**Desafios**

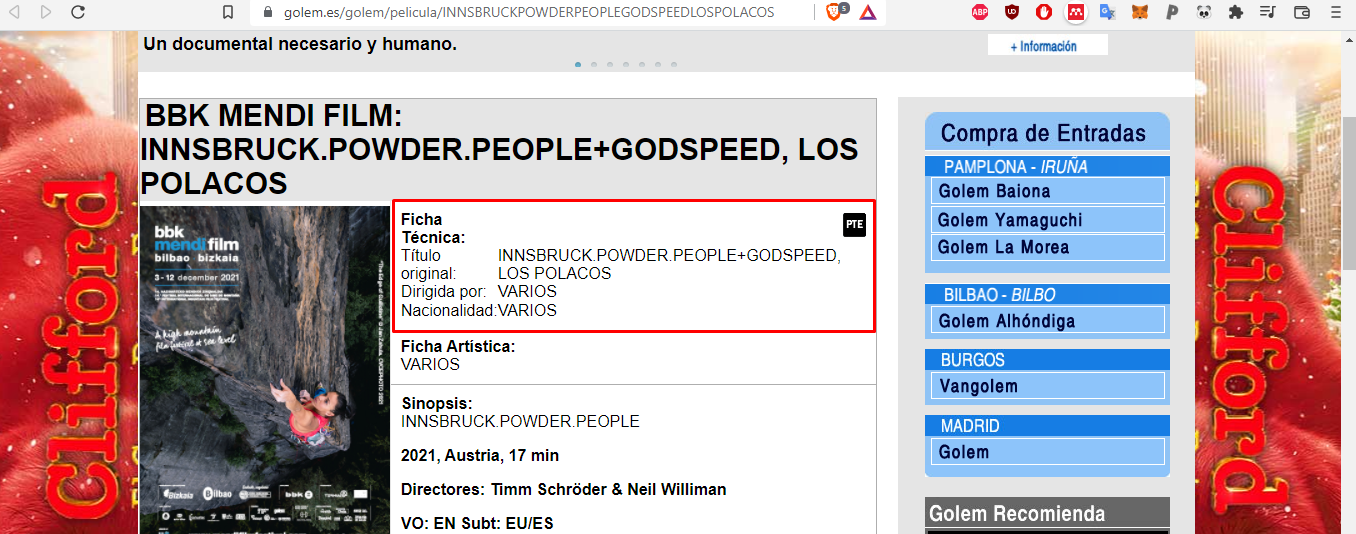
Establecer los parámetros que son iguales y diferentes de ambas web:

|  |  |
| --- | --- |
| **Golem** | **Brodway** |
| Titulo original | Titulo original |
| Director | Director |
| Duración | Duración |
| Nacionalidad | Nacionalidad |
| Reparto | Reparto |
| Sinopsis | Sinopsis |
| Imagen portada | Imagen portada |
|  | Sesiones |
|  | Año |
|  | Guión |
|  | Trailer |

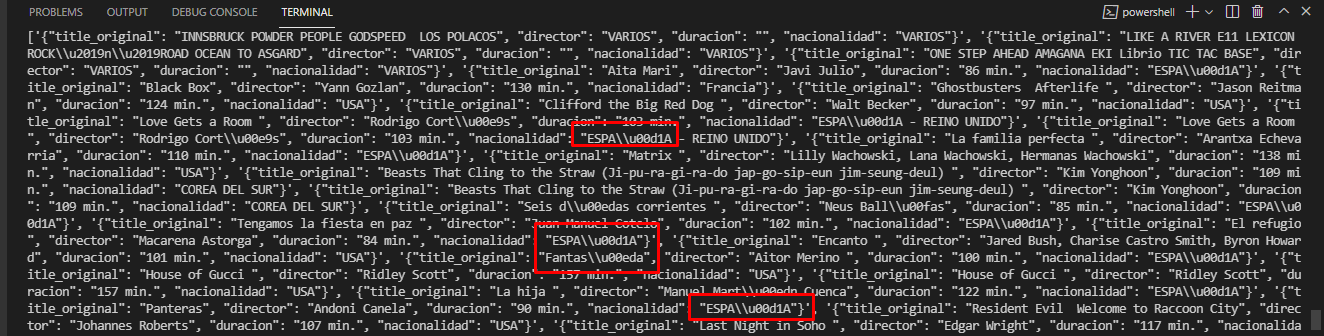
Armar los json con el contenido de los títulos, algunos tenían comas o caracteres especiales que dificultaban la tarea de convertirlos en json

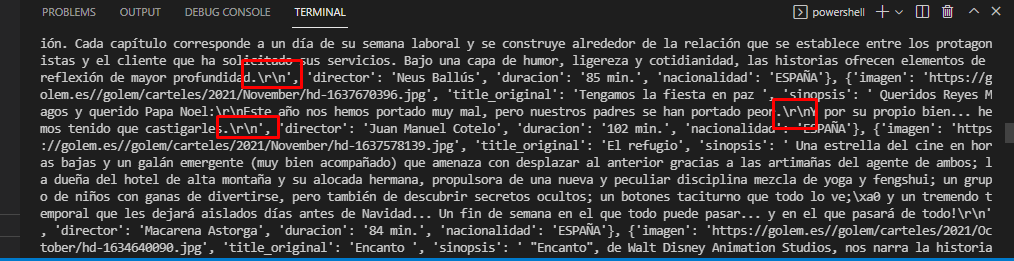


Algunas películas no proporcionaban datos como la duración o por cual directo fue dirigido y había que validar datos inexistentes.



Problemas con palabras en Unicode, había que convertirlos en utf-8



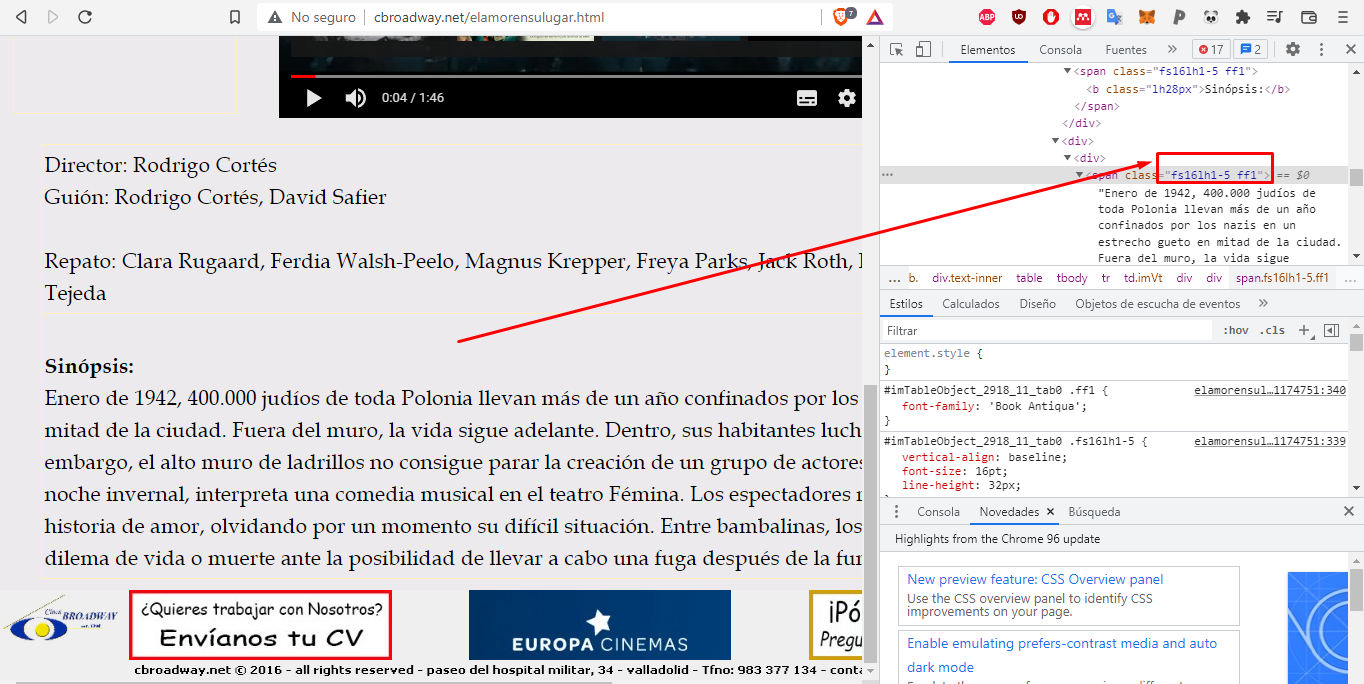


* No hay una forma óptima para validar que las películas no se repitan, porque no tienen un id para identificarlos y compararlos por el título de la película es muy ineficiente debido a que pueden existir títulos de películas que difieran en al menos un carácter o estén en otro idioma lo que provocaría que se repitan las películas.
* Se debe crear dos scrapper por cada web, debido a que cada web está programada de diferente manera.
* Los scrappers dejarían de funcionar si el programador del sitio web cambia el html de la web.

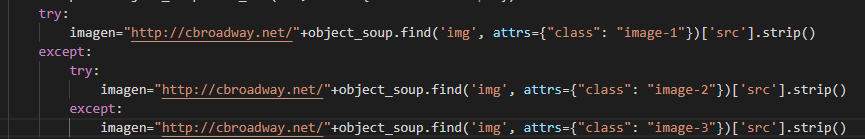
En ocasiones la librería daba error de conexión



No logré obtener la sinopsis de la web Brodway, debido a que la clase fs16lh1-5 ff1 se repetía en muchas etiquetas html y no había manera de encontrarlo.



En la web de brodway, para obtener la imagen de portada de la película, ciertas páginas tenían diferentes clases para la etiqueta img, dificultando asi la obtención de la imagen.



En la web de brodway, no se pudo obtener la url del trailer debido a que el iframe se carga de manera dinámica es decir asincrónica.

