un fichero que no tiene el contenido del fichero original (y por tanto pesa menos, solo 11 bytes), si no que redirige a ese archivo, permitiendo enlazar distintos directorios o ficheros en ordenadores en remoto. Para crear un link simbólico, el comando es **ln -s fichero-original fichero-simbolico**. Si se borra el fichero original, el enlace simbólico ya no tiene utilidad. Para borrar un enlace, se puede borrar como un fichero cualquiera con **rm**.

Todos los ficheros tienen un nombre y una extensión, y siempre se debe escribir completo. Los directorios no llevan nunca extensión. Se recomienda no utilizar espacios en blanco en los nombres tanto de los ficheros como de los directorios porque dificulta la navegación.

1.2. Mostrar contenidos en pantalla

clear

El comando **clear** borra los comandos de la pantalla para dejarla vacía.

cat

Para mostrar en pantalla el contenido de un fichero, se puede emplear el comando **cat fichero**. No obstante, este comando no diferencia cambio de página, si no que muestra todo el contenido por pantalla aunque ocupe más. También se puede utilizar este comando con varios ficheros para que se muestren uno detrás de otro.

less

Para mostrar el contenido de un fichero en la pantalla por partes, se emplea **less**. De esa forma no hay que subir y bajar en la pantalla, si no que se desplaza mediante el uso del espacio para ir por páginas enteras, **enter** para ir línea a línea, y la tecla **q** para salir (quit).

head
tail

El comando **head** muestra en pantalla las primeras 10 líneas de un fichero por defecto. Se le puede añadir un parámetro para indicar el número de líneas que se desea ver: **head -15 fichero** muestra las primeras 15 líneas. De forma similar, **tail** muestra las últimas 10 líneas de un fichero.

1.3. Buscar contenidos de ficheros

grep

El comando **grep** es fundamental al utilizarse mucho con los pipes y las redirecciones (véase más abajo). Se utiliza **grep cadena-buscada fichero**, y el resultado son todas las líneas que incluyen la cadena de caracteres que se ha buscado, tanto como palabra suelta como dentro de otra palabra. Utilizando **grep -w**, se filtran aquellas líneas en las que la cadena forma palabras individuales (si se busca **with**, descarga **within**, por ejemplo). Este comando es sensible a las mayúsculas y minúsculas. Para ignorar eso, se añade el parámetro **-i**. El parámetro **-v** muestra aquellas líneas

que no incluyan la cadena, `-n` incluye al inicio de cada línea el número de dicha línea, y `-c` el número de líneas que se deberían mostrar.

wc

El comando `wc` cuenta distintas cosas en un fichero: líneas, palabras, y bytes. Para obtener solo una de esas mediciones, se pueden añadir los parámetros `-l`, `-w`, `-c` respectivamente. Tradicionalmente, la cantidad de bytes indica la cantidad de caracteres, pero puede haber pequeñas diferencias por la codificación. Para obtener la cantidad de caracteres, se emplea `grep -m`.

1.4. Redireccionamientos y pipes

>

> >

Tras ejecutar un comando, el resultado que sale en pantalla es la **salida estándar**. Sin embargo, se puede **redireccionar la salida** a otro lugar mediante el símbolo `>`. Linux trata igual la salida estándar y un fichero, pero la salida estándar es la opción por defecto, por lo que nosotros debemos redireccionar la salida cuando lo queremos guardar en un fichero. Hay que tener cuidado porque si se redirecciona una salida a un fichero que ya existe, se sobrescribe el contenido del mismo. Si lo que queremos es que el contenido nuevo se añada detrás del contenido ya presente en un fichero, se deben emplear los dos símbolos `> >`.

<

También existe un **redireccionamiento de entrada**, siendo la entrada estándar lo que se escribe por teclado. Para redireccionar la entrada, se emplea el símbolo `<`.

Nunca se debe emplear el mismo fichero de entrada y de salida. Al leer un fichero, el programa va línea a línea, por lo que si se procesa la primera línea y se guarda, se sobrescribe el fichero que se usaba de entrada, afectando a la continuidad del programa.

gedit

El comando `gedit fichero` abre un editor de texto plano que nos permite crear un fichero y escribir su contenido. Para guardar el contenido que hemos escrito, pulsa `ctrl s` y para cerrar `ctrl q`. Aunque el editor de texto también muestre el contenido de un fichero que ya existe previamente, si el tamaño del fichero es muy grande, no es recomendable leer su contenido mediante este comando, si no mediante otros mencionados previamente como `less`.

sort

El comando `sort` ordena una entrada por teclado o un fichero alfanuméricamente. Si un fichero está separado en líneas, compara todos los caracteres de cada línea y empieza a ordenar por el primero. A este comando se le puede pasar directamente ficheros o se le puede redireccionar por entrada. El parámetro `-r` muestra el resultado en orden inverso, y `-u` filtra las líneas repetidas en un archivo. Para ordenar una sola parte, el parámetro `-k` permite elegir los campos a comparar.

|

Hay ocasiones que para obtener un resultado paso a paso, hay que crear ficheros temporales que modificar y a los que acceder. Esto se puede evitar mediante el uso de **pipes** o barras verticales (`|`), que conectan los comandos antes y después del pipe. De esta forma, se omite el paso intermedio y se utiliza la salida del primer comando como entrada del segundo.

1.5. Wildcards y convenciones en nombres de archivo

*
?

Los wildcards en informática son como comodines. Existen dos tipos: el asterisco y el símbolo de interrogación. El **asterisco** va a representar cualquier número de caracteres en el nombre de un fichero o directorio. Por ejemplo, si ponemos prueba*, puede resultar en el directorio pruebas y en los ficheros prueba.txt y prueba.pdf. Por el contrario, el **interrogante** representa exactamente un carácter. Así, al poner prueba?, el resultado será exclusivamente el directorio pruebas, pero no los ficheros.

A la hora de nombrar ficheros y directorios, se deben **evitar los símbolos especiales** tales como /, *, & y %. También es muy recomendable **evitar espacios entre caracteres**. En resumen, a la hora de poner un nombre a un archivo, se deberían usar solo caracteres alfanuméricos junto a barras bajas y puntos. Tradicionalmente, los nombres empiezan en minúscula y pueden terminar en un punto seguido de un grupo de letras que indican el contenido de un archivo (por ejemplo, poner al final de todos los archivos en código Python .py).

1.6. Seguridad del sistema de archivos (permisos de acceso)

Cada archivo y directorio tiene **permisos de acceso asociados**, que se pueden comprobar en la primera columna al realizar `ls -lht`. Se trata de una cadena de 10 símbolos d, r, w, x, -. La d solo estará presente en primera posición e indica que se trata de un directorio. Si se trata de un fichero, en primera posición hay un guion. Los 9 símbolos restantes se agrupan en grupos de 3 en 3 y representan los **permisos del usuario, del grupo al que pertenece el usuario (y no es el usuario) y de todos los demás** respectivamente. Las opciones son:

- r (read): permisos de lectura y copiado de un archivo y de listar el contenido de un directorio.
- w (write): permisos de escritura y modificación de un archivo y de crear y borrar los archivos del directorio o mover archivos a él.
- x (execution): permisos de ejecución de un archivo cuando sea apropiado, es decir, cuando sea ejecutable. Por ejemplo, los comandos de linux como `ls` o `mv` son ficheros que se encuentran en `/usr/bin/` y que se pueden ejecutar por todos a la hora de escribir esos comandos en la terminal.

chmod

Si en lugar del carácter aparece un guion, significa que ese permiso no está dado. Si eres el propietario de un fichero, se pueden cambiar los permisos mediante el comando `chmod`. Para indicar a quién se le quiere cambiar el permiso, se pone u (usuario), g (grupo), o (otros), a (todos). Para quitar un permiso, se escribe un guion (-) seguido del permiso que se quiere quitar y del fichero; para dar un permiso, se escribe un más (+) y el permiso seguido del nombre del fichero. Un igual (=) expresa los permisos que se quieren. Para poner permisos diferentes, se escriben separados por comas (pero

sin espacios), por ejemplo `chmod g+wx,o+w fichero`. También se puede ejecutar el comando `chmod 777 fichero` para darle todos los permisos a todos. Es importante tener en cuenta también los permisos de los directorios, ya que para poder modificar un fichero, no solo se deben tener los permisos para modificarlo, si no que a su vez debe estar en un directorio para el que se tengan los permisos para modificarlo.

A la hora de crear ficheros, Bash proporciona al usuario autor los permisos de lectura y escritura, pero no de ejecución. Esto es importante a la hora de crear archivos ejecutables con extensión `.sh`.

1.7. Procesos y trabajos

Un **proceso** es un programa que se ejecuta y que recibe un **identificador (PID)** por parte del sistema operativo. Hay procesos que los lanzamos nosotros, pero hay otros que funcionan por detrás como por ejemplo buscar conexiones Wifi.

top El comando `top` muestra todos los procesos que se están llevando a cabo en tiempo real. La terminal es un proceso, y cada vez que nosotros lanzamos un proceso desde la terminal, por lo que se genera un proceso con un proceso padre (la terminal). Los procesos no pueden ejecutarse a la vez (van secuencialmente), pero el sistema operativo va gestionando los procesos y sus tiempos para que parezca que van en paralelo. Los sistemas operativos modernos pueden contener varios núcleos o kernel que permite su funcionamiento a la vez, pero dentro de cada núcleo o kernel los procesos van de forma secuencial. Para salir de `top`, se debe ejecutar `ctrl c`.

ps El comando `ps` muestra todos los procesos que se han lanzado desde esa terminal. Añadiendo el parámetro `-ef`, se incluyen los procesos padres (lanzados por root) con el usuario que lanzó el proceso, el PID y el PID padre.

sleep El proceso `sleep` está parado el tiempo que se le indique. En realidad, va a estar más tiempo debido a la gestión del tiempo del sistema operativo que va alternando distintos procesos de forma intermitente. Por ello, aunque ese proceso sí dure el tiempo determinado, el usuario tiene que esperar algo más. Si se escribe `sleep 10s &`, se estará ejecutando de fondo (en el background), y lo que se devuelve a la terminal es el PID. Esto permite seguir utilizando la terminal para otros procesos mientras tanto. Si se va enviado un proceso en el foreground cuando se quería en el background, se puede abortar (con `ctrl c`) y volver a ejecutar bien o se puede detener (con `ctrl z`) y poner `bg` para enviarlo al background. Si por el contrario se quiere reiniciar un proceso suspendido, se pone `fg`. Esto reinicia el último proceso suspendido. Para especificar uno concreto, se debe poner su número de trabajo (no el PID): `fg $1`.

kill Para matar un proceso de forma eficiente, se puede usar el comando `kill -9` seguido del identificador del proceso. De esta forma se asegura que también se eliminen los subprocesos generados por ese proceso y se cierran todos los recursos. Sin embargo, no es posible matar los procesos de otros usuarios.

1.8. Otros comandos útiles de UNIX

- **df**: informa del espacio libre del sistema.
- **du**: saca los kilobytes utilizados por cada subdirectorio. Esto puede ser útil cuando se ha superado el almacenamiento para ver qué directorio tiene más archivos.
 - **-s**: muestra solo un resumen
- **gzip y gunzip**: gzip reduce el tamaño de un fichero utilizando el compresor Zip, resultando en un fichero con extensión .gz. gunzip descomprime el fichero .gz. Para esto, es necesario tener permisos de lectura.
- **tar**: combina varios archivos en uno único que puede o no estar comprimido.
 - **-c**: crea un archivo.
 - **-x**: extrae un archivo.
 - **-v**: muestra el progreso de un archivo.
 - **-f**: nombre de archivo.
 - **-t**: ver el contenido del archivo.
 - **-j**: comprime el archivo mediante bzip2.
 - **-z**: comprime el archivo mediante gzip.
 - **-r**: añade o actualiza archivos o directorios a archivos existentes.
 - **-C**: directorio especificado
 - **-exclude=**: excluye los archivos que vayan a continuación de la orden principal.
- **zcat**: lee archivos de extensión .gz sin la necesidad de descomprimirlos previamente.
- **cut**: extrae porciones de texto de un fichero al seleccionar columnas o caracteres.
 - **-d**: delimitador en comillas simples (por ejemplo, **-d ' | '**).
 - **-f**: campo o columna que se quiere extraer (**-f1** o, si se quieren varias, **-f1,2**).
- **echo**: muestra en la salida estándar lo que se ponga a continuación. Aunque funcione directamente, es recomendable escribir las cadenas de caracteres entre comillas dobles.
- **tr**: reemplaza o borra caracteres de una cadena. No admite un fichero de entrada, si no que debe venir por un pipe o utilizando el redireccionamiento de entrada.
 - **-s**: sustituye la secuencia que se ponga a continuación entre comillas simples con una única ocurrencia.
 - **-d**: elimina los caracteres proporcionados.

- **touch**: toca el fichero, es decir, accede a él, cambiando la fecha de acceso. Si el fichero no existe, se crea el fichero sin contenido (tamaño 0).
- **date**: imprime en pantalla el día y la hora. Se puede personalizar la forma en la que se muestra.
- **file**: muestra la información de uno o varios ficheros como su codificación.
- **stat**: muestra el archivo o el estado del archivo. También es posible acceder a la meta-información del fichero mediante el parámetro `-c` seguido de una secuencia válida.
- **basename y dirname**: **basename** elimina las partes del path en el nombre de un archivo para dejar únicamente el nombre propio. También puede eliminar el sufijo de un nombre. Por el contrario, **dirname** elimina el nombre del archivo para dejar únicamente el path.
- **diff**: compara el contenido de dos ficheros y muestra las diferencias entre ambos.
- **find**: busca en los directorios ficheros y directorios con un cierto nombre, fecha, tamaño u otro atributo que se quiera especificar. Por ejemplo, para buscar todos los archivos en el directorio actual que sean de texto plano, se puede poner `find . -name "*.txt" -print`. Para buscar únicamente directorios, se debe poner `-type d` y para buscar únicamente ficheros `-type e`.
- **xargs**: se utiliza usualmente para combinar comandos.
- **history**: muestra el historial de comandos.
- **ssh**: permite acceder a un servidor de Linux remoto.
- **nohup**: hace que haya procesos de fondo al trabajar desde un servidor remoto y que no se detengan cuando cerremos sesión.
- **gnome-terminal**: crea una nueva terminal.

1.9. Variables

En bash, las únicas variables que se puede crear son con **cadena de caracteres** (x=ejemplo). Si se pone un número, el intérprete interpreta el número como un carácter ("3" en Python). Aunque las comillas no sean necesarias, sí son recomendables (sí son obligatorias cuando hay espacios en blanco entre caracteres). Es importante **no dejar espacios** entre la variable, el igual y la cadena, ya que se interpretaría el nombre de la variable como un comando y puede dar un error. Para acceder al valor de la variable, se debe poner con un símbolo de dólar antes del nombre: `echo $x`. Si la variable que se accede no existe, la salida estará en blanco. Las variables que creamos se borran al apagar el ordenador. Para prevenir esto, se pueden guardar las variables en el fichero de configuración `.bashrc`.

El sistema crea automáticamente algunas variables denominadas **variables de ambiente** cada vez que se enciende el ordenador. Esto es importante saberlo para no cambiar el valor de estas variables. Ejemplos son:

- HOME: path completo del usuario hasta la carpeta personal o home directory.
- USER: usuario en el que nos encontramos.
- PATH: directorios separados por dos puntos donde el intérprete va a buscar los ficheros de los comandos si no se encuentran en el directorio de trabajo.
- ?: guarda un 0 si el último comando ha sido correcto y un entero diferente de 0 si la última acción no ha sido ejecutada correctamente (ha dado un error). Es importante remarcar que aquí, comando verdadero es 0, mientras que en Python y otros lenguajes de programación, True es cualquier número que no sea 0.

export En una terminal nueva, las variables que se habían creado no existen. Para ello, se puede usar el comando `export nombre-variable` para que terminales nuevas sí puedan acceder a las variables en ese estado. Si la variable se modifica posteriormente, si no se vuelve a exportar, la terminal nueva no tiene la actualización. A su vez, como la terminal nueva es un proceso nuevo, no le puede exportar sus variables al proceso padre.

readonly Para evitar que una variable se modifique, se puede utilizar el comando `readonly nombre-variable`.

Para guardar en una variable la salida de un comando, se debe emplear la expresión `x=$(comando)`. Entre los paréntesis y el comando sí se pueden dejar espacios para facilitar la lectura humana. Si hay una errata en el comando y da error, la asignación a la variable no se realiza y la variable se queda sin contenido.

Se puede crear una variable con el contenido de otras variables. Teniendo `x=3` y `y=4`, si se escribe `z=$x+$y`, al poner `echo $z`, el resultado es `3+4`. Esto se debe a que bash no interpreta el símbolo `+` como suma matemática. Para poder hacer eso, la declaración de variable y la suma se debe realizar entre doble paréntesis: `((z=x+y))`.

unset Para eliminar variables, se usa el comando `unset nombre-variable`.

I.10. Scripts

Los scripts son ficheros ejecutables de código. La extensión de estos ficheros depende del lenguaje de programación: en el caso de Python, los scripts tienen la extensión `.py`, y en el caso de bash, la extensión `.sh`. Para que un fichero se pueda ejecutar, se deben tener los permisos de ejecución. Además, la primera línea del código debe contener el directorio del intérprete que se debe utilizar. Tradicionalmente, los intérpretes se encuentran en `/usr/bin` o en `/bin`. Por ejemplo, la primera línea de un script de Python debe ser `#!/usr/bin/python3` y de Bash `#!/bin/bash`. Una vez hecho esto, desde la propia terminal se puede ejecutar el script.

I.10.1. Variables especiales en scripts

Es posible pasar argumentos a un script para personalizar la salida. Para asignar algo indeterminado a una variable, se emplea `$1 - $9`, de forma que al ejecutar el script, se deban poner los argumentos que pasará a reemplazar los huecos en ese orden. Es buena

práctica que en el script los argumentos pasados por teclado se guarden en variables con nombres que tengan sentido y luego sean esas variables las que se empleen en el resto del código. De esa forma, el código queda más claro visualmente y es más fácil de modificar sin equivocaciones.

Otras variables especiales son:

- \$0: el nombre del script de bash.
- \$#: la cantidad de argumentos que se han pasado al script.
- \$@: todos los argumentos que se han pasado al script.
- \$\$: el PID del script actual.
- HOSTNAME: nombre de la máquina en la que se está ejecutando el script.
- SECONDS: la cantidad de segundos desde que el script se inició.
- RANDOM: devuelve un número aleatorio diferente cada vez que se refiera a esto.
- LINENO: devuelve el número de la línea actual en el script de bash.

I.11. Branching

Las sentencias condicionales `if` y `case` ayudan a tomar decisiones en los scripts de bash. Permiten decidir si ejecutar ciertos fragmentos de código en base a ciertas condiciones.

I.11.1. Expresiones condicionales

Una expresión condicional básica hace que, si una condición es verdadera, se ejecute el siguiente código. Si la condición resulta ser falsa, entonces esas acciones no se desarrollan.

```
if [ condición ]
then
    código a ejecutar
fi
```

Los posibles operadores que se pueden utilizar en la condición son:

- `! expresión`: la expresión es falsa
- `-z string`: la longitud de `string` es 0, es decir, está vacía.
- `string1 = string2`: `string1` es igual que `string2`.
- `string1 != string2`: `string1` no es igual que `string2`.
- `integer1 -eq integer2`: `integer1` es numéricamente igual que `integer2`.

- `integer1 -gt integer2`: `integer1` es numéricamente mayor que `integer2`.
- `integer1 -lt integer2`: `integer1` es numéricamente menor que `integer2`.
- `-d fichero`: el fichero existe y es un directorio.
- `-e fichero`: el fichero existe.
- `-r fichero`: el fichero existe y tiene el permiso de lectura.
- `-s fichero`: el fichero existe y su tamaño es mayor que 0, es decir, no está vacío.
- `-w fichero`: el fichero existe y tiene el permiso de escritura.
- `-x fichero`: el fichero existe y tiene el permiso de ejecución.
- `fichero1 -nt fichero2`: es True si `fichero1` es más nuevo que `fichero2`.
- `fichero1 -ot fichero2`: es True si `fichero1` es más antiguo que `fichero2`.

Es posible conectar varias condiciones mediante el uso de `and` y `or`. Para ello, la forma preferida es combinar condiciones individuales con los operadores `&&` para `and` y `||` para `or`.

```
if [ condición 1 ] && [ condición 2 ]; then
    código a ejecutar
fi
```

Otra opción sería escribir los operadores en la propia condición con `-a` para `and` y `-o` para `or`. Para invertir la condición (buscar que algo no sea `x`), se utiliza `!` antes de la condición.

Las expresiones condicionales pueden requerir ciertas acciones si una condición es verdadera, y otras si es falsa. Para ello, se emplea `else`:

```
if [ condición ] ; then
    código a ejecutar si la condición es verdadera
else
    código a ejecutar si la condición es falsa
fi
```

En algunas ocasiones puede haber distintas condiciones que deriven en distintas acciones.

```
if [ condición ]
then
    código a ejecutar
elif [ otra condición ]
then
    otro código a ejecutar
else
    código diferente
fi
```

I.11.2. Bucles

Los bucles permiten repetir una serie de comandos hasta que se llega a un objetivo. Son útiles para ahorrar en código redundante y automatizar tareas repetitivas.

I.11.2.1. For loops

Los bucles for toman cada elemento de una lista en orden, lo asignan a una variable y se ejecuta el comando hasta que se acaben los objetos de la lista.

```
for variable in lista
do
    código a repetir
done
```

También es posible procesar una serie de números, pero para ello deben estar en formato C: {inicial..final..pasos}.

I.11.2.2. While loops

Los bucles while permiten repetir una lista de comandos mientras que la condición sea cierta y se detiene cuando la condición pasa a ser falsa:

```
while [ condición ]
do
    código a repetir
done
```

I.11.2.3. Until loops

El bucle until es similar al while, pero repite la lista de comandos hasta que la condición sea cierta, es decir, mientras sea falsa y se detiene cuando se hace verdadera.

```
until [ condición ]
do
    código a repetir
done
```

Capítulo II

Programación en Python

Los lenguajes de programación son la forma de comunicarnos con los ordenadores. Algunos lenguajes tienen bajo nivel de abstracción que trabajan en binario y por ello son más similares al propio sistema. Por otro lado, están los lenguajes de alto nivel de abstracción que son más cercanos al idioma humano. Python es el lenguaje con el mejor **balance entre potencia y simplicidad**.

La máquina ejecuta un programa ejecutable (.exe), que está escrito en ensamblador, el lenguaje del ordenador. Para algunos lenguajes de programación, se requiere de un compilador para que genere a partir de nuestro código un ejecutable que entienda la máquina, como por ejemplo C. En el caso de Python, lo que requiere es un **intérprete**, que genera a tiempo real las instrucciones del ordenador para ejecutar nuestro código. El intérprete solía ser más lento que el compilador al tener que estar traduciendo a tiempo real el código. No obstante, a día de hoy, Python puede alcanzar un 90 % del tiempo que tardaría un programa escrito en C en ejecutarse. Esto se debe a que hay problemas paralelizables, es decir, que se pueden dividir en partes y realizar al mismo tiempo. El rendimiento de un programa depende de entrada-salida, es decir, la lectura de los datos que se encuentran en el disco duro, y eso depende de la tecnología del disco duro, no del lenguaje de programación.

Python tiene un **gran soporte de librerías**, que son programas escritos por otras personas y que se pueden utilizar por los demás. Los comandos de Python se pueden escribir directamente desde la terminal (escribiendo primero `python3` y luego salir con `ctrl d`), pero eso no es cómodo. Es mejor crear primero un fichero fuente con extensión `.py` que luego pueda utilizar el intérprete. Para ello, se escribe `pico nombre.py`. Esto abre un editor de texto donde podemos escribir el código. Se guarda con `ctrl s` y se cierra con `ctrl x`. Para que el intérprete lo lea, se debe poner en la consola `python3 nombre.py`. En caso de que haya alguna errata o algún error en el código, a la hora de ejecutar el código por el intérprete se va a indicar en qué línea del código fuente se encuentra y qué problema hay. También existen **programas o IDEs** que son entornos donde poder editar el código fuente de forma más intuitiva. Ejemplos son Visual Studio Code, PyCharm, Spider, etc.

El código de Python se va ejecutando secuencialmente comando a comando.

II.1. Variables

Las **variables** son contenedores donde guardar un contenido al que poder referirse mediante el nombre de la variable. El contenido puede ser desde números a cadenas de texto, lo que define el tipo de variable (el tipo de contenido que contiene una variable). En otros lenguajes, es necesario declarar una variable antes de poder usarla (se crea antes de meterle un contenido). Esto en Python no es necesario; la **creación de tipos es implícita por el intérprete en función del primer valor que se le dé a la variable**. Así, si se escribe directamente `age = 35`, el intérprete le da el tipo `integer` (número entero) a la variable `age`. Los **tipos en Python son implícitos y dinámicos** (pueden cambiar), lo que puede ser muy cómodo, pero también puede dar lugar a errores. Los tipos son:

- Integer: número entero
- Float: número con decimales
- String: cadena de caracteres, ya sean letras o números, que van entre comillas simples o dobles. Una vez que están definidas, ya no se pueden modificar.
- Bool: booleanos, es decir, `True` o `False`.
- List: lista
- Tuple: tupla
- Dictionary: diccionario

En cuanto a los nombres de las variables, no pueden contener espacios en blanco ni empezar por un número. Python es case-sensitive, por lo que hay una distinción entre mayúsculas y minúsculas.

II.2. Primeras nociones

El comando `print` permite imprimir algo por pantalla.

El comando `input` le pide información al usuario que debe aportar por teclado.

```
name = input('Enter your name: ')\nprint('Hello, ', name)
```

Ese código está dividido en dos líneas de código. La primera línea es bloqueante, ya que le pide al usuario insertar por teclado su nombre. Una vez dado, ese nombre se asigna a la variable `name` que se emplea para imprimir por pantalla `Hello` y el nombre que se le ha insertado.

Aunque los tipos de variables sean implícitos, se puede realizar un **casting**, es decir, especificar el tipo de la variable:

- `str(variable)`: convierte en tipo string.

- `int(variable)`: convierte en tipo integer.
- `float(variable)`: convierte en tipo float.

Los comentarios en Python, marcados con `#`, sirven para poner anotaciones para otras personas o nosotros mismos y que el intérprete va a omitir. Es buena práctica poner comentarios para explicar las partes del código.

```
#Código de prueba
print('Esto es una prueba') #print
```

II.3. Expresiones condicionales

En Python, se pueden escribir condicionales: si algo es verdadero/falso, ocurre algo. Esto se hace mediante `if` y `else` y sirve para cambiar el flujo del programa o tomar decisiones.

```
num_pensado = 4
num = input("Elige un número: ")

if num_pensado == num:
    print("Enhorabuena, has acertado")
else:
    print("Sigue intentando")
```

Si se cumple la condición del `if`, el código que se ejecuta es el que se encuentra indentado y se continúa sin ejecutarse el bloque `else`. Si no se cumple el `if`, el código que se ejecuta es el que se encuentra indentado tras `else`, y se omite el bloque del `if`.

En el caso de que haya más condicionales, se puede utilizar `elif` seguido del siguiente condicional, siempre y cuando la condición sea de verdadero o falso:

```
num_pensado = 4
num = input("Elige un número: ")

if num_pensado == num:
    print("Enhorabuena, has acertado")
elif num_pensado < num:
    print("No has acertado, el número es más bajo")
else:
    print("No has acertado, el número es más grande")
```

II.3.1. Operadores condicionales

Los operadores condicionales son aquellos que resultan en `True` o `False`. Los operadores son:

- `x == y`: x es igual a y. Un solo igual asigna el valor de una variable a otra, dos iguales comprueban si son lo mismo.
- `x != y`: x no es igual a y.
- `x < y`: x es menor que y.
- `x <= y`: x es menor o igual que y.
- `x > y`: x es mayor que y.
- `x >= y`: x es mayor o igual que y.

II.3.2. Operadores lógicos

Para poder incluir varios condicionales, se utilizan los operadores lógicos **and** y **or**.

El operador **and** implica que las dos condiciones que se unen deben ser **True** para que el resultado sea **True**. Para el operador **or**, basta con que una condición sea **True** para que el resultado sea **True**.

X	Y	X and Y	X or Y
True	True	True	True
True	False	False	True
False	True	False	True
False	False	False	False

II.4. Bucles o loops

Los bucles permiten repetir ciertas sentencias para ahorrar código redundante.

II.4.1. For loops

Los for loops permiten repetir un código un **número determinado de veces**. La estructura típica de un for loop es:

```
for variable in range(número inicial, número final, incremento):  
    código que se quiere repetir
```



```
for variable in otra_variable:  
    código a repetir
```

Es importante remarcar que el código que se quiere repetir debe estar indentado. Si no se pone número inicial, automáticamente empieza en 0. De igual forma, el incremento es 1 por defecto. El número final no está incluido. Por ello, si sólo se pone `range(número)`, se podría interpretar que se repite el bucle se repite el número indicado de veces. Por ejemplo, utilizando el ejemplo anterior, se podría escribir:

```
num_pensado = 4

for i in range(5):
    num = input("Elige un número: ")
    if num_pensado == int(num):
        print("Enhorabuena, has acertado")
        break
    else:
        print("Sigue intentando")
```

La `i` es una variable que empieza en 0 y con cada iteración aumenta en uno hasta llegar a 4. Esto se hace automáticamente por el intérprete. Si se acierta el número antes, se sale del bucle mediante `break`.

II.4.2. While loops

Los bucles `while` se repiten un **número indeterminado de veces**. Este bucle depende de una condición:

```
while condición:
    código que se quiere repetir
```

En estos bucles, es importante que la condición se modifique dentro del bucle para no crear un bucle infinito.

```
num_pensado = 4
fallo = True

while fallo == True:
    num = input("Elige un número: ")
    if num_pensado == num:
        print("Enhorabuena, has acertado")
        fallo = False
    else:
        print("Sigue intentando")
```

Así, cuando el usuario acierta el número, la variable `fallo` pasa a ser falsa y el bucle se interrumpe.

II.5. Cadenas o strings

Las cadenas son cualquier secuencia de caracteres o números que se encuentran entre comillas simples o dobles. También hay cadenas vacías que no tienen ninguna secuencia entre las comillas. Las cadenas se pueden concatenar mediante el `+`. También se puede repetir el contenido de una cadena al multiplicarlo con las veces que se quiere repetir.

```
'AB' + 'cd' # ABcd
'Hi' * 2 # HiHi
```

len Se puede calcular la longitud de una cadena mediante len().

```
cadena = "Hola"
len(cadena) #4
```

in Para comprobar si una cadena contiene algo, se emplea el operador in. Para ver si no contiene algo, se usa not in. Esto realmente es una expresión booleana, ya que el resultado es verdadero o falso.

```
secuencia = "ATC"
if "G" in secuencia:
    print("La secuencia contiene guanina.")
else:
    print("La secuencia no contiene guanina.")
```

II.5.1. F-strings

A la hora de imprimir texto, es común alternar cadenas con valores de variables. Esto se puede hacer intercalando las cadenas entre comillas con las variables o con los denominados f-strings (format-strings). Los f-strings son expresiones con una sintaxis más fluida que permite darle formato a cadenas. Para imprimir el valor de variables, basta con poner el nombre de la variable entre llaves ({}). En caso de querer imprimir una llave en un f-string, se debe poner doble.

```
variable = 1
print(f"El valor de la variable es {variable}.")
# Output: El valor de la variable es 1.
print(f"El valor de la variable es {{{variable}}}.")
# Output: El valor de la variable es {1}.
```

II.5.2. Indexación y slicing

Las cadenas se pueden indexar, es decir, referir a una posición en la misma cadena. Esto se realiza escribiendo la cadena (o la variable que contiene la cadena) seguido de la posición que se quiere entre corchetes. La primera posición es un 0, y las subsiguientes van incrementando en uno. A esto se le conoce como **indexación positiva**.

```
cadena = "hola"
cadena[0] #h
cadena[1] #o
cadena[2] #l
cadena[3] #a
```

La **indexación negativa** se refiere a acceder a la posición desde el final de la cadena. El último carácter es -1, y los siguientes van incrementando en 1.

```
cadena = "hola"
cadena[-1] #a
cadena[-2] #l
cadena[-3] #o
cadena[-4] #h
```

En algunas ocasiones, queremos acceder a un fragmento de una cadena. A esto se le conoce como **slicing**, y se emplea con las posiciones: [primero:último:paso]. La última posición no está incluida. Los pasos son opcionales, ya que tienen un valor por defecto de 1. Si se omite el primer valor, se empieza desde el primer carácter por defecto. Si se omite el último valor, se termina en el último carácter. Así, si se pone [:], se incluye todo el contenido.

```
cadena = "hola"
cadena[1:3] #ol
cadena[:2] #hl
```

II.5.3. Métodos de cadenas

- `lower()`: devuelve la cadena con todas las letras en minúscula.
- `upper()`: devuelve la cadena con todas las letras en mayúscula.
- `replace(s1, s2)`: reemplaza la primera cadena con la segunda.
- `count(x)`: cuenta el número de veces que se encuentra x.
- `index(x)`: devuelve la posición del primer x.
- `isalpha()`: devuelve True si todos los caracteres de una cadena son letras.
- `eval()`: evalúa la expresión matemática dentro de la cadena cuando el resultado es numérico.

II.6. Estructuras de datos

En Python hay cuatro estructuras de datos: listas, diccionarios, tuplas y conjuntos o sets.

II.6.1. Listas

La lista es un **conjunto ordenado de objetos**. Pueden ser números, cadenas, booleanos, y se pueden mezclar los distintos tipos. El contenido de una lista se escribe entre corchetes y separando los distintos objetos por comas.

```
lista = [False, 1, 2, "tres"]
```

Las listas guardan algunas similitudes con las cadenas:

- Son indexables y pueden sufrir slicing.
- len muestra la cantidad de elementos que se encuentran en cadena.
- El operador in muestra si la lista contiene algo.
- Se puede concatenar listas mediante + y multiplicarlas con *.
- Se pueden crear listas vacías con [].
- Se puede iterar sobre ellas.

```
codones = ['AAA', 'GCT', 'CCA']
for codon in codones:
    for nucleotido in codon:
        print(nucleotido)
```

Algunos métodos específicos para trabajar con listas son:

- append(x): añade x al final de la lista.
- sort(): ordena la lista de forma alfabética si solo contiene letras y de pequeño a grande si solo contiene números.
- count(x): devuelve el número de veces que aparece x en la lista.
- index(x): devuelve el índice del primer valor de x.
- reverse(): revierte el orden de la lista.
- remove(x): elimina el primer x en la lista.
- pop(i): elimina el item en el índice i y devuelve su valor.
- insert(i,x): inserta x en el índice i de la lista.

II.6.2. Diccionarios

Los diccionarios son **conjuntos no ordenados de parejas de clave y valor**. Las claves deben ser únicas e inmutables de cadenas o números, ya que sirven para acceder al valor que tengan asignado.

```
#Diccionario de dos elementos
year = {'Enero':31, 'Febrero':28}
```

Se puede iterar sobre los diccionarios, pero el elemento que se devuelve es la clave. Para que se devuelva el valor, se debe **indexar por la clave**:

```
for mes in year:
    print(mes) # Imprime las claves
    print(year[mes]) #Imprime los valores
```

II.7. Leer ficheros

Para abrir un fichero en Python, se emplea `open(path/del/fichero)`. Para obtener las líneas que tiene el fichero, se utiliza el método `readlines()`.

```
fichero = open(ejemplo.txt)
contenido = fichero.readlines()
```
