

INSTRUCTIVO DESCARGAREVISTASJS

Este código diseñado en JavaScript es un Scraper que recopila, extrae y descarga las imágenes de las revistas solicitadas en Global News.

Requisitos previos instalados:

Para el correcto funcionamiento de este Scraper es necesario tener previamente instaladas las siguientes herramientas y librerías:

- **node.js y npm:** descarga conjunta desde el sitio oficial: <https://nodejs.org/>
- **express:** npm install express
- **puppeteer:** npm install puppeteer
- **axios:** npm install axios
- **sharp:** npm install sharp
- **pdf-lib:** npm install pdf-lib
- **socket.io:** npm install socket.io

Alcance del Scraper

Las revistas que pueden ser descargadas con este código principalmente son las requeridas hasta la fecha, interviniendo en más de 15 medios descargables. El código aún se encuentra en desarrollo, por lo que se espera que, en un futuro, se encuentren incluidas muchas más revistas. Para el caso actual, estos son algunos de los medios disponibles:

- Semana
- El Diario
- ADN
- El Espectador
- Rolling Stone
- Style America
- Bogotá Nocturna
- Agencia de Viajes
- La Tribuna
- Portavoz
- Etc...

Configuración del Scraper:

La ejecución es bastante sencilla, dónde por medio de un doble clic al archivo **DescargaRevistasJS.bat** ubicado en el escritorio del pc podemos ejecutar el código. Si este bat por alguna razón falla, otra alternativa es dirigirse a la carpeta raíz ubicada normalmente en documentos, ejecutar un cmd ubicado en la ruta del código y correr el siguiente comando: **npm run start**.

La versión actual a la fecha se encuentra ejecutándose en la IP222, en el puerto 3000, por lo que fácilmente se puede acceder a la interfaz de usuario desde el navegador con el host: **192.168.1.222:3000**. Tener en cuenta que el acceso está limitado para la red dentro de la empresa, no se puede acceder desde otra que no sea esta.

Páginas soportadas:

Los medios pueden estar alojados en distintas páginas web, por lo que muchos de ellos comparten algunas páginas y facilitan la descargas de estos mismo, sin embargo, si las revistas no se encuentran alojados dentro de las páginas disponibles, **NO** será posible la descarga. Para este caso en particular, se debe realizar la actualización del código con la nueva página.

Las páginas disponibles a la fecha son:

- Flipsnack
- Camacol
- Calameo
- Isuu
- Tabloide
- Yumpu
- Google Drive
- Fliphtml5
- Anyflip
- Heyzine
- Ladevi
- Semana
- Eldiario
- Anyflip
- Clarin
- Rollingstone

Uso del Scraper:

Para hacer uso en la descarga de revistas, simplemente se debe copiar el link, pegarlo en el área de texto y darle enter o clic en el botón descargar, comenzará la descarga simplemente es esperar y al finalizar el proceso se obtendrá el PDF de la revista.



Recomendaciones de uso:

Para una correcta ejecución del código y descarga de los archivos es necesario tener en cuenta ciertas recomendaciones:

1. La revista debe estar en la página 1.
2. Copiar el link directamente desde el navegador, ya que, de esta manera, garantizamos que esté funcionando y tenga el HTTP o HTTPS incluido.
3. Garantizar que el link accede directamente a la página.
4. La ruta de respaldo se encuentra en: <\\192.168.1.153\Area-Tecnologia\DESCARGAREVISTASJS>.
5. Verificar el orden de las páginas, si hay que modificarlas, se recomienda el uso de la página [iLovePDF](#).
6. Las páginas a ordenar son:
 - Fliphtml5
 - Anyflip
 - Heyzine
 - Ladevi
 - Semana
 - Eldiario
 - Yumpu (encriptadas)
 - Clarin
7. Anyflip tiene una página de publicidad que no la toma, buscarla, tomar screenshot y meterla manualmente.
8. Yumpu encriptado tiene agrega una página adicional de todas las páginas de la revista en miniatura, se debe eliminar esta página.

Errores en la ejecución:

El código se encuentra diseñado a prueba de errores, por lo que operacionalmente no debería de intervenir algún error, sin embargo, los posibles errores pueden suceder por factores como:

1. Conexión inestable o poca velocidad de internet.
2. Página web no disponible.
3. Actualización de la página web.
4. Link mal ingresado.

En cualquier caso, para corregir el scraper luego de un error, es necesario volver a ejecutar el código, posteriormente el error será corregido dentro del código.