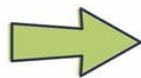


SCRAPPER TO AN API

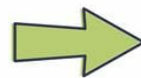
Booking



HTML WEBSITES



WEB SCRAPING



DATA

DESARROLLO DE APLICACIONES PARA LA CIENCIA DE DATOS
ESCUELA DE INGENIERÍA INFORMÁTICA Y MATEMÁTICAS
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
CIENCIA E INGENIERÍA DE DATOS
SEGUNDO CURSO

ADONAI OJEDA MARTÍN

14/01/2023

VERSIÓN I

ÍNDICE

1. Resumen.....	4
2. Recursos Utilizados.....	5
3. Diseño.....	6
4. Conclusiones.....	7
5. Líneas Futuras.....	8
6. Bibliografía.....	9

RESUMEN

Una web HTML nos permite poder acceder a su contenido. Se conoce como scrapping a la acción de acceder mediante programas software para extraer información de sitios web a través de su HTML y guardar los datos en una base de datos, en un sitio web o en cualquier otro lugar.

Acceder a hoteles de Las Palmas de Gran Canaria ha sido nuestra finalidad en nuestro scrapper, obteniendo de ellos su nombre, localización, descripción, rating, opinión general, comentarios, y su link.

A continuación, se ha recogido estos datos para estructurarlos en una api y a través de operaciones rest poder acceder a ellos en formato Json.

Finalmente, podemos ver esta información estructurada en Json y podemos estudiar diferencias entre hoteles o cualquier curiosidad que tengamos sacando conclusiones dependiendo de nuestros intereses.

RECURSOS UTILIZADOS

La práctica ha sido realizada en el entorno de desarrollo IntelliJ IDEA CE, con la versión java 11 jdk.

Asimismo, para su desarrollo y obtener todas las versiones anteriores y posteriores del proyecto se ha subido a un repositorio de GitHub, siendo esta la primera versión, el scrapping a hoteles de Las Palmas de Gran Canaria de Booking.

Finalmente, Java Doc 11 ha sido la herramienta de documentación usada.

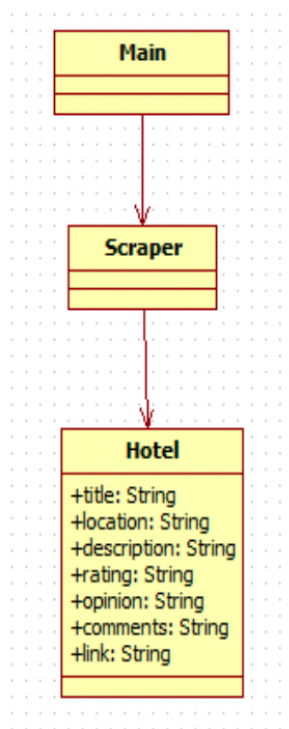
DISEÑO

En primer lugar, hemos definido un paquete cuyo nombre es “org.es.ulpgc.dacd.apiScraper”, que contiene las clases con las que ha sido desarrollada la práctica.

En segundo lugar, hemos creado la clase “Hotel” que es la clase POJO, que contiene los atributos y métodos necesarios e influyentes en el desarrollo de la práctica.

En tercer lugar, la clase “Scraper”, que crea la conexión a Booking recogiendo el HTML y la información de este relevante para nuestra práctica, además contiene métodos que nos harán falta en la clase principal.

Finalmente, la clase principal “Main”, que define el localhost al que conectarse, las posibles request que le puedes hacer con sus respectivas response y finalmente el formato en que se devuelve.



CONCLUSIONES

En conclusión, esta práctica nos ha servido para investigar acerca de lo que es el scrapping, su uso, y su desarrollo en código adquiriendo habilidades antes desconocidas. Además de poder conocer y afianzar conceptos de cómo hacer una api rest.

LÍNEAS FUTURAS

Nuestra API podría ser de gran utilidad en el mercado, ya que muchos hoteles a través de esta recopilación de datos podrán estudiar y examinar a otros hoteles a través de sus ratings, comentarios, ..., comparándose y viendo que podrían mejorar y ofrecer al cliente para destacar.

Por otro lado, esta recopilación de datos es de Las Palmas de Gran Canaria, podríamos usar esta práctica para examinar otros territorios o incluso comparar hoteles entre islas, por ejemplo, Gran Canaria y Tenerife, hacer un estudio a medida y recomendar porque un sitio es mejor que el otro o viceversa, que los turistas lo agradecerán.

Finalmente, también podríamos examinar otras webs HTML, no solo Booking, sino Trivago, por ejemplo.

BIBLIOGRAFÍA

https://www.booking.com/searchresults.es.html?label=gen173nr-1FCAEoggl46AdIM1gEaEaIAQGYAQq4ARfIAQzYAQHoAQH4AQulAgGoAgO4Apio8J0GwAIB0glkM2RjODgwMTUtNWZjYy00NTlmLWlzMGQtNzQ5MmZkZTE2NWQw2AIG4AIB&sid=1c54fe47c864ff28f788bc14a7b9f120&sb_lp=1&src=region&error_url=https%3A%2F%2Fwww.booking.com%2Fregion%2Fes%2Fgran-canaria.es.html%3Flabel%3Dgen173nr-1FCAEoggl46AdIM1gEaEaIAQGYAQq4ARfIAQzYAQHoAQH4AQulAgGoAgO4Apio8J0GwAIB0glkM2RjODgwMTUtNWZjYy00NTlmLWlzMGQtNzQ5MmZkZTE2NWQw2AIG4AIB%26sid%3D1c54fe47c864ff28f788bc14a7b9f120%26&ss=Gran+Canaria&is_ski_area=0&ssne=Gran+Canaria&ssne_untouched=Gran+Canaria®ion=754&checkin_year=&checkin_month=&checkout_year=&checkout_month=&group_adults=2&group_children=0&no_rooms=1&bh4u_keep_filters=&from_sf=1&dest_type=city&dest_id=-388528

<https://jarroba.com/scraping-java-jsoup-ejemplos/>

<https://www.youtube.com/watch?v=yw7B85174JQ>

<https://oracle-max.com/creando-un-api-rest-con-java-y-netbeans-que-devuelva-un-json-en-su-request/>

<https://learn.microsoft.com/es-es/azure/cognitive-services/bing-news-search/java>