



7 de enero de 2023

# Proyecto Scraper

Desarrollo de Aplicaciones para  
ciencia de datos

**Susana Suárez Mendoza**

GRADO EN CIENCIA E INGENIERÍA DE DATOS

2º CURSO

ESCUELA DE INGENIERÍA INFORMÁTICA

Universidad de Las Palmas de Gran Canaria

**Resumen.**

La finalidad d este proyecto es proporcionar diferente información sobre diferentes hoteles encontrados en Booking. Consta de las siguientes partes:

1. La información es obtenida mediante la realización de un scraper a dicha página web obteniendo su documento HTML.
2. Toda esta información puede ser consultada por una API desarrollada para la consulta de la misma. Las respuestas de cada petición vendrán dadas en el formato JSON.

## ❖ Recursos utilizados.

Este proyecto ha sido desarrollado en el entorno de programación llamado IntelliJ Edu utilizando el lenguaje de programación llamado Java para llevar a cabo toda la implementación.

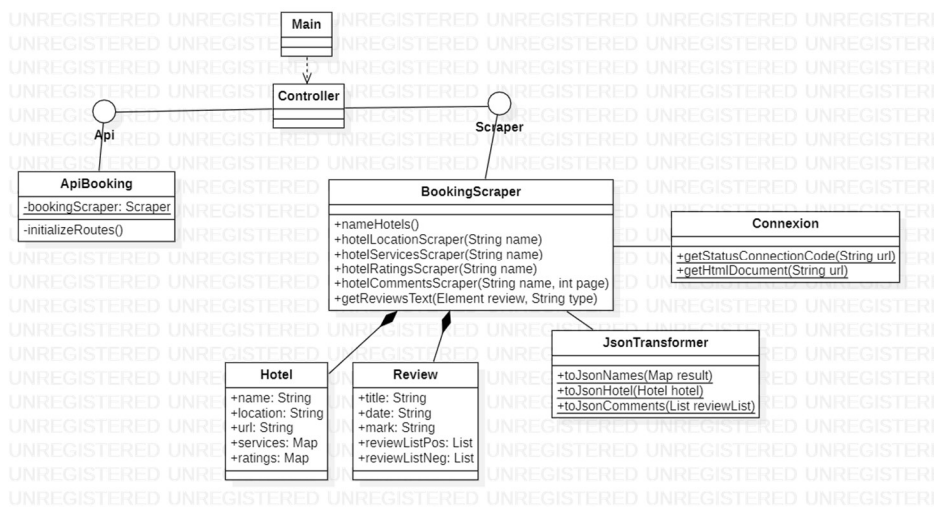
La documentación utilizada han sido las diferentes Katas realizadas en clases prácticas para desarrollar la API con Spark y la inspección de la página web de Booking para conocer las diferentes clases que contienen toda la información.

## ❖ Diseño.

### • Principios de diseño:

- Responsabilidad única. Cada clase contenida en este proyecto solo tiene una responsabilidad.
- Abierto a la ampliación, cerrado a la modificación. Si en un futuro además de realizar Scraper a Booking se quiere realizar a otra página web, sería posible debido a la Interfaz existente.
- Inversión de la dependencia. Los módulos de bajo nivel dependen de las abstracciones y no de módulos de alto nivel con el uso de Interfaces.

### • Diagrama de clases:



## ❖ Conclusiones.

De esta tarea he aprendido a realizar scraping a una página web por lo que, si en un futuro tuviese que realizar scraper a cualquier otra página, me vería capaz de realizarlo. También he aprendido a realizar una API con Spark y la ejecución de la misma.

## ❖ Líneas futuras.

Para poder evolucionar este producto se podría añadir más scrapers a otras páginas web además de proporcionar diferentes apis para cada una de ellas.

