



Booking

SCRAPER

Developing a scraper that takes information from different hotels in Booking.

Desarrollo de Aplicaciones para Ciencia de Datos / 2ºCurso

Grado en Ciencia e Ingeniería de Datos

Escuela de Ingeniería Informática / ULPGC

Resumen

Este proyecto se basa en realizar un “scraping” de una web acerca de hoteles para obtener información de estos, de tal modo se creará de una API Rest que sea capaz de mostrar información sobre un hotel ya sea su ubicación, servicios, comentarios y valoraciones.

Para este proyecto, cabe mencionar que hemos utilizado tanto las librerías de Jsoup y Spark.

En cuanto al funcionamiento del programa, este está dividido en tres paquetes: La “api”, que contiene la clase “WebService” dedicada a la creación del web service encargado de solicitar la información del hotel deseado a la API haciendo uso de Spark, es decir, se implementa los métodos “get”.

El paquete “model”, que contiene las clases “Comment”, “Hotel” y “Location”, donde se definen las variables del hotel con sus respectivos atributos.

Y por último el paquete “scraper”, donde se implementan los métodos para obtener la información de los hoteles de la página web de Booking, los cuales son guardados acto seguido en unas listas para que posteriormente el “WebService” puede acceder a dicha información.

Finalmente, tenemos tanto el “main” donde deberemos insertar la url de Booking del lugar que quieras información con “&offset=” para que este funcione, como su “controller” que se utiliza para no sobrecargar las clases de métodos que no les corresponden.

Índice

Recursos utilizados.....	Pág 3
Diseño.....	Pág 3
Conclusiones.....	Pág 4
Líneas futuras.....	Pág 4
Bibliografía.....	Pág 4

Recursos utilizados

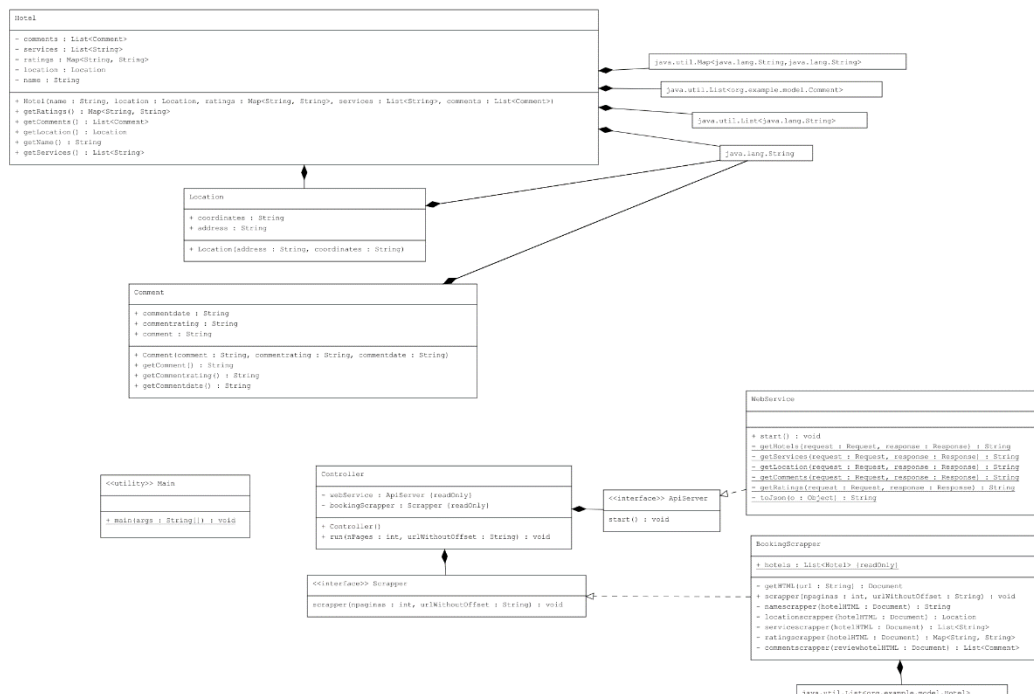
Intelli J: Entorno de desarrollo del programa.

GitHub: Repositorio online con el código del programa.

Microsoft Word: Editor de texto para la memoria.

Diseño

Para este proyecto se ha tenido en cuenta el principio de responsabilidad única por clase, en el que cada clase debe solo ocuparse de una funcionalidad. También empleamos la separación por módulos y, además, se ha hecho uso del principio de inversión de la dependencia al usar interfaces.



Conclusiones

En este trabajo he aprendido a como realizar un scraper de una web de manera estructurada separando en diferentes paquetes las distintas funcionalidades del programa, me he dado cuenta de que hay que tener cuidado con los campos que se deben seleccionar de las urls para obtener los datos ya que pueden dar problemas, pero en general el proyecto me ha parecido muy interesante.

Líneas futuras

Este proyecto se le podría dar una mayor utilidad y una posible forma de comercializarlo si se desarrollara una especie de app o interfaz que permitiera seleccionar un hotel concreto y se nos mostrara la información sobre el mismo y una especie de análisis acerca de los comentarios (análisis de sentimiento) para poder valorar la experiencia del cliente en el hotel, esto sería muy útil para las compañías hoteleras que podrían utilizar este servicio.

Bibliografía

<https://www.booking.com/>

<https://stackoverflow.com/>

<https://mvnrepository.com/>

David García Díaz

13/11/2023