SCRAPING WEB

Scraping con Webscrapper

¿Qué es el scraping?

El Web scraping nos permite extraer información de la web de forma estructurada para poder manejarla como datos y tratarla para su análisis y visualización.

Existen diferentes técnicas para hacer scraping. Todas ellas utilizan algún software o aplicación, bien directamente en el navegador bien usando algún lenguaje con Python que, mediante script, captura la información disponible en una web.

El scraping usa técnicas automatizadas para extraer la información publicada en la web en una escala de automatización que va desde las acciones manuales ("copiar y pegar") hasta sistemas de reconocimiento de información semántica, pasando por aplicaciones de web scraping, parsers HTML, algoritmos de minería de datos, etc.

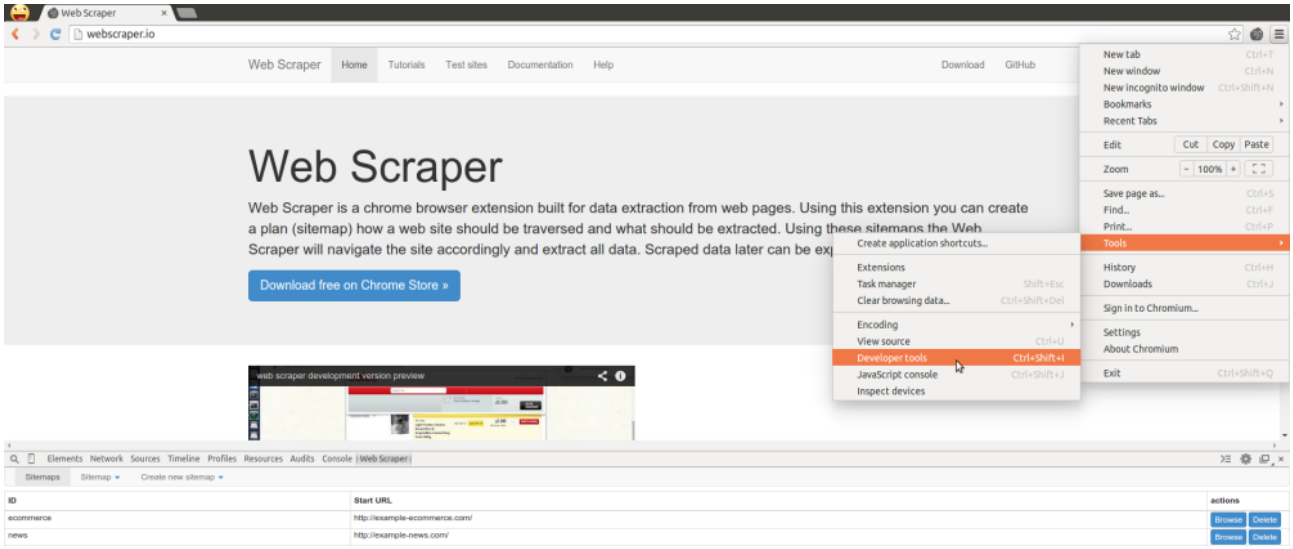
No todas las web permiten ser "scrapedas". En ocasiones, los administradores de los sitios web limitan esta posibilidad mediante técnicas como bloqueos de IPs, limitar accesos con el fichero robots.txt, etc.

Qué es Web scraper y cómo usarlo

Web Scraper es una extensión para navegador que resulta muy útil para raspar contenido de una web ya que puede ejecutarse directamente desde el navegador web sin necesidad de instalar otro software adicional.

Para utilizarlo debemos:

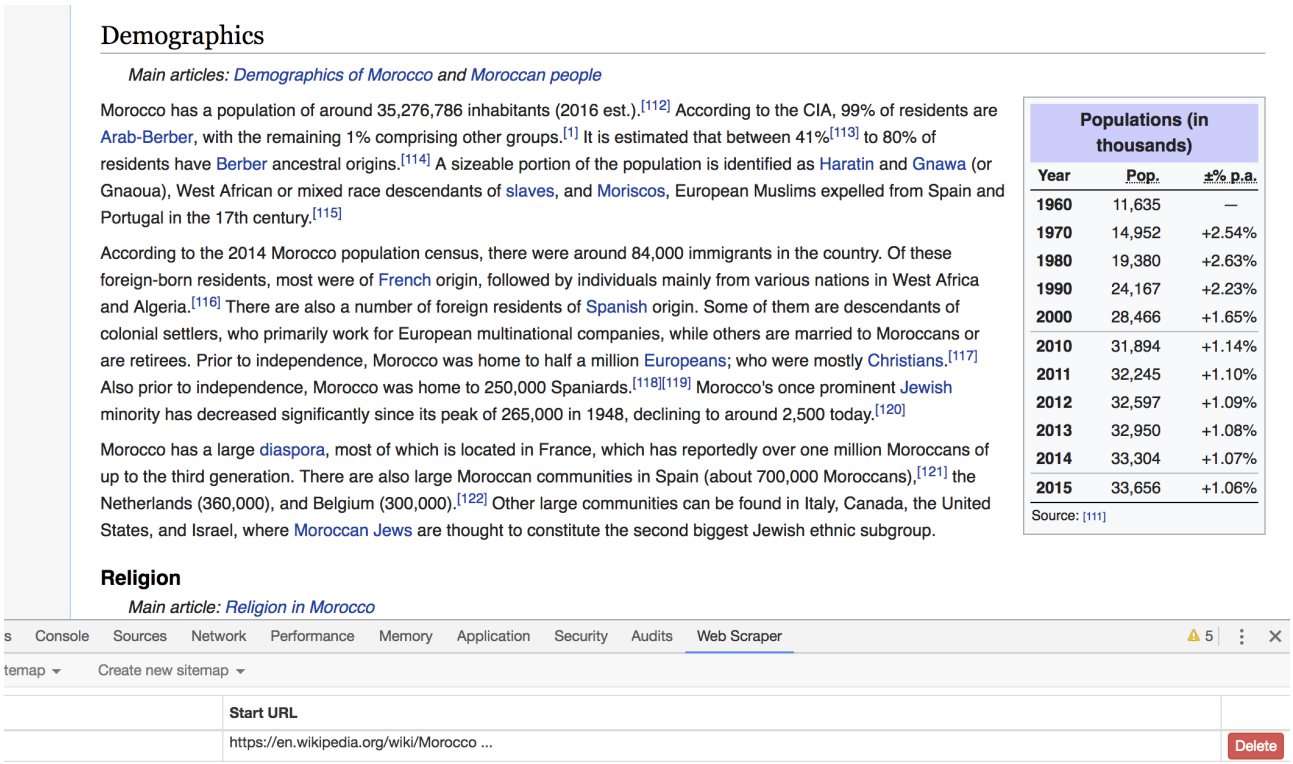
1. Descargarnos el plugin para Chrome e instalarlo.
2. Para acceder al plugin debemos abrir en Chrome la consola de herramientas de desarrolladores. Para acceder al DevTools debes seccionar el menú de la esquina superior derecha del explorador, luego seleccionando More Tools > Develope Tools. Alternativamente puedes usar Ctrl+Shift+I o Cmd+Opt+I en Mac, y en Windows presionar la tecla F12



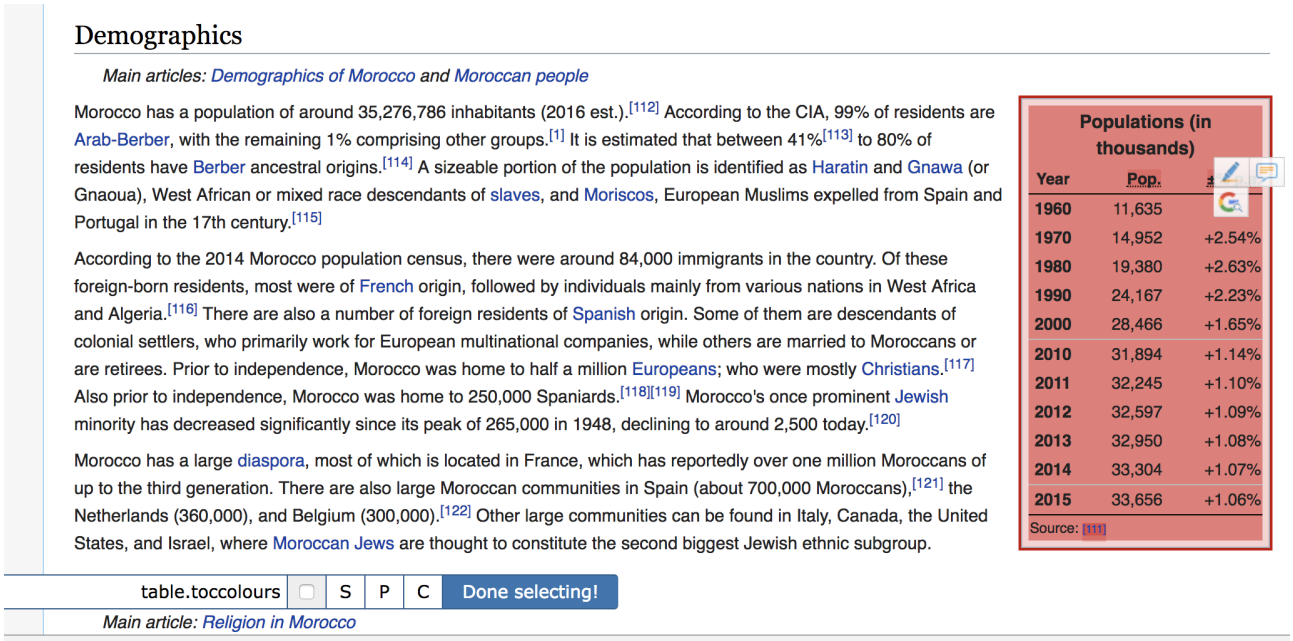
Acceder a la consola de herramientas de desarrolladores mediante menú

Ejemplo de uso: Capturar datos de la Wikipedia

Veamos un caso práctico de uso. Pongamos por caso que queremos descargar un conjunto de datos de una entrada de la wikipedia de un país. Habitualmente, encontramos información en tablas como, por ejemplo, de la población. Vamos a ver cómo extraer una tabla.



Web Scraper en DevTools



Demographics

Main articles: Demographics of Morocco and Moroccan people

Morocco has a population of around 35,276,786 inhabitants (2016 est.).[112] According to the CIA, 99% of residents are Arab-Berber, with the remaining 1% comprising other groups.[1] It is estimated that between 41%[113] to 80% of residents have Berber ancestral origins.[114] A sizeable portion of the population is identified as Haratin and Gnawa (or Gnaoua), West African or mixed race descendants of slaves, and Moriscos, European Muslims expelled from Spain and Portugal in the 17th century.[115]

According to the 2014 Morocco population census, there were around 84,000 immigrants in the country. Of these foreign-born residents, most were of French origin, followed by individuals mainly from various nations in West Africa and Algeria.[116] There are also a number of foreign residents of Spanish origin. Some of them are descendants of colonial settlers, who primarily work for European multinational companies, while others are married to Moroccans or are retirees. Prior to independence, Morocco was home to half a million Europeans; who were mostly Christians.[117] Also prior to independence, Morocco was home to 250,000 Spaniards.[118][119] Morocco's once prominent Jewish minority has decreased significantly since its peak of 265,000 in 1948, declining to around 2,500 today.[120]

Morocco has a large diaspora, most of which is located in France, which has reportedly over one million Moroccans of up to the third generation. There are also large Moroccan communities in Spain (about 700,000 Moroccans),[121] the Netherlands (360,000), and Belgium (300,000).[122] Other large communities can be found in Italy, Canada, the United States, and Israel, where Moroccan Jews are thought to constitute the second biggest Jewish ethnic subgroup.

Populations (in thousands)

Year	Pop.	#
1960	11,635	
1970	14,952	+2.54%
1980	19,380	+2.63%
1990	24,167	+2.23%
2000	28,466	+1.65%
2010	31,894	+1.14%
2011	32,245	+1.10%
2012	32,597	+1.09%
2013	32,950	+1.08%
2014	33,304	+1.07%
2015	33,656	+1.06%

Source: [111]

table.toccolours | S | P | C | Done selecting!

Main article: Religion in Morocco

- Accede a la página de la wikipedia de Marruecos
- Localiza la tabla Populations (in thousands)
- Abre del DevTools y accede a la pestaña de Web Scraper

- Crea un mapa del sitio seleccionado “Create new sitemap”. Indica el nombre que quieras en “Sitemap name”, por ejemplo “wiki-marruecos” y en “Start URL” incluye la URL de la página web, en este caso “https://en.wikipedia.org/wiki/Morocco”
- Una vez creado el sitemap que hemos llamado wiki-marruecos debemos crear el selector desde el que vamos a extraer la información, en este caso, la tabla «Populations (in thousands)».
- Pon un nombre a Id, por ejemplo «tabla-poblacion».
- En Type selecciona «Table».
- En Selector, pulsa el botón Select y luego, en la página web, pulsa sobre toda la tabla. Quedará seleccionada en rojo. En el popup que aparece, pulsa sobre «Done selecting»
- En «Header row selector» seleccionar Select y, en la página web, marcar la fila de encabezado (Year, Pop…)y en el popup que aparece, pulsa sobre “Done selecting”
- Marca el cuadro de «Multiple» para que cargue todas las demás filas de datos
- Salva el selector con «Save selector»
- Nos aparecerá la información del selector guardado y podremos comprobar qué datos se van a cargar pulsado el botón «Data preview»

ID	Selector	type	Multiple	Parent selectors	Actions
tabla-poblacion	table.toccolours	SelectorTable	yes	_root	Element preview | Data preview | Edit | Delete

- Si queremos, podríamos repetir el proceso creando nuevos selectores. Estos pueden ser tablas, pero también podemos crear selectores html, imágenes, etc.
- En este caso vamos a «raspar» los datos y a descargarlos directamente en csv.
- En el menú de Web Scraper de nuestro Sitemap “Sitemap wiki-marruecos” seleccionamos la opción «Scrape»

- Cuando finalice, en ese mismo menú seleccionamos la opción «Export Data as CSV». Se descargará la tabla y ahora ya podremos trabajar con ella en un editor de hojas de cálculo o importarla a un servicio web para crear una infografía, etc., como en este ejemplo creado con Infogram.

Evolución de la población de Marruecos desde 1960

Aprender más

La web de Web Scraper tiene un **completo manual con videotutoriales** en el que puede seguirse, con más detalle, cómo trabajar con esta herramienta.

Navigate multi-level navigation to scrape all items in an e-commerce site.

Scrape an e-commerce site with pagination.

Grupo Ciberimaginario | Manuel Gertrudix - Alejandro Carbonell | 2025/2026 | Esta obra está bajo una Licencia Creative Commons Atribución 4.0 Internacional. Los contenidos citados se ajustan a lo regulado en el art. 32 del TRLPI de España