# Developing a Scraper that publishes data through an API

## Desarrollo de Aplicaciones

CIENCIA E INGENIERÍA DE DATOS
ESCUELA DE INGENIERÍA INFORMATICA
UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

Alejandra Ruiz de Adana Fleitas

Entrega: sábado, 7 de enero de 2023

## Resumen:

El proyecto se divide en varias clases y 2 interfaces los cuales se encuentran dentro de 3 carpetas llamadas: *"database"*, *"model"* y *"webservice"*. Dentro de las diferentes carpetas encontramos las clases *"DMLTranslater"*, *"SqliteBookingDatabase"*, *"Service"*, *"Assessment"*, *"Comment"*, "*Location*", *"Scraping*", *"ScrapingApi"* y *"JsonTransformer"*; y las interfaces *"BookingDatabase"* y *"BookingSource"*; además fuera de las carpetas encontramos la clase *"Main"* el cual se encargará de que todo el programa se ejecute.

El trabajo consiste en extraer datos de Booking de manera automatizada, a través de un scraping, guardándolos en clases POJO para luego implementar un servicio web que exponga estos datos en formato JSON. Lo primero que haremos será, a través de la clase "*Scraping*", extraer los datos de la página del Ac Hotels en Booking mediante su URL, pasando estos datos a parámetros, permitiéndonos conocer los datos necesarios para poder inicializar los objetos correspondientes a las clases *"Location"*, *"Service"*, "*Assessment*" y "*Comment"*. A continuación, crearemos la clase "*JsonTransformer*", en la cual representaremos los objetos que queremos mostrar en formato JSON. Por último, en la clase "*ScrapingApi*" utilizaremos la librería Spark para crear un servicio web que expondrá peticiones GET a través de una URL, mostrando un JSON con toda la información.

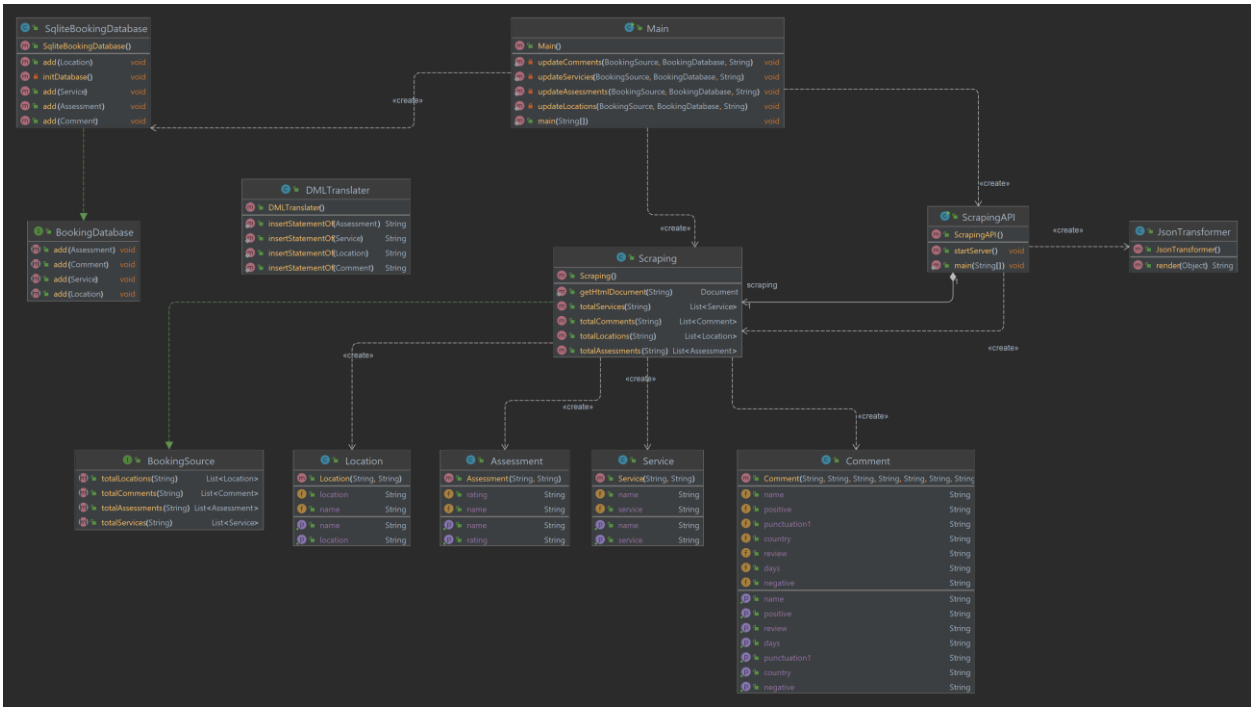
Para finalizar, mencionar que, como complemento del trabajo, también he desarrollado una base de datos en la que se encuentran cuatro tablas con los datos recogidos del "Scraping". De esta manera, después de recoger los datos pertinente a través del "Scraping", he creado una base de datos, la cual contiene cuatro tablas y en las cuales, a través *"DMLTranslate"*, se irán añadiendo los atributos de cada objeto uno por uno.

## ÍNDICE:

## Recursos utilizados:

Como entorno de desarrollo he utilizado el programa intellij, como herramienta de control he utilizado Git y por último como herramienta de documentación he utilizado Word.

## Diseño:




## Conclusiones:

En líneas generales este trabajo me ha parecido más fácil de realizar que el anterior, principalmente porque el anterior al ser el primer trabajo no sabia ni como empezar, ni como manejarme bien con el programa de Intellj. Sin embargo, con este proyecto ha sido todo lo contario, ya que desde el primer momento he sabido como tenía que enfocarlo y, aunque he tenido algunas dificultades he sabido como resolverlas con bastante solvencia.

A lo largo del presente proyecto he aprendido como hacer un scraping, lo cual me parece muy importante de cara al futuro, y como hacer un Api Rest. Lo más que me ha costado de este proyecto es realizar la Api Rest, ya que, aunque habíamos visto ejemplos en clase, como nunca había intentado hacer una por mi misma no sabía realmente como se hacía. También me ha servido para utilizar y practicar los conocimientos aprendidos en clase y trabajos anteriores, afianzando así el contenido, contenido como pueden ser las sentencias SQL, la deserialización mediante librerías gson, la creación de una base de datos o el uso de clases POJO.

Por último, si volviera a hacer este proyecto lo enfocaría de la misma manera ya que me parece una manera bastante clara, sencilla y efectiva de hacerlo.

## Líneas futuras:

Booking es una de las mayores plataformas de reserva de hoteles, vuelos, alquiler de coches, actividades y servicios, siendo de esta manera una de las plataformas donde se realiza un mayor número de valoraciones y, por tanto, donde se puede obtener más información.

En un futuro, para crear una aplicación útil y funcional utilizando como base un scraping de Booking, se podría establecer una base de datos gigante con todas las valoraciones y comentarios que hay sobre un hotel, actividad, etc; añadiéndole a dicha base nuevos datos recogidos, por ejemplo, cada 4 meses. De esta manera, tendríamos una forma de conocer las opiniones que tienen sobre un servicio nuestros clientes, pudiendo hacer estudios estadísticos para saber en qué aspectos tendríamos que mejorar. Así mismo, el actualizar los datos cada 4 meses nos permitiría saber cuáles han sido las valoraciones negativas más recientes y, tras hacer el estudio estadístico, las más repetidas, y por tanto las más relevantes a tener en cuenta, ya que no tendrá la misma importancia una valoración de hace 4 meses a una de hace 2 años. Al mismo tiempo, también nos permite ver nuestro rango de mejora ya que podemos comprobar si los siguientes 4 meses seguimos teniendo los mismos tipos de valoraciones negativas, si ya las hemos solucionado o si tenemos otras nuevas.

La forma de desarrollar este proyecto sería utilizando la misma metodología que hemos utilizado en este trabajo e implementándole además un “Timer Task” para que se actualicen los datos cada cierto tiempo.

## Bibliografía


- https://jarroba.com/scraping-java-jsoup-ejemplos/
- https://www.youtube.com/watch?v=kn6Xob8SAZo&ab_channel=DavidPachecoJimenez
- Y recursos del campus para