

Developing a Scraper that publishes data through an API

Desarrollo Aplicaciones para Ciencia de Datos

2º curso

Grado en Ciencia e Ingeniería de Datos

Escuela de Ingeniería Informática y Matemáticas

Universidad de Las Palmas de Gran Canaria

Resumen

En este proyecto desarrollamos un código que nos permite, mediante una API en la que interactúa el usuario, y una dirección url que he proporcionado yo manualmente en el mismo, obtener los primeros 25 hoteles de una ciudad y la ubicación, servicios, ratings y comentarios de uno elegido por el usuario de estos 25, devolviendo todo en formato JSON a la API.

Índice

1. Recursos utilizados	4
2. Diseño	4
3. Conclusiones	4
4. Líneas futuras	5
5. Bibliografía	5

1. Recursos utilizados

Para el desarrollo de este proyecto se ha utilizado como entorno de programación:

- IntelliJ

Herramienta de control de versiones:

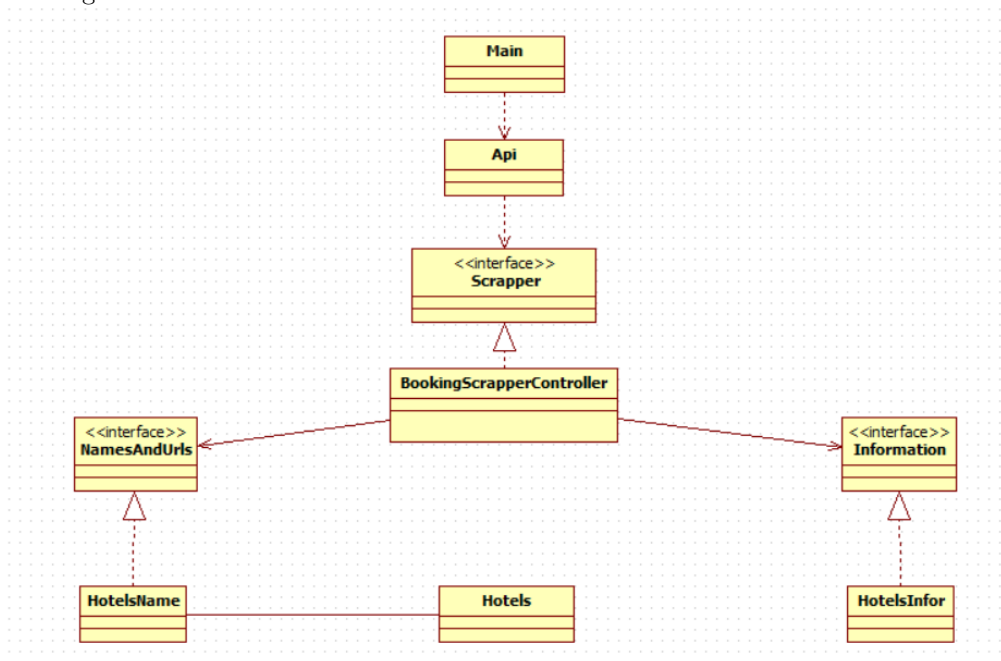
- Git
- GitHub

Herramienta de documentación:

- LaTeX

2. Diseño

Para el diseño de este proyecto se ha intentado seguir los principios SOLID y una arquitectura MVC; haciendo que cada clase solo sea responsable de una cosa, el uso de interfaces para encapsular métodos genéricos, modularidad de las clases y abstracción de todo lo que no fuese necesario para un algoritmo genérico.



3. Conclusiones

La principal lección que he aprendido durante el desarrollo de este proyecto ha sido la utilización de las interfaces, que aunque no prometo que estén bien del todo o que se podría mejorar, las he empezado a entender mejor y he querido incluirlas sin salirme de lo necesario.

Si tuviese que volver a empezar este proyecto, antes de pasar a la parte del código pasaría un tiempo diseñando el diagrama de clases, para mejorar desde un principio el encapsulamiento, abstracción y coherencia, y de esta forma facilitarme la programación.

4. Líneas futuras

Poder comercializar este producto es algo complicado, ya que a nivel legal el scraping no está muy bien visto. Sin embargo, en el caso de que tuviésemos permisos para hacerlo, deberíamos implementar en el proyecto la función de poder decidir la ciudad de la que queremos obtener los datos de los hoteles, utilizar las interfaces que tenemos para añadir otros sistemas de búsqueda sin ser Booking, por ejemplo tripAdvisor o parecidos. Y quizás, como última idea, utilizar bases de datos para el almacenamiento de información, de esta forma no tener que estar continuamente extrayendo toda esta información de sus páginas.

5. Bibliografía

<https://aep2223.ulpgc.es/course/view.php?id=2098>

<https://www.booking.com/index.es.html>

Jesús María Matos Torres

12 de enero de 2023