

Seminario de profundización Web Scraping

Andrés Camilo Serna

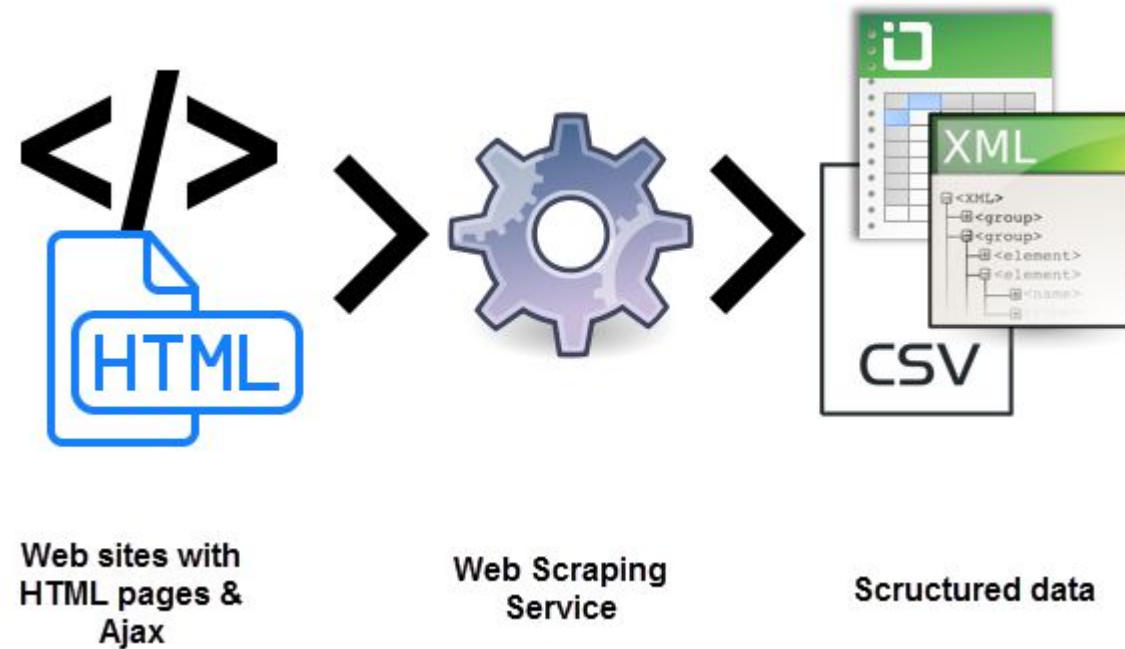
acserna@danalyticspro.co

Agenda

- Web Scraping – teoría
 - Tipos
- Ejemplo Web Scraping
- Ejercicio Web Scraping - HTML
- Ejercicio Web Scraping - API
- Caso de aplicación: Redes
- Ejercicio Redes

Web Scraping

Web Scraping



Ejemplo - Web Scraping

Se nos ha solicitado realizar la extracción automática de información de computadores de la siguiente página web.

https://www.flipkart.com/laptops/~buyback-guarantee-on-laptops-/pr?sid=6bo%2Cb5g&uniqBStoreParam1=val1&wid=11.productCard.PMU_V2

Es de interés conocer el nombre, precio y valoración (rating) de cada computador.

Ejemplo - Web Scraping

Se nos ha solicitado realizar la extracción automática de información de televisores de la siguiente página web.

[https://listado.mercadolibre.com.co/televisores#D\[A:televisores\]](https://listado.mercadolibre.com.co/televisores#D[A:televisores])

Es de interés conocer el nombre, precio, precio anterior y marca de cada televisor.

Ejemplo - Web Scraping

Usted como científico de datos trabaja en una empresa de consultoría que se encarga de realizar investigación a personas, durante su jornada laboral es menester realizar la debida diligencia (Due Diligence en inglés) a la empresa “**GRUPO OHANNA S.A.**”.

Su jefe le comunica que la información asociada a esta empresa se puede encontrar en la página web:

<https://opencorporates.com>

Se requiere conocer las personas que trabajan en dicha empresa y a su vez las empresas que están asociadas con cada una de dichas personas.

Finalmente, se requiere una representación gráfica (red) de estas relaciones.

Caso de aplicación - Redes

<https://lasillavacia.com/silla-llena/red-cachaca/enredo-desenredado-del-poder-colombia-72127>

Ejercicio de redes

<https://github.com/acserna/grafos>