



Fuentes de Datos
Reporte Proyecto 1
EquipoB
Ainé Fernández Gregorio
Victor Emiliano Ortiz Mayen

El programa del proyecto consiste en cinco scripts de bash `set_env`, `get_bookshelves_list`, `get_bookshelves`, `get_bookshelves_titles` y `get_books`, cada uno cumple una función distinta que ayuda a completar el objetivo del proyecto.

set_env

El primer script le da al usuario la posibilidad de construir un nuevo directorio para el proyecto si así lo desea o continuar con uno que él quiera. Si contesta que sí quiere crear un nuevo directorio para el proyecto, se tiene que checar si el directorio o path insertado es válido y si está vacío. Para esto, no tuvimos muchas dificultades, ya que preguntamos si el directorio existe, si el directorio no existe preguntamos por otro hasta que nos de un directorio existente con un `while`. Luego usamos `ls` con `-A` para checar si el directorio está vacío o no vacío.

get_bookshelves_list

En este script se descarga la lista de los bookshelves del sitio web de Gutenberg. Para este script era necesario un `curl` seguido de un `sed` para extraer la información correcta de la página web, pero el `sed` original nos causó problemas y no pudimos lograr que funcionara así que terminamos modificando la estructura del `sed` para lograr el output deseado.

get_bookshelves

El script le da al usuario una lista de bookshelves para descargar y le pide que seleccione uno, si el que seleccionó es un bookshelf válido entonces el script descarga la lista de títulos en ese bookshelf. Para esto necesitamos el url del bookshelf de la id que proporcionó el usuario, otra vez tuvimos dificultades y decidimos que era mejor poner el url completo y agregarle el id que seleccionó el usuario hasta el final y así el script funcionó.

get_bookshelves_titles

El script checa si la id que seleccionó el usuario pertenece a un nuevo bookshelf, descarga los títulos y los guarda en un archivo. Aquí también modificamos el `sed` para que fuera al url, trajera los títulos y finalmente los guardara en un archivo con el nombre de `titles.txt`

get_books

Este script descarga `n` libros al azar del bookshelf seleccionado. En `books_rdm` hicimos que nos diera un número aleatorio entre 0 y `n_books-1`, y al resultado se le suma 1 para que este entre 1 y `n_books`, luego con un `sed` buscamos ese número de línea en el archivo de `titles` para obtener el id del libro y con ese id después construimos el url para descargar el texto de ese libro, todo esto dentro de un `for` para que se repita `n` veces.

Conclusiones

Lo que aprendimos de este proyecto es que nos dimos cuenta de la utilidad y las ventajas que tiene bash y la terminal para la organización de directorios y creación de archivos, ya que es más fácil hacerlo mediante bash que mediante otros lenguajes; aunque es un lenguaje con muchas claves y no es tan intuitivo como otros lenguajes de programación.