

PEC 2: Primeros scripts

El objetivo de estos ejercicios consiste en poner en práctica la gestión básica del sistema operativo GNU/Linux desde la línea de comandos, como por ejemplo la creación y modificación de directorios y ficheros, gestión de permisos, variables de entorno, etc., entre otras operaciones.

Para la elaboración de esta PEC es necesario **utilizar el usuario creado en la PEC 1** (es posible utilizar cualquier otra máquina). Deben demostrarse las operaciones a través de **capturas de pantalla** con el usuario creado en la PEC 1, de otro modo se considerará una PEC plagiada.

Esta PEC está enfocada en demostrar que el alumno ha adquirido la capacidad de trabajar en la línea de comandos de GNU/Linux y con Bash. Todavía **no se puede utilizar grep ni sed ni awk**, aunque sí **jq** o **xmlstarlet**.

Formato de entrega: fichero en formato zip que contenga el fichero de texto y las capturas de pantalla (**sin recortar**, donde **se vea claramente el usuario de la UOC creado en la PEC 1**, y un **máximo de 3 capturas por ejercicio**) con la nomenclatura siguiente:

- *Capturas de pantalla:* en los ejercicios donde se indique explícitamente, se usará la nomenclatura **PEC2_NumeroEjercicio_NumeroCaptura.png**, por ejemplo, si necesitáis tres capturas para el primer ejercicio: **PEC2_1_1.png**, **PEC2_1_2.png**, **PEC2_1_3.png**
- *Fichero de texto:* fichero en texto plano llamado **PEC2_Respuestas.txt**

Enunciado

El listado de tareas a realizar es el siguiente:

Configuración del DNS del sistema (1.5 puntos)

Escoger cuatro servidores DNS públicos del país donde estéis conectados (por ejemplo de <https://public-dns.info/>, que sean 100% fiables).

- Comprobar con el comando `dig` los siguientes hosts de Internet para saber cuál de los servidores DNS tiene menos tiempo de latencia.
 - `datascience.recursos.uoc.edu`
 - `www.rediris.es`
 - `www.kaggle.com`

Adjuntar capturas con la ejecución de los comandos.

- Mostrar con *capturas de pantalla* los pasos para utilizar dicho servidor en vuestro sistema operativo.

Análisis básico de un conjunto de datos mediante la línea de comandos (2 puntos)

Descargar el conjunto de datos llamado Indian Liver Patient Dataset (<https://www.openml.org/d/1480>) y responder a las preguntas siguientes usando los comandos del sistema operativo adecuados:

- ¿Con qué comando se puede saber el tamaño del fichero?
- ¿Cuántas líneas contiene?
- ¿Con qué comando podemos mostrar “solamente” las primeras siete líneas del fichero?
- ¿Cómo podemos mostrar “solamente” las líneas 200-220?
- ¿Con qué comando podemos mostrar solamente la segunda columna?
- ¿Y solamente la tercera columna pero con los valores ordenados y sin repeticiones?
- Descargad todos los ficheros el conjunto de datos Indian Liver Patient Dataset y guardadlos en una carpeta en vuestro \$HOME llamada liver. ¿Con que comando, estando en nuestra carpeta \$HOME, podemos averiguar el MIME TYPE y la Codificación de caracteres utilizada de cada uno de los ficheros?

Escribid cada comando en el documento y adjuntad una captura con su ejecución (una captura puede servir para más de un comando).

Creación de un informe en Bash (4 puntos)

A partir del conjunto de datos Indian Liver Patient Dataset (<https://www.openml.org/d/1480>):

- Con los comandos `cat`, `cut`, `sort`, `uniq`, `tr`, `wc`, `diff` (no hace falta usarlos todos) obtener todos los valores y su recuento, ordenados de menor a mayor (según el número de apariciones), de la columna "Albumin". Escribid el comando en el documento y adjuntad una captura con su ejecución (*1 punto*).
- Crear un script `script3.sh` en Bash que muestre sólo la edad, el sexo, el número de "Albumin" y el total de proteínas de los registros con un valor de exacto de "Albumin" de 3, separados por sexo. Es decir, primero debe mostrar a todas las mujeres y luego a todos los hombres. Adjuntad el script y una captura con su ejecución. **Recordar que no es posible usar `grep` ni `sed` ni `awk` (3 puntos)**.

La salida debe ser como en el fragmento de la captura siguiente:

```

Albumin = 3
Age Sex Albumin Proteins
*****

Women
-----
38,Female,3, 5.6
38,Female,3, 5.6
32,Female,3, 6
31,Female,3, 6
19,Female,3, 5.5
38,Female,3, 7
20,Female,3, 6.1
32,Female,3, 7
42,Female,3, 6.7

Men
-----
72,Male,3, 7.4
60,Male,3, 6.3
33,Male,3, 5.4
60,Male,3, 6.8
60,Male,3, 7.4
60,Male,3, 7
72,Male,3, 6.2
38,Male,3, 6
49,Male,3, 5.7
41,Male,3, 6.1
46,Male,3, 6
33,Male,3, 6.2
33,Male,3, 7
33,Male,3, 6.5
38,Male,3, 6
44,Male,3, 5.6

```

*Captura de pantalla de la
salida del script en Bash del
ejercicio 3b*

Escribimos a continuación el contenido de cada campo:

V1. Age of the patient. Any patient whose age exceeded 89 is listed as being of age "90".

V2. Gender of the patient

V3. Total Bilirubin

V4. Direct Bilirubin

V5. Alkphos Alkaline Phosphatase

V6. Sgpt Alanine Aminotransferase

V7. Sgot Aspartate Aminotransferase

V8. Total Proteins

V9. Albumin

V10. A/G Ratio Albumin and Globulin Ratio

Class. Liver patient / not patient

Manipulación de ficheros JSON en la línea de comandos con jq (2 puntos)

Actualmente JSON es el formato de facto para la distribución de la información. En la línea de comandos, el programa jq es la herramienta idónea para su tratamiento.

A partir del documento JSON siguiente sobre las elecciones al parlamento europeo (<https://www.europarl.europa.eu/election-results-2019/data-sheets/json/2019-2024/election-results/parties.json>) de 2019, obtener, usando jq:

1. El listado de todos los códigos de los países presentes en el documento. Escribid el comando en el documento y adjuntad una captura con su ejecución.
2. El listado de todos los acrónimos de los partidos de España presentes en el documento. Escribid el comando en el documento y adjuntad una captura con su ejecución.

Enlaces de interés

<https://public-dns.info/>

<https://www.openml.org/d/1480>

<https://www.europarl.europa.eu/election-results-2019/data-sheets/json/2019-2024/election-results/parties.json>

Formato de entrega

Entregar un único fichero en formato zip que contenga:

- Capturas de pantalla de los ejercicios donde se solicite, asegurándose que:
 - Están sin recortar y se lee el resultado completo de la ejecución del comando.
 - Se incluye como máximo un total de 3 capturas de pantalla por ejercicio.
 - Aparezca siempre claramente el usuario de la UOC creado en la PEC 1.
 - *Es necesario respetar la nomenclatura indicada con anterioridad.*

- Un fichero de texto plano llamado **PEC2_Respuestas.txt** que contenga las respuestas con las instrucciones.

Puntuación

Ejercicio 1: 1.5 *puntos*.

Ejercicio 2: 2 *puntos*.

Ejercicio 3: 4 *puntos*.

Ejercicio 4: 2 *puntos*.

Valoración global de la PEC 2: 0.5 *puntos* (respetan los aspectos formales de entrega descritos en el enunciado, concisión, corrección ortográfica, etc.)

NOTA: No se aceptará en ningún caso la entrega de la PEC2 después de la fecha máxima de entrega (31-10-2021 a las 23:59:59). *Si por alguna razón pensáis que no vais a poder entregar a tiempo, consultadlo con vuestro profesor siempre con anterioridad.*

Las capturas de pantalla donde no aparezca el nombre de usuario creado en la PEC anterior invalidarán el ejercicio completamente (no recibirán puntuación alguna).