

---

# PROYECTO SCRAPPING DE HOTELES(BOOKING)

---

Desarrollo de Aplicaciones para Ciencia de Datos (DACD)



2º GRADO DE CIENCIA E INGENIERÍA DE DATOS (GCID)  
ESCUELA DE INGENIERÍA INFORMÁTICA (EII)  
Universidad de las Palmas de Gran Canaria (ULPGC)

## **Autor**

José Gabriel Reyes Rodríguez

## **Fechas de desarrollo del proyecto**

desde el 02-12-2022 hasta el 08-12-2022

## **Versiones**

1

## **Revisiones de la memoria**

0 revisiones (por tutoría)

3 revisiones (propias)

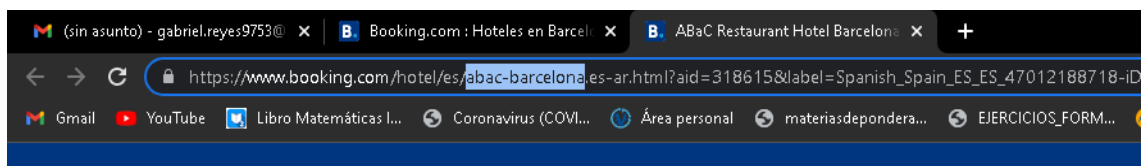
## Índice:

Resumen:.....	3
Recursos utilizados.....	4
Entornos de desarrollo.....	4
Herramientas de control de versiones.....	4
Herramientas de documentación .....	4
Diseño.....	4
Conclusiones .....	5
<b>Líneas futuras</b> .....	5
Bibliografía .....	5

## Resumen:

Para este trabajo, se ha desarrollado un scrapper para la página Booking.com en donde podemos tomar los datos más relevantes de los hoteles que existen dentro de esta página en España, para esto, esta implementación lo que hace principalmente es en base a una url “base” para todos los hoteles y el label facilitado en la petición API se consiguen todos los datos solicitados por el usuario. El programa funciona de tal manera que, con cada ejecución, se hace un scrapping de la página solicitada in situ, es decir, no se persisten datos en la aplicación. En cada petición, el programa no tarda más de 4 segundos en cada una de las peticiones. Las ventajas que ofrece esta metodología es la capacidad de poder ofrecer información de cualquier hotel mientras este sea de España, como punto negativo a esto, al no persistir los datos, si se hace la misma petición numerosas veces, para cada una se obtendría la información desde Booking.com y no desde una base de datos o un Datalake, Por otra parte, el programa está pensado únicamente para serializar los datos en JSON ya que no se trabaja con clases POJO o clases “record” para hacer objetos, sino que todos los datos son ingresados en mapas del tipo Hash o LinkedHashMap y estos mapas al estar completos son pasados a formato JSON para su entrega.

Para ejemplificar lo que sería necesario que el usuario aportase en su petición como hotel sería la parte seleccionada en la siguiente imagen:



Esto se ha hecho de esta manera ya que esta label y el nombre exacto del hotel no tienen porqué coincidir en Booking, por eso es necesario el nombre exacto del label.

Como en el enunciado no se especificaba para cuantos hoteles debía hacerse, decidí optar por esta solución ya que estaría disponible para todos los labels de booking.

## Recursos utilizados

## Entornos de desarrollo

IntelliJ IDEA versión educativa edición 222.4167.41

## Herramientas de control de versiones

Git

## Herramientas de documentación

Microsoft Word.

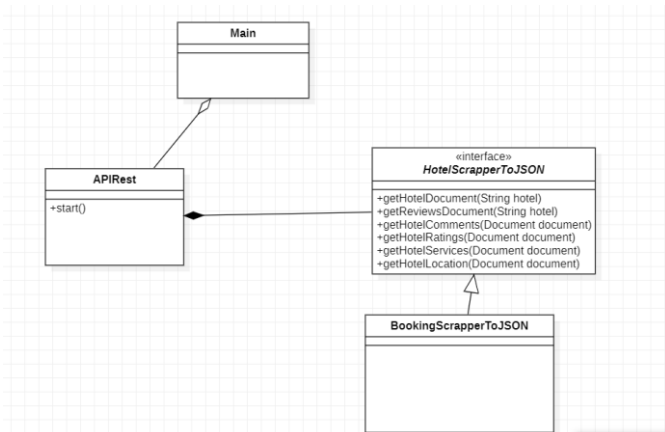
## Extras:

Talented API tester

## Diseño

Para el diseño, considero que se sigue un MVVM ya que no existe un controlador como tal, sino que el propio Web service cuenta con una función start que inicia la API y a su vez las peticiones son las que llaman a la clase de BookingScrapperToJSON para ejecutarse, además, lo más parecido a un modelo del que se dispone serían los diferentes mapas utilizados para su conversión a JSON que estarían en la clase BookingScrapperToJson.

Diagrama de clases:



## Conclusiones

Con este proyecto, he aprendido a usar la spark y gestionar diferentes peticiones con parámetros incluidos en el path y a estructurar ligeramente mejor el programa en relación con el anterior trabajo de spotify, donde en este trabajo si he incluido una interfaz para la base de datos y he podido comprobar la utilidad de las mismas.

Si volviese a empezar de cero, quizás me plantearía utilizar en vez de mapas para recoger los datos, objetos de clases “record” para cada petición (como objetos tipo comentarios, etc...) aunque desconozco qué método sería más eficaz en cuánto a tiempo se refiere.

## Líneas futuras

Con este tipo de proyectos, sería interesante poder hacer peticiones sobre el hotel con mejor valoración en ciudades específicas o el hotel con mejores servicios en una ciudad determinada o en base a las propias estrellas del hotel. Por último, a partir de estas peticiones más avanzadas, poder desarrollar una página web para hacer rankings de hoteles en base a las peticiones anteriormente mencionadas o un ranking de mejores hoteles calidad-precio en un área determinada.

## Bibliografía

url para sacar los datos generales con un ejemplo:

<https://www.booking.com/hotel/es/abac-barcelona.es.html>

url para sacar los datos de los comentarios con un ejemplo:

<https://www.booking.com/reviews/es/hotel/abac-barcelona.es.html>

Para pasar de mapa a JSON:

[How can we convert a map to the JSON object in Java? \(tutorialspoint.com\)](https://www.tutorialspoint.com/java/java_map_to_json.htm)