

## Proyecto scraper

# Booking.com

Desarrollo de Aplicaciones para Ciencia de Datos

2º año de Grado de Ingeniería en Ciencia de Datos.

Escuela Universitaria de Informática. Universidad de Las Palmas de Gran Canaria.



Irene Guerra Déniz

Versión: v1.0

Fechas de revisión: 28/12/2022, 11/01/2023, 13/01/2023

## Resumen:

Este proyecto consiste en recoger información de hoteles a través de scraping y mostrar esa información por medio de una API REST. El scraping se realiza a la página web Booking.com.

En primer lugar, se obtiene el nombre del hotel que solicita el usuario a través de la API y se realiza una búsqueda en Google con una url que contiene “booking” y el nombre del hotel al lado. De esta manera se accede a la página de Booking del hotel en específico, y de ahí se scrapean las distintas secciones (ubicación, teléfono, servicios, valoraciones y comentarios). Para el teléfono se ha tenido que acceder excepcionalmente a Kayak, ya que es el único buscador de hoteles que otorga la información telefónica de ellos. Si el hotel se encuentra en Booking pero no en Kayak se mostrará toda la información menos el teléfono.

Cada vez que el usuario haga una solicitud, deberá poner qué información quiere específicamente de las secciones que se mencionaron con anterioridad, para así solo scrapear la parte solicitada y ahorrar tiempo de ejecución. La información se le pasa al Webservice de distinta manera según el campo solicitado. Si se quiere la información básica se pasa un objeto de tipo Hotel que contiene el teléfono y un objeto de tipo Location (con coordenadas y dirección), si se solicitan los servicios se le devolverá un HashMap con el nombre de la categoría del servicio como clave, y con una lista de los servicios de esa categoría como valor, si se piden las valoraciones se devuelve otro HashMap con la categoría a valorar como clave y la puntuación como valor, y si se quiere conocer los comentarios se accederá a una url distinta de booking donde se concentran específicamente los comentarios del hotel. Esta información se devolverá como una lista de tipo Review, que posee como atributos el título del comentario, el comentario positivo y el comentario negativo. Toda esta información cuando llega a la clase Webservice se pasa a Json para plasmarla en la API.

En principio el proyecto busca cualquier hotel que se encuentre en Booking, para probar el funcionamiento se han probado hoteles como el AC Hotels, Gloria Palace Amadores, Baobab, Riu Plaza España, entre otros. Además, se pueden escribir los nombres con espacios o guiones entre las separaciones, por lo que en principio no hay problema con el formato en el que se escriba el nombre del hotel en la API.

## Índice

1. Recursos utilizados
  - a) Entorno de desarrollo
  - b) Herramienta de control de versiones
  - c) Herramienta de documentación
2. Diseño
  - a) Patrones y principios de diseño
  - b) Diagrama de clases
3. Conclusiones
4. Líneas futuras
5. Bibliografía

## Recursos utilizados

### Entorno de desarrollo

Se ha utilizado IntelliJ para desarrollar el proyecto y la extensión de Google Chrome Talend API Tester para usar y testear la API.

### Herramienta de control de versiones

Para el control de versiones se ha empleado Git, y se ha subido el proyecto a GitHub.

### Herramienta de documentación

Para la documentación del proyecto se ha utilizado Microsoft Word y luego se ha exportado a PDF.



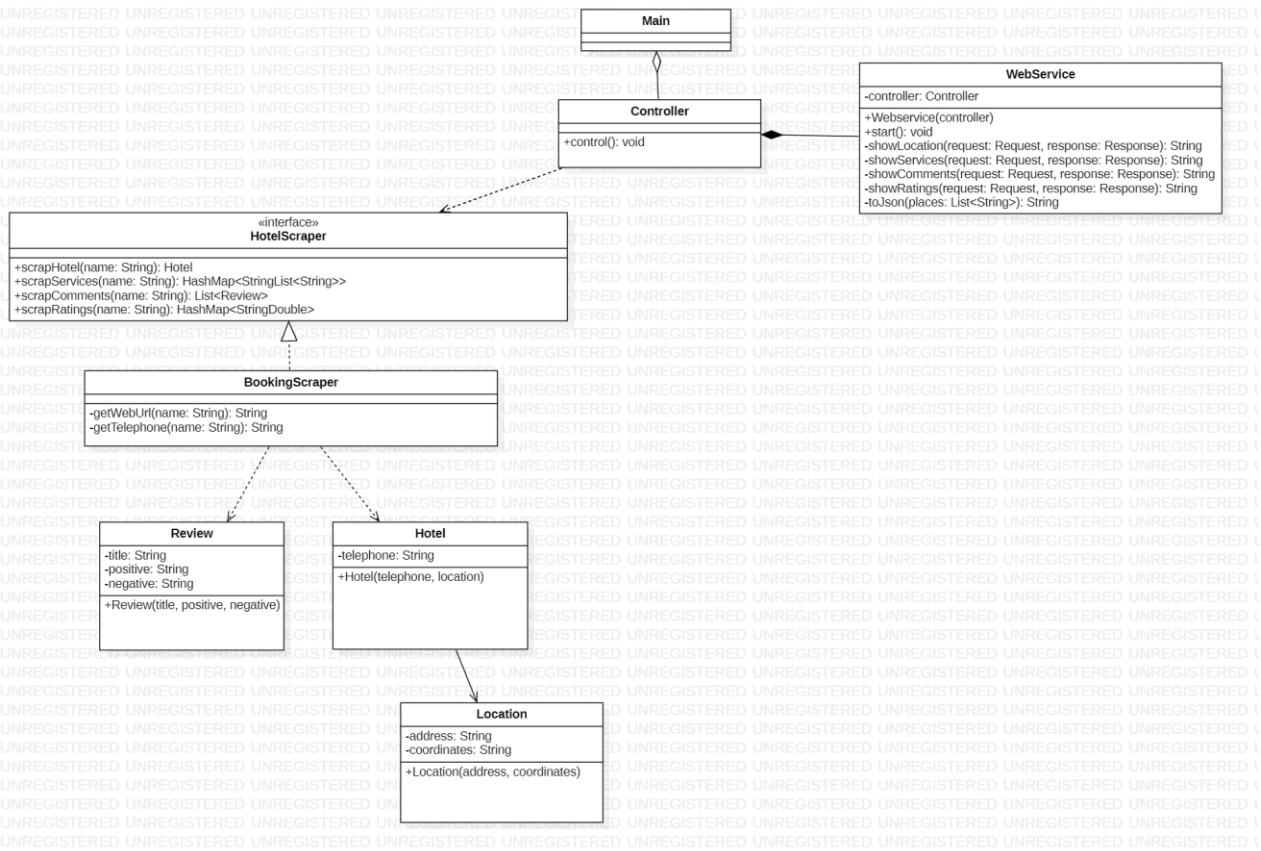
## Diseño

### Patrones y principios de diseño

Se ha aplicado el estilo de arquitectura software model-view-controller (modelo, vista, controlador) sin tener definida la vista ya que no se posee una interfaz de usuario en el proyecto.

Además, se han intentado implementar los principios de diseño SOLID.

Diagrama de clase



## **Conclusiones**

Con la realización de este proyecto he practicado scraping encontrado una manera de buscar en Google páginas para poder acceder a toda la información que necesito. De esta manera el usuario puede encontrar información de casi cualquier hotel, en vez de unos pocos predefinidos. También he puesto en práctica el desarrollo de una API REST y como extraer información de la solicitud (en este caso el nombre del hotel) y enviar una respuesta según el path introducido.

## **Líneas futuras**

Para mejorar el proyecto de cara al futuro se podrían almacenar los datos que el usuario va solicitando, tratarlos como hoteles favoritos o buscados, y guardarlos en una base de datos. Así, de esta manera, poder comprobar las similitudes y diferencias entre hoteles e incluso estudiarlos para filtrarlos según distintas categorías (como calidad-precio, mejores comentarios, mayores puntuaciones, mayor cantidad de servicios, etc). Con esto se podría facilitar la elección de un posible hotel para el usuario.

## **Bibliografía**

<https://stackoverflow.com/>

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

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

<https://www.kayak.es/>