

Developing a Scraper that publishes data through an API (optional)

Desarrollo de Aplicaciones para Ciencia de Datos

Segundo Curso

Grado de Ingeniería en Ciencia de Datos

Escuela Universitaria de Informática

Universidad de Las Palmas de Gran Canaria

The screenshot shows a TripAdvisor review for a property in Mogán, Gran Canaria. The review is by user 'Garceitor' from February 2022, with 9 contributions and 3 useful votes. It is marked as the 'Opinión favorita del establecimiento' (Favorite opinion of the establishment). The review is titled 'Maravilloso' (Amazing) and describes a stay in November, praising the location, cleanliness, and amenities. Below the review, a list of 'Servicios de habitación' (Room services) is shown, including air conditioning, cleaning service, safe, seating area, coffee maker, American kitchen, cable/satellite TV, and bathroom/shower. The property has a 5.0 'Excelente' (Excellent) rating based on 116 opinions, ranking it 1st out of 76 special accommodations in Mogán. A breakdown of ratings for location, cleanliness, service, and value is provided. The address is listed as Antónico El Molinero, 12, 35149, Mogán, Gran Canaria España.

Garceitor escribió una opinión (feb 2022)
9 contribuciones • 3 votos útiles

Opinión favorita del establecimiento

Maravilloso

«Estuvimos en noviembre pasamos unos días estupendos en estos apartamentos zona tranquila todo muy limpio asignación de hamacas según te registras y solo para adultos cerca de Mogán playa paseando 15m un lugar para volver si quieres descansar y al mismo tiempo disfrutar no lo dudes»
Leer más ▾

Servicios de habitación

- Aire acondicionado
- Servicio de limpieza
- Caja fuerte
- Zona de asientos
- Cafetera/tetera
- Cocina americana
- TV por cable o satélite
- Bañera/ducha

5,0 Excelente 116 opiniones

1 de 76 alojamientos especiales en Mogán

- Ubicación
- Limpieza
- Servicio
- Relación calidad-precio

Antónico El Molinero, 12, 35149, Mogán, Gran Canaria España

Resumen

En este proyecto tendré la oportunidad de aprender y desarrollar mis conocimientos sobre los temas inculcados en las clases tanto teóricas como prácticas.

Consistirá en la realización de un “scraper” utilizando la Web de TripAdvisor y, además, el desarrollo de un web service para mostrar la información recolectada.

Para la realización de este scraper y web service me apoyaré en el entorno de programación IntelliJ para usar Java y en la Web de TripAdvisor.

Contenido

Resumen..... 2

Recursos utilizados..... 4

Diseño..... 5

Conclusiones 6

Líneas Futuras 7

Bibliografía 8

 Documentación 8

 Imágenes 8

Recursos utilizados

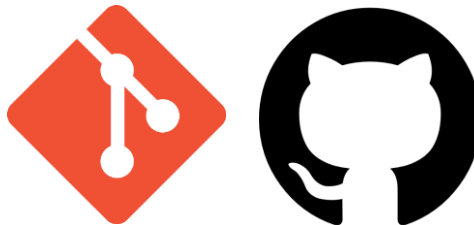
El entorno de desarrollo con el que he usado para programar en Java ha sido **IntelliJ**.



El portal web aprovechado es <https://www.tripadvisor.es/>



La herramienta de control de versiones usada ha sido **Git** y **GitHub**.



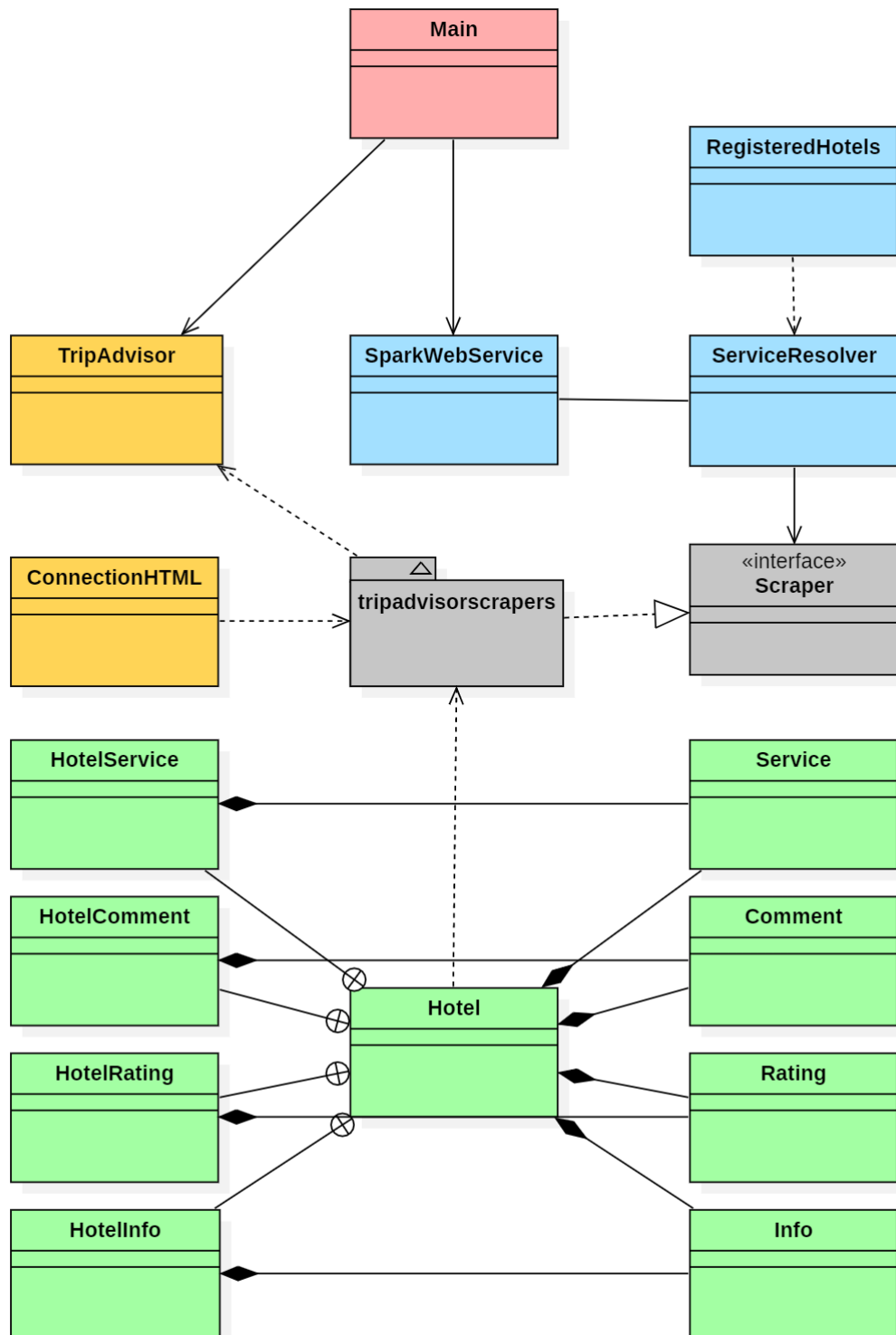
La herramienta de documentación empleada en esta memoria es **Microsoft Word**.



Diseño

Los patrones y principios de diseño seguidos han sido los recomendados por el profesorado en el documento de “*Guía de estilo Java y Clean Code.pdf*” subido al Campus Virtual además del principio de diseño de “Separation of corners”.

Diagrama de Clases



Conclusiones

Mi experiencia personal durante el trabajo de este proyecto ha sido bastante gratificante gracias a la capacidad que he tenido documentarme en Internet por mí mismo, para el entendimiento del código y su posterior expansión.

Del mismo modo, también me siento satisfecho ya que este tipo de actividades me acercan más al mundo real de la programación, siendo capaz de ver, en esta vez con claridad, la cantidad de aplicaciones que se pueden llevar a cabo con ella; conectando, en este caso, Java y TripAdvisor.

Si empezara de nuevo este proyecto, lo único que llevaría a cabo de otra manera sería la optimización del tiempo empleado, ya que he ido afinando mis habilidades para buscar la información necesaria; y añadir el apartado opcional de crear un método en el web service para obtener todos los nombres de hoteles en una ciudad, el cual no he hecho por falta de tiempo; no obstante, quizás también sumaría más hoteles para ofrecer su información.

Líneas Futuras

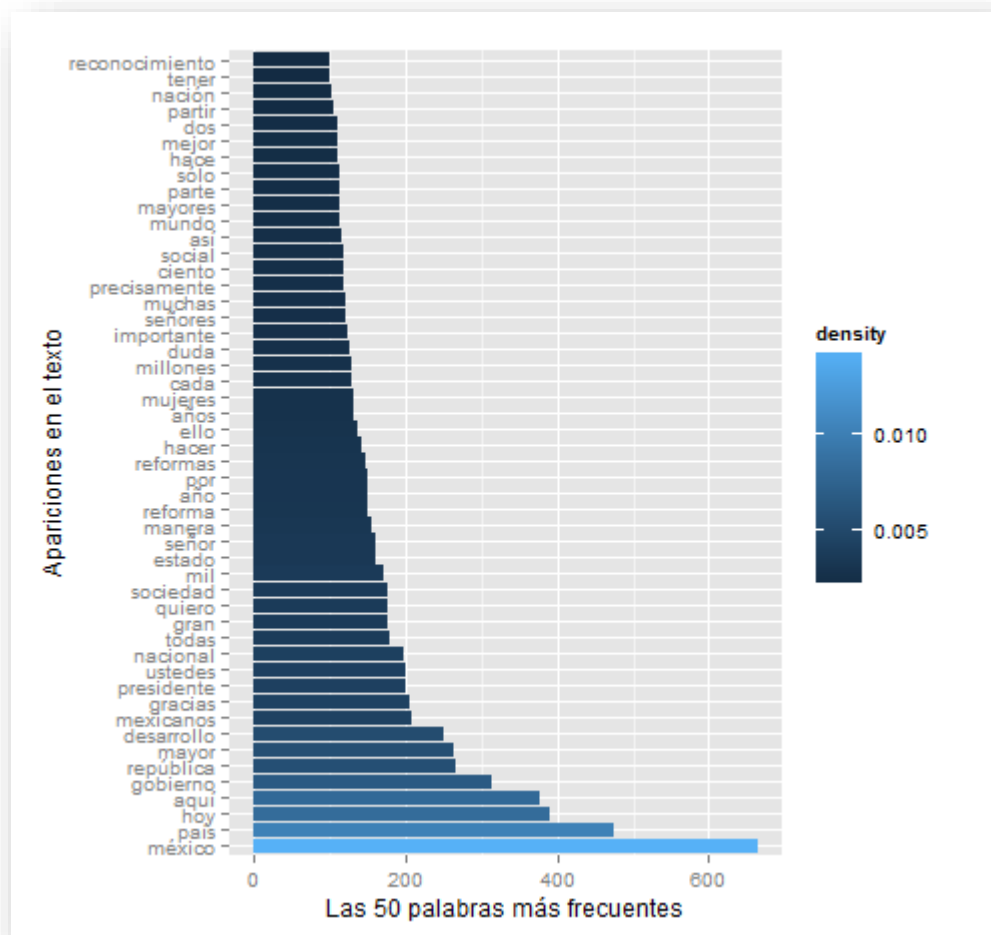
Personalmente, si tuviera que seguir desarrollando este producto para comercializarlo, lo llevaría al ámbito del seguimiento de hoteles y haría estudios sobre, por ejemplo, la satisfacción de los clientes de estos.

Una manera de hacerlo sería fijándonos los comentarios que podemos encontrar en la Request de “get /hotels/:name/comments”.

Así podríamos ver como un hotel en concreto es valorado por sus clientes a medida que pasa el tiempo, ya que estos datos se van actualizando cada vez que un usuario añade una reseña.

Con estos datos, y usando alguna librería externa como la de NLTK (procesamiento del lenguaje natural), podríamos ponernos en contacto con algún hotel y ofrecer nuestros servicios de seguimiento estadístico.

Un posible gráfico de esto podría ser el siguiente:



En nuestro ámbito, podríamos ver palabras relacionadas con los comentarios de nuestro hotel de estudio; de este modo, sabríamos en que campos nos relacionan más nuestros clientes.

Bibliografía

Documentación

Scraping:

<https://jarroba.com/scraping-java-jsoup-ejemplos/>

Imágenes

TripAdvisor:

https://play-lh.googleusercontent.com/1VJj_XWx9gt2dHi82fZLK-6UOzadeN122yD11vX2l718plfJDSFmaChjry7AgKCJgQ

<https://gofishing.is/wp-content/uploads/2013/12/banner-tripadvisor-2013.png>

IntelliJ:

https://upload.wikimedia.org/wikipedia/commons/thumb/9/9c/IntelliJ_IDEA_Icon.svg/2048px-IntelliJ_IDEA_Icon.svg.png

Git:

<https://git-scm.com/images/logos/downloads/Git-Icon-1788C.png>

GitHub:

<https://cdn-icons-png.flaticon.com/512/25/25231.png>

Word:

<https://i.pinimg.com/originals/be/6a/8e/be6a8e264fa97e41b64d789fc0ba9a53.png>

Gráfico:

<https://dlegorreta.wordpress.com/tag/nltk/>

Pablo Medina Sosa

Fecha: 07/01/23

Versión: V1.0

