

Línea de comandos Base de datos

Adolfo De Unánue

20 de noviembre de 2019

Índice general

Problema 1	1
Problema 2	2
Problema 3	2
Problema 4	2

Algunos comandos útiles

Problema 1

Crea un archivo llamado problema-1.sh

En una línea de bash:

1. Descarga el libro de *The Time Machine* de **H. G. Wells**,
2. Remover el encabezado (desde la primera línea hasta la línea que contiene ***** START OF THIS PROJECT GUTENBERG EBOOK THE TIME MACHINE *****
3. Remover la licencia (desde ***** END OF THIS PROJECT GUTENBERG EBOOK THE TIME MACHINE ***** hasta el final del archivo)
4. Convertirlo a minúsculas
5. Extraer las palabras
6. Ordenarlas
7. Eliminar duplicados y contarlos
8. Ordenar de mayor a menor
9. Mostrar el *top 10* de palabras con su frecuencia

HINT: En sed el final del archivo se identifica con el caracter \$. La primera línea la puedes identificar con el número.

Analizando datos: Comandos avanzados

Problema 2

Usando el archivo `UFO-Nov-Dic-2014.psv`

- ¿Cuántos avistamientos por estado hay? (Guárdalo en el archivo `problema-2a.sh`)
- ¿Cuántos avistamientos no tienen forma de esferoide? (Guárdalo en el archivo `problema-3a.sh`)

Problema 3

Guárdalo en el archivo `problema-3.sh`

Usando el archivo `UFO-Nov-Dic-2014.psv`.

Describe estadísticamente (max, min, mean) los tiempos de duración de la observación

Bash programming

La mayor parte del tiempo usaremos el `shell`, para hacer pequeños *scripts*, pero existen ocasiones en las cuales es necesario tratar al `shell` como un lenguaje de programación¹

Problema 4

Crearemos un *script* que analice todos los datos (de 1990 a la fecha) de la página web de UFOs.

El cascarón de este archivo se encuentra en `scripts/ufo-analysis.sh`. Úsalo como punto de partida.

Agrega las funciones:

1. `clean_data` Convierte a minúsculas, separador a `|`
2. `concat_data` Concatena todos los meses en un solo archivo.
3. `calculate_stats` Calcula los conteos por estado, color, forma, año, mes y hora
4. Crea un *script* de python o R que genere las gráficas de estos conteos, guarda estos archivos.

¹En este punto es bueno preguntarse si no deberías hacerlo mejor en otro lenguaje de programación, como python.