

Laboratorio de datos: web scraping y Procesamiento de Lenguaje Natural

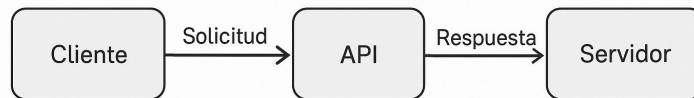
Clase 8c. Introducción a APIs



¿Qué es?

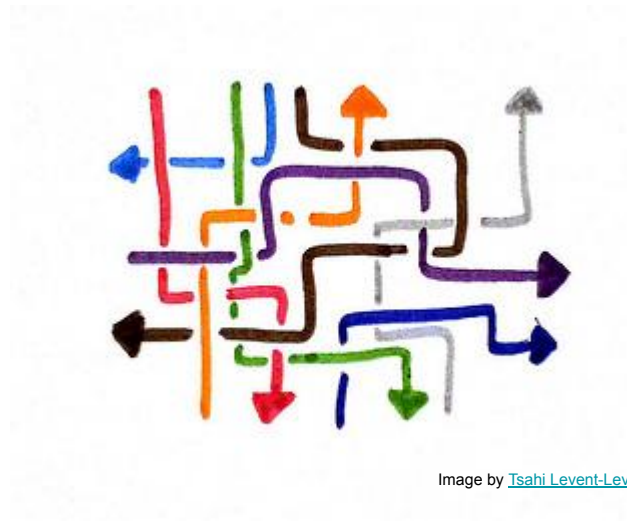
- Una API (Application Programming Interface) es un conjunto de reglas y protocolos que permite que diferentes programas se comuniquen entre sí.
- Elementos centrales:
 - Solicitud (request): lo que pedimos al servicio.
 - Respuesta (response): lo que el servicio nos devuelve.
 - Formato estandarizado: generalmente JSON o XML.

Funcionamiento general de una API



Usos posibles

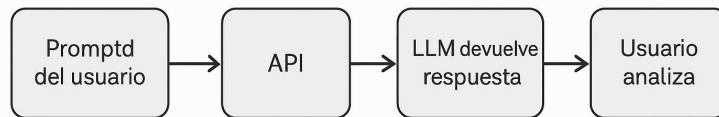
- **Acceso a datos:** bases estadísticas, censos, encuestas, información meteorológica, redes sociales.
- **Automatización de tareas:** por ejemplo, descargar diariamente un dataset sin intervención manual.
- **Servicios de terceros:** traducción automática, geolocalización, procesamiento de imágenes, modelos de lenguaje.
- **Ventaja principal:** las APIs permiten trabajar con información y servicios sin necesidad de descargar programas completos o manejar manualmente los datos.



Vamos al Notebook

APIs y LLMs

- Los LLMs (como GPT, Gemini, Claude, LLaMA) exponen una API que recibe texto de entrada (prompt) y devuelve texto generado como salida (completion).
- Elementos clave en la consulta a un LLM mediante API:
 - **Endpoint:** la URL a la que se envía la solicitud.
 - **Clave de autenticación:** código personal para acceder al servicio.
 - **Parámetros:** prompt, temperatura, número máximo de tokens, etc.



Gemini (LLM de Google)

- Familia de LLMs accesibles mediante la Google AI Studio API.
- Uso de clave de autenticación de Google Cloud.
- Respuestas en formato JSON.
- Posibilidad de controlar parámetros como temperature, maxOutputTokens, etc.



Una consulta a un LLM genérico

```
library(httr)
library(jsonlite)

url <- "https://api.llm-ejemplo.com/v1/chat"
api_key <- "TU_API_KEY"

prompt <- list(
  model = "llm-demo",
  messages = list(
    list(
      role = "user",
      content = "Explicá qué es una clase social en términos sencillos.")
    )
  )

respuesta <- POST(
  url,
  add_headers(Authorization = paste("Bearer", api_key)),
  body = toJSON(prompt, auto_unbox = TRUE),
  encode = "json"
)

resultado <- fromJSON(content(respuesta, "text"))
cat(resultado$choices[[1]]$message$content)
```

Vamos al Notebook