

Cuestionario

1) Evaluar limitaciones que pueda tener el sitio a la hora de acceder a los datos.

El sitio consta de una cantidad muy amplia de libros divididos en varias categorías, creemos que esto puede ser una limitación porque no podríamos obtener toda la información en solo un día. Por lo que pensamos en achicar nuestro alcance a solo la categoría Ficción y a solo las dos primeras páginas de los libros subidos más recientemente.

2) ¿Cuándo se va a *scrapear*?

Se va a *scrapear* diariamente, alrededor de las 9 AM para que podamos monitorearlo en nuestro horario laboral.

3) ¿Qué preprocesamiento hará falta hacer?

No es necesario un preprocesamiento de los datos extraídos, debido a que con el uso correcto del XPath ya podemos obtenerlos de la forma que los vamos a plasmar en la base de datos.

4) ¿Dónde y en qué formato guardaremos los datos?

Los datos se guardarán en una base de datos, en una tabla que nos permita poder plasmar y ordenar la información *scrapeada*.

5) ¿Cómo se actualizará el dashboard?

El dashboard se actualizará diariamente con los últimos datos encontrados, mostrando así algunos gráficos que representen ciertos cambios y estadísticas que se hayan realizado en la página en ese tiempo.

6) ¿Qué podemos hacer si falla el proceso?

Si falla el proceso se puede deber a que la página posee un límite de pedidos que se le pueden hacer. En ese caso pasaremos al otro dominio de Libgen que es

idéntico en todo sentido solo con otra URL. En el peor de los casos, donde fallen ambas URLs por el mismo error, pasaremos a implementar un timeout para luego volver a solicitar la información.