



Almacenamiento y captura de datos

Claudio Aracena

GobLab - Universidad Adolfo Ibáñez
Chatbot Chile



Dudas de la clase anterior

- Cómo llevar a producción
 - Modificar script para guardar datos en sistema de archivos o base de datos
 - Generar archivo .py y llevar a servidor
 - Instalar librerías y correr script
 - Adicional: Utilizar crontab (sistema de calendarización) o llamar mediante API
- Docker y Selenium
 - <https://stackoverflow.com/questions/61805008/using-selenium-standalone-chrome-in-docker-compose-connecting-with-pythons-selenium>
 -



Contenidos

- Captura de datos desde archivos
- Base de datos
- Captura y almacenamiento de datos en BD
- Captura de datos de la Web (Web scraping)
- **Captura de datos de API (ej: Twitter)**
- Captura y almacenamiento en arquitecturas Big data

Códigos y clase en:

<https://github.com/caracena/almacenamiento-captura-datos>

Clase de hoy



Captura de datos de API

- Application Programming Interface (API)
- Conexión API vía requests
- Conexión API vía librerías





Application Programming Interface (API)

Una API es una forma estandarizada de compartir data en la web. Muchos sitios web comparten data utilizando APIs.

Algunos ejemplos de APIs son:

- Twitter API
- Google Maps API
- Github API
- Biblioteca del Congreso Nacional API



Conexión API vía requests

Una de las formas de conectarse a una API es mediante peticiones HTTP. Dependiendo lo que indique la documentación de la API se deberán realizar requerimientos GET, POST, PUT o DELETE.

Conexión API vía librerías

Otra forma de realizar peticiones a una API es utilizar alguna librería construida para esos fines. Generalmente, existen librerías de conexión para aplicaciones populares como Twitter, Google, Facebook, entre otras.