

2023

SCRAPPER TO API

2º Curso de Ciencia e Ingeniería de Datos
Escuela de Ingeniería
Informática Universidad de Las Palmas de
Gran Canaria

Esta es la memoria sobre el proyecto de Scrapping to API

Esta aplicación consiste en un programa que se encargará de extraer los datos de la página booking.com en formato JSON y ofrecerlos en forma de peticiones a través de una API REST. Para enviar solicitudes a dicha API REST, esta aplicación crea un servidor donde tiene configurada la petición GET, que mostrará los datos solicitados en la ruta web utilizando el protocolo HTTP, en formato JSON.

Índice: Recursos utilizados..... ¡Error! Marcador no definido.

Diseño 2

Conclusiones 3

Líneas futuras..... 3

Bibliografía 4

Recursos Utilizados:

Entre los recursos utilizados encontramos como entorno de desarrollo la aplicación IntelliJ IDEA Community Edition en su versión 2022.2.2. No se utilizó ninguna herramienta de control de versiones y la herramienta de documentación utilizada es Word.

Diseño

La estructura del código consiste en una carpeta llamada “scraper” que contendrá varias clases que conforman el programa final. Siendo la clase “Main” la cabeza de todo, esta se encargará de tomar las URL necesarias y proporcionárselas a la clase de “Scraping”, la cual accederá a ellas y procederá a la extracción de datos mediante varias funciones. Para la extracción de los datos ha sido necesario crear cuatro clases, donde cada una representará una petición distinta, tal y como se describe a continuación:

- La clase “Location” identifica a un nombre y una localización, siendo estos los datos de la cabecera, que en este caso corresponde al hotel Club Magic Life Fuerteventura, obtenida de la página web de booking.com.
- La clase “Service” se encargará de tomar el nombre y características del servicio en cuestión.
- La clase “Comments” utilizará una URL distinta a la general del programa, puesto que la sección de comentarios de booking.com se encuentra en una pestaña distinta. El programa tomará datos como el nombre, país, puntuación y la reseña de los usuarios, así como sus comentarios positivos y negativos y fecha de alojamiento en el hotel.
- La clase “Ratings” se encargará de tomar los valores de puntuación general y su categoría. Por ejemplo: Comida – 9,1.

El “Main” también se encargará de llamar a la clase “APIRest” la cual se encargará de inicializar el servidor y de obtener toda la información mencionada de la página web ya contenida en la clase “Scraping” para realizar un *port* en diferentes *paths* para cada petición GET. Dicha clase llamará a la clase “JsonTransformer”, que se encargará de transformar la información obtenida a un formato JSON.

Bibliografía

- https://www.youtube.com/watch?v=mxNtUllYuNM&ab_channel=LD20 (para el desarrollo del scrapping)
- <https://openai.com/blog/chatgpt/> (el uso de el chat gpt, herramienta que me recomendó Jose Juan, de donde aprendí a hacer relaciones entre clases y la utilización de varios métodos en general)