



Gestao de Ficheiros

Filesystem - Gestao de ficheiros.

o Filesystem, "Sistema de Ficheiros", refere-se à organizacao, estrutura de ficheiros num computador. Em UNIX, ficheiros estao organizados em **directorias**. Directorias podem conter ficheiros e outras directorias.

A directoria de base em UNIX é chamada A `root`, raíz (a nao confundir, mas a ideia é a mesma). é a directoria onde vamos se em windows procurarmos "O meu PC > OS (C)".

```
jeff@Nosferatu: ~/example
anaconda3  example      notes      R
Desktop    examples.desktop  Pictures   snap
Documents  igv                  Programs   Templates
Downloads  Music               Public     Videos

jeff@Nosferatu:~$ cd example/
jeff@Nosferatu:~/example$ ls
poem.txt  prose.txt  there_is_nothing_in_here
jeff@Nosferatu:~/example$ man ls
jeff@Nosferatu:~/example$ ls -l
total 8
-rw-rw-r-- 1 jeff jeff 470 Jun 11 01:09 poem.txt
-rw-rw-r-- 1 jeff jeff 0 Jun 11 01:09 prose.txt
drwxrwxr-x 2 jeff jeff 4096 Jun 11 01:09 there_is_nothing_i
n_here
jeff@Nosferatu:~/example$ ls -al
total 20
drwxrwxr-x 3 jeff jeff 4096 Jun 11 01:09 .
drwxr-xr-x 42 jeff jeff 4096 Jun 11 01:09 ..
-rw-rw-r-- 1 jeff jeff 17 Jun 11 01:10 .Im_hiding
-rw-rw-r-- 1 jeff jeff 470 Jun 11 01:09 poem.txt
-rw-rw-r-- 1 jeff jeff 0 Jun 11 01:09 prose.txt
drwxrwxr-x 2 jeff jeff 4096 Jun 11 01:09 there_is_nothing_
in_here
jeff@Nosferatu:~/example$
```

Notem que quando fazemos `ls -la`, temos duas coisas que aparecem : `.` e `..`

- `.` é a directoria corrente, onde nós estamos.
- `..` é a directoria acima (ou anterior).

Se carregarmos `.` ficamos no mesmo sitio, `..` vamos uma directoria acima!

Outros comandos uteis

- `pwd` - a directoria corrente no sistema de ficheiros.
- `cp` "copy" - copiar um ficheiro, damos o nome do ficheiro e o ficheiro de destino.
- `mv` "move" - mover um ficheiro. o primeiro argumento é o ficheiro em questao, o segundo é a directoria de destino → cuidado para nao substituir ficheiros!

- `mkdir` "make directory" - criar directoria.
- `rm` - remover ficheiro.
- `chown` - mudar a propriedade de um ficheiro.

Exercício

1. criar uma directoria `forca_saude`

2. fazer download o ficheiro:

https://github.com/SantosJGND/UNIX_intro/blob/main/UNIX_intro/forca_saude_unix.zip

3. usando a linha de comandos, copiar este ficheiro para a nova directoria

4. correr o comando:

```
unzip forca_saude_unix.zip
```

Ficheiros e conteudos

Alguns comandos úteis

- `cat` "concatenar" - abrir um ficheiro e mostrar na command line.

```
jeff@Nosferatu: ~/example
jeff@Nosferatu:~/example$ cat poem.txt
Mary had a little lamb,
It's fleece was white as snow;
And everywhere that Mary went
The lamb was sure to go.

He followed her to school one day
Which was against the rule;
It made the children laugh and play,
To see a lamb at school.

And so the teacher turned him out,
But still he lingered near;
And waited patiently about
Till Mary did appear

"What makes the lamb love Mary so?"
The eager children cry;
"Why, Mary loves the lamb, you know,"
The teacher did reply.

jeff@Nosferatu:~/example$
```

- `head` - imprimir as 10 primeiras linhas de um ficheiro.
- `tail` - imprimir as 10 ultimas linhas de um ficheiro.
- `more` - imprimir de forma corrida - `Enter` para continuar, `q` para sair.
- `file` - obter informações adicionais sobre um ficheiro.

Exercício

Entrar na pasta UNIX/general e explorar os ficheiros. Usar `more`, `tail` e `file`.

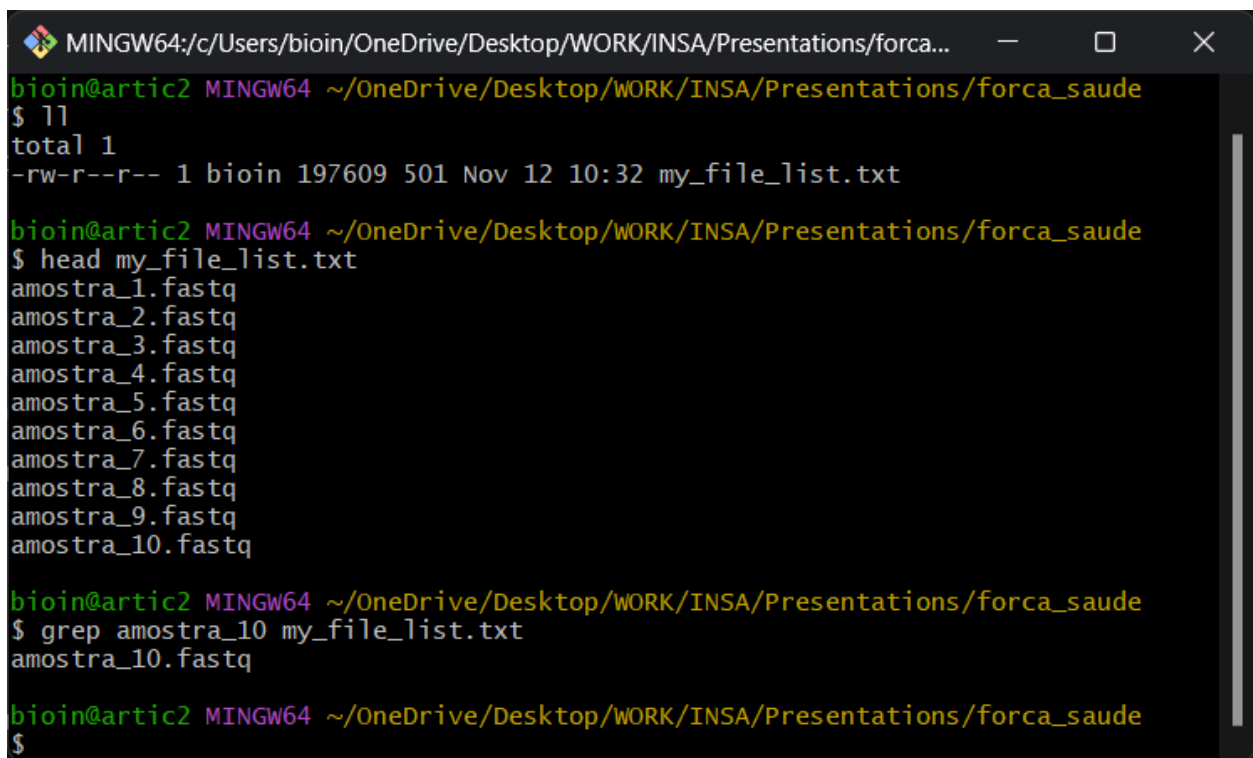
Grep - Encontrar qualquer coisa num ficheiro.

"Grep" é um acrónimo para "*Global Regular Expression Print*", e serve para encontrar um padrão de texto num ficheiro.

Formato:

```
grep [options] [regexp] [filename]
```

Por exemplo, se tem um ficheiro com nomes de amostras, e querem averiguar se uma amostra especifica se encontra no ficheiro:

A screenshot of a terminal window titled 'MINGW64: c:/Users/bioin/OneDrive/Desktop/WORK/INSA/Presentations/forca...'. The prompt is 'bioin@artic2 MINGW64 ~/OneDrive/Desktop/WORK/INSA/Presentations/forca_saude'. The user enters '\$ ll', showing file details for 'my_file_list.txt'. Then they enter '\$ head my_file_list.txt', listing 10 'amostra_*.fastq' files. Finally, they enter '\$ grep amostra_10 my_file_list.txt', which returns 'amostra_10.fastq'.

```
MINGW64: c:/Users/bioin/OneDrive/Desktop/WORK/INSA/Presentations/forca...
bioin@artic2 MINGW64 ~/OneDrive/Desktop/WORK/INSA/Presentations/forca_saude
$ ll
total 1
-rw-r--r-- 1 bioin 197609 501 Nov 12 10:32 my_file_list.txt

bioin@artic2 MINGW64 ~/OneDrive/Desktop/WORK/INSA/Presentations/forca_saude
$ head my_file_list.txt
amostra_1.fastq
amostra_2.fastq
amostra_3.fastq
amostra_4.fastq
amostra_5.fastq
amostra_6.fastq
amostra_7.fastq
amostra_8.fastq
amostra_9.fastq
amostra_10.fastq

bioin@artic2 MINGW64 ~/OneDrive/Desktop/WORK/INSA/Presentations/forca_saude
$ grep amostra_10 my_file_list.txt
amostra_10.fastq

bioin@artic2 MINGW64 ~/OneDrive/Desktop/WORK/INSA/Presentations/forca_saude
$
```

- Se quisermos encontrar um padrao que inclua espacos, usamos "".

Alguns dos argumentos mais usados :

- `grep -v` ignora o padrao em vez de o incluir.
- `grep -i` ignora maiúsculas e minúsculas.

Exercicio

Entrar na pasta `UNIX/general/lyrics/`.

- Qual das estrophes tem a "Nariz"?
- Qual das estrophes não tem a palavra "produtividade"
-

Compressão

Em bioinformática, é comum lidar com ficheiros muito grandes. Uma das soluções é a compressão.

Ficheiros individuais

A maneira mais comum de comprimir ficheiros é `gzip`. este comand cria ficheiros com a extensão `.gz`. Lanca se da forma:

```
gzip [nome do ficheiro]
```

para descomprimir, junta-se o argumento `-d`.

```
gzip -d [nome do ficheiro]
```

```
bioinf@artic2: ~/INSA/presen x + v
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll
total 12
drwxr-xr-x 2 bioinf bioinf 4096 Nov 13 09:43 ./
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../
-rw-r--r-- 1 bioinf bioinf 12 Nov 13 09:43 my_first_file.txt
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ gzip my_first_file.txt
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll
total 12
drwxr-xr-x 2 bioinf bioinf 4096 Nov 13 09:50 ./
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../
-rw-r--r-- 1 bioinf bioinf 50 Nov 13 09:43 my_first_file.txt.gz
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ gzip -d my_first_file.txt.gz
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll
total 12
drwxr-xr-x 2 bioinf bioinf 4096 Nov 13 09:50 ./
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../
-rw-r--r-- 1 bioinf bioinf 12 Nov 13 09:43 my_first_file.txt
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ |
```

- Existem outras maneiras de comprimir ficheiros: `bzip` e `xz`, com as suas próprias extensões.
- Em algumas aplicações bioinformáticas, é frequente usar uma versão alternativa ao `gzip`, chamada `bgzip`. O `bgzip` é similar, cria ficheiros maiores, mas permite um acesso mais rápido ao interior.

Às vezes é necessário ler um ficheiro comprimido (por exemplo para averiguar antes de trabalhar).

- Em linux, existem comandos `zmore` e `zcat`. Estes são equivalentes aos comandos vistos acima.

Comprimir vários ficheiros.

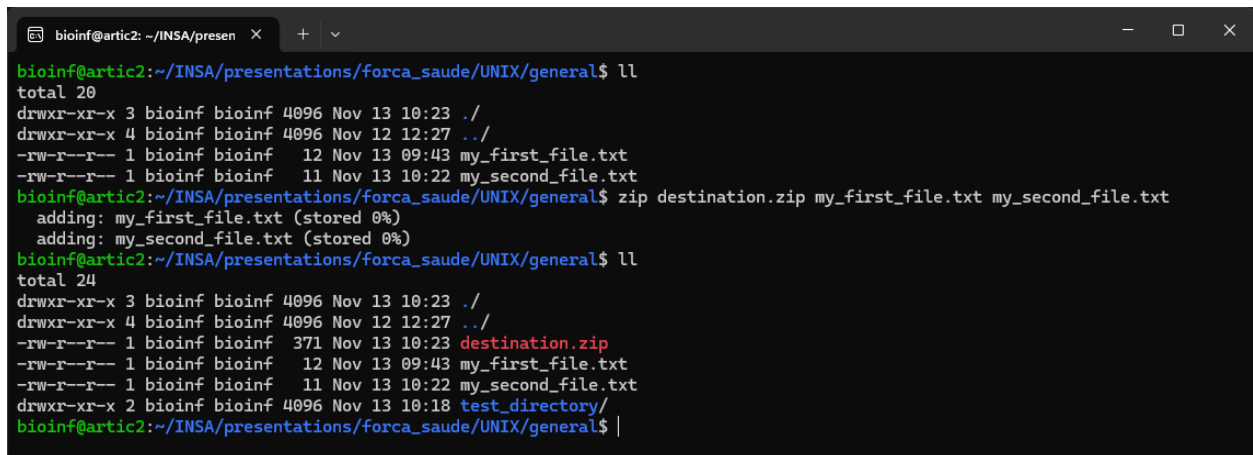
Muitas vezes queremos comprimir não só um ficheiro mas vários, comprimir uma directoria. Para isto em UNIX temos acesso ao `zip` e `tar`. ambos frequentemente utilizados. comando `zip` é o mesmo que em Windows.

- frequentemente, eu uso `zip` se vou partilhar com um colega ou abrir no Windows

- Uso `tar` quando o ficheiro nao vai sair do UNIX.

Comprimir com `zip`

```
zip [nome do ficheiro final - .zip] [ficheiro1 ficheiro2 etc..]  
zip -r [nome do ficheiro final - .zip] [directoria]
```



```
bioinf@artic2: ~/INSA/presen x + v  
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll  
total 20  
drwxr-xr-x 3 bioinf bioinf 4096 Nov 13 10:23 ./  
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../  
-rw-r--r-- 1 bioinf bioinf 12 Nov 13 09:43 my_first_file.txt  
-rw-r--r-- 1 bioinf bioinf 11 Nov 13 10:22 my_second_file.txt  
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ zip destination.zip my_first_file.txt my_second_file.txt  
  adding: my_first_file.txt (stored 0%)  
  adding: my_second_file.txt (stored 0%)  
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll  
total 24  
drwxr-xr-x 3 bioinf bioinf 4096 Nov 13 10:23 ./  
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../  
-rw-r--r-- 1 bioinf bioinf 371 Nov 13 10:23 destination.zip  
-rw-r--r-- 1 bioinf bioinf 12 Nov 13 09:43 my_first_file.txt  
-rw-r--r-- 1 bioinf bioinf 11 Nov 13 10:22 my_second_file.txt  
drwxr-xr-x 2 bioinf bioinf 4096 Nov 13 10:18 test_directory/  
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ |
```

Para descomprimir, utilizamos o comando `unzip`


```
bioinf@artic2: ~/INSA/presen x + v
adding: test_directory/my_first_file.txt (stored 0%)
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll
total 28
drwxr-xr-x 3 bioinf bioinf 4096 Nov 13 10:25 ./
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../
-rw-r--r-- 1 bioinf bioinf 371 Nov 13 10:23 destination.zip
-rw-r--r-- 1 bioinf bioinf 12 Nov 13 09:43 my_first_file.txt
-rw-r--r-- 1 bioinf bioinf 11 Nov 13 10:22 my_second_file.txt
-rw-r--r-- 1 bioinf bioinf 384 Nov 13 10:25 test_compress.zip
drwxr-xr-x 2 bioinf bioinf 4096 Nov 13 10:18 test_directory/
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ unzip test_compress.zip
Archive: test_compress.zip
replace test_directory/my_first_file.txt? [y]es, [n]o, [A]ll, [N]one, [r]ename: yes
extracting: test_directory/my_first_file.txt
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll
total 28
drwxr-xr-x 3 bioinf bioinf 4096 Nov 13 10:25 ./
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../
-rw-r--r-- 1 bioinf bioinf 371 Nov 13 10:23 destination.zip
-rw-r--r-- 1 bioinf bioinf 12 Nov 13 09:43 my_first_file.txt
-rw-r--r-- 1 bioinf bioinf 11 Nov 13 10:22 my_second_file.txt
-rw-r--r-- 1 bioinf bioinf 384 Nov 13 10:25 test_compress.zip
drwxr-xr-x 2 bioinf bioinf 4096 Nov 13 10:25 test_directory/
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ |
```

ver mais exemplos em <https://www.geeksforgeeks.org/zip-command-in-linux-with-examples/>

Comprimir com **tar**

O formato para comprimir com tar é o mesmo que com o zip: **primeiro o ficheiro de saída, depois o a directoria que queremos comprimir!**

```
tar [options] [output.tar] [input/]
```

tar precisa sempre de argumentos. vamos sempre usar um destes dois:

- **c** : "Criar" para comprimir
- **x** : "eXtrair" para descomprimir um arquivo.

Em qualquer caso vamos querer criar um ficheiro de saída, por isso vamos usar a opção:

- `f` : Creates an archive with given filename.
- Caso estejamos a descomprimir não é preciso especificar um nome de ficheiro de saída.

Finalmente, se quisermos usar (quer seja para comprimir ou descomprimir) uma camada adicional de compressão, podemos utilizar o argumento:

- `z` : Comprimir usando `gzip`. Extensão resultante `.tar.gz`.

```
bioinf@artic2: ~/INSA/presen X + v
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll
total 28
drwxr-xr-x 3 bioinf bioinf 4096 Nov 13 10:48 ./
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../
-rw-r--r-- 1 bioinf bioinf 371 Nov 13 10:23 destination.zip
-rw-r--r-- 1 bioinf bioinf 12 Nov 13 09:43 my_first_file.txt
-rw-r--r-- 1 bioinf bioinf 11 Nov 13 10:22 my_second_file.txt
-rw-r--r-- 1 bioinf bioinf 384 Nov 13 10:25 test_compress.zip
drwxr-xr-x 2 bioinf bioinf 4096 Nov 13 10:25 test_directory/
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ tar -cf my_tar.tar test_directory
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ tar -czf my_tar.tar.gz test_directory/
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ ll
total 44
drwxr-xr-x 3 bioinf bioinf 4096 Nov 13 10:49 ./
drwxr-xr-x 4 bioinf bioinf 4096 Nov 12 12:27 ../
-rw-r--r-- 1 bioinf bioinf 371 Nov 13 10:23 destination.zip
-rw-r--r-- 1 bioinf bioinf 12 Nov 13 09:43 my_first_file.txt
-rw-r--r-- 1 bioinf bioinf 11 Nov 13 10:22 my_second_file.txt
-rw-r--r-- 1 bioinf bioinf 10240 Nov 13 10:49 my_tar.tar
-rw-r--r-- 1 bioinf bioinf 188 Nov 13 10:49 my_tar.tar.gz
-rw-r--r-- 1 bioinf bioinf 384 Nov 13 10:25 test_compress.zip
drwxr-xr-x 2 bioinf bioinf 4096 Nov 13 10:25 test_directory/
bioinf@artic2:~/INSA/presentations/forca_saude/UNIX/general$ |
```

Exercicio

Criar uma pasta com os ficheiros my_first_file.txt e my_second_file.txt.

comprimir a pasta usando `zip` e `tar`