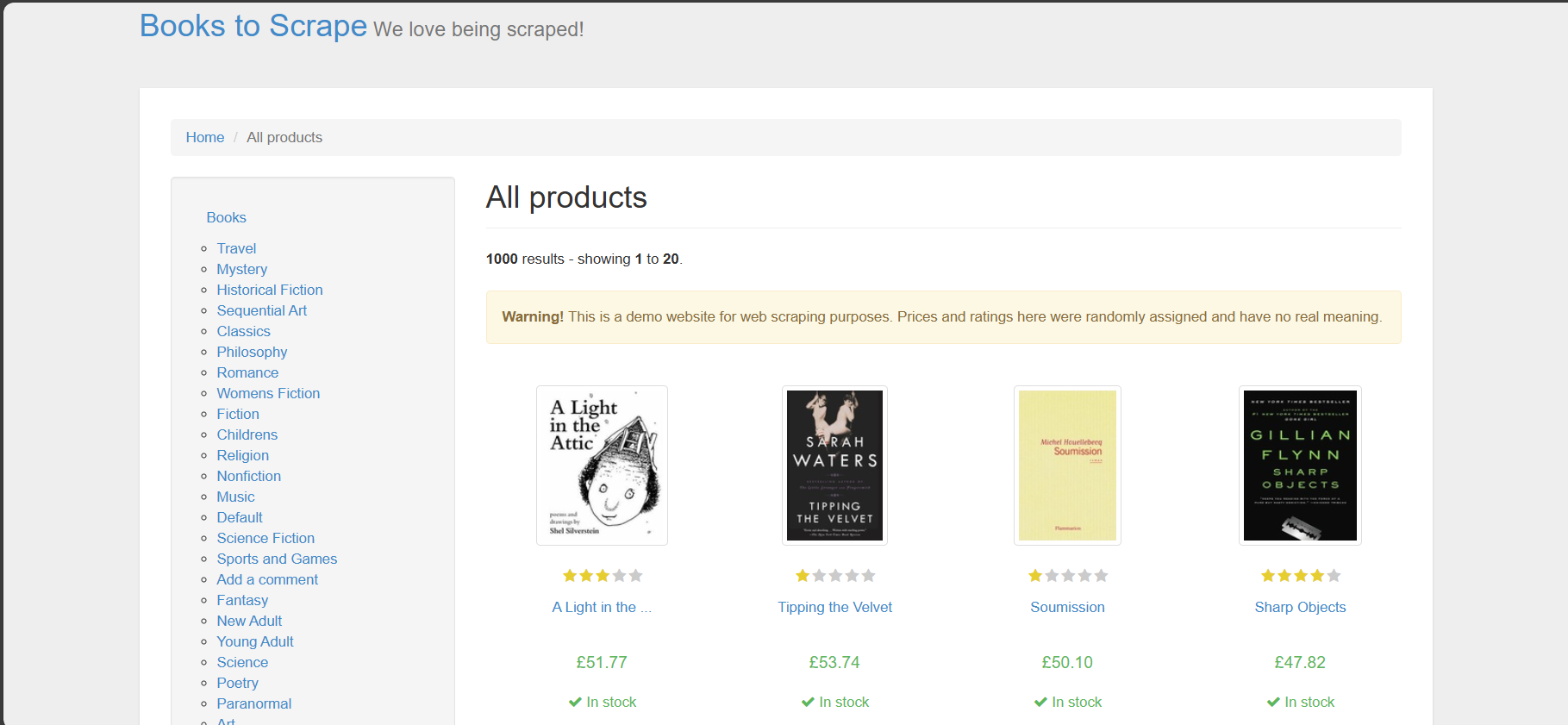
**Personal Assignment – Week02: CRAWLER DATA**

**Link project:** <https://github.com/dhlananhh/BigData>

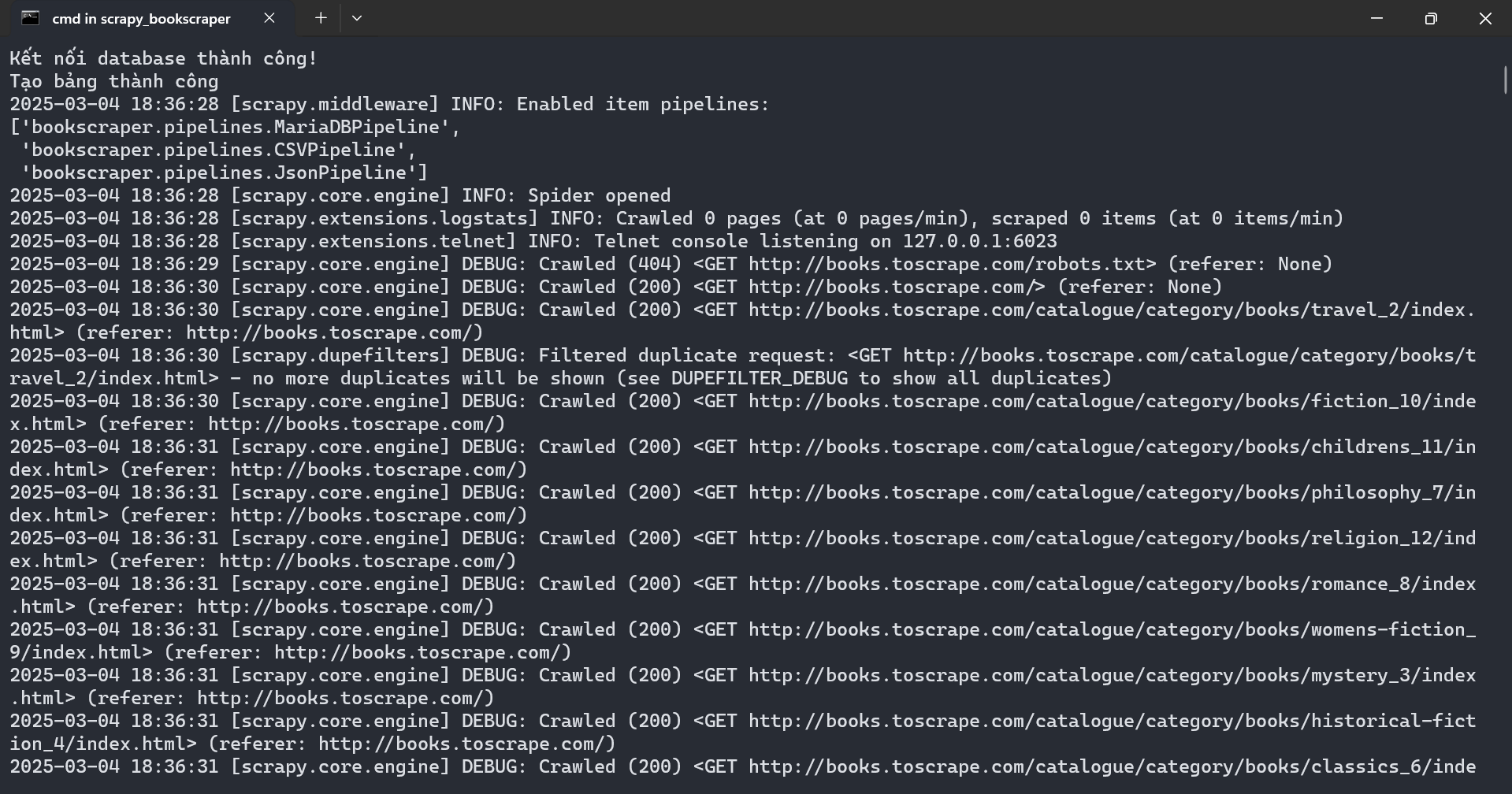
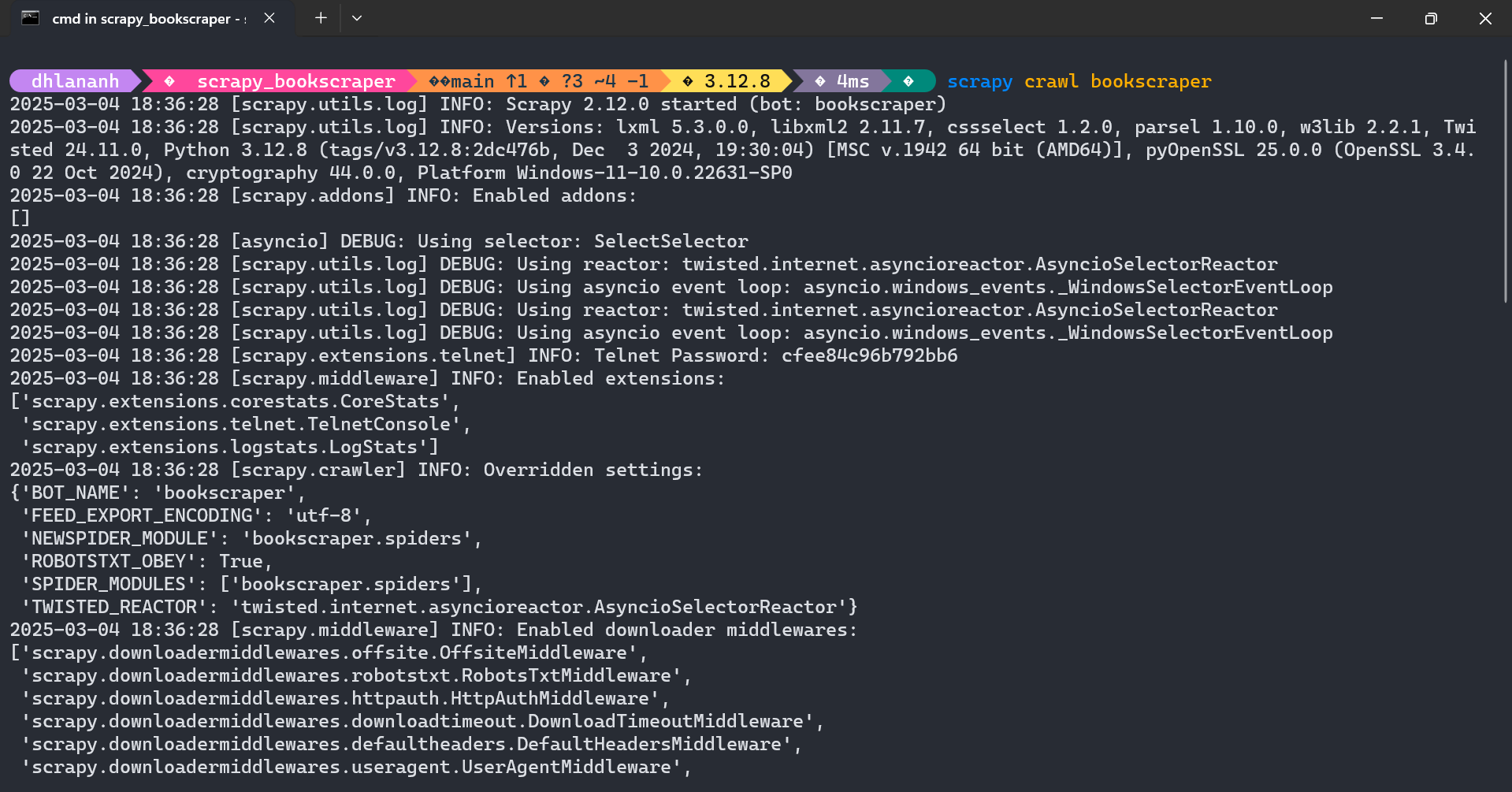
**Đề tài:** Xây dựng hệ thống thu thập dữ liệu tự động từ fictional bookstore website, sử dụng Docker và Scrapy

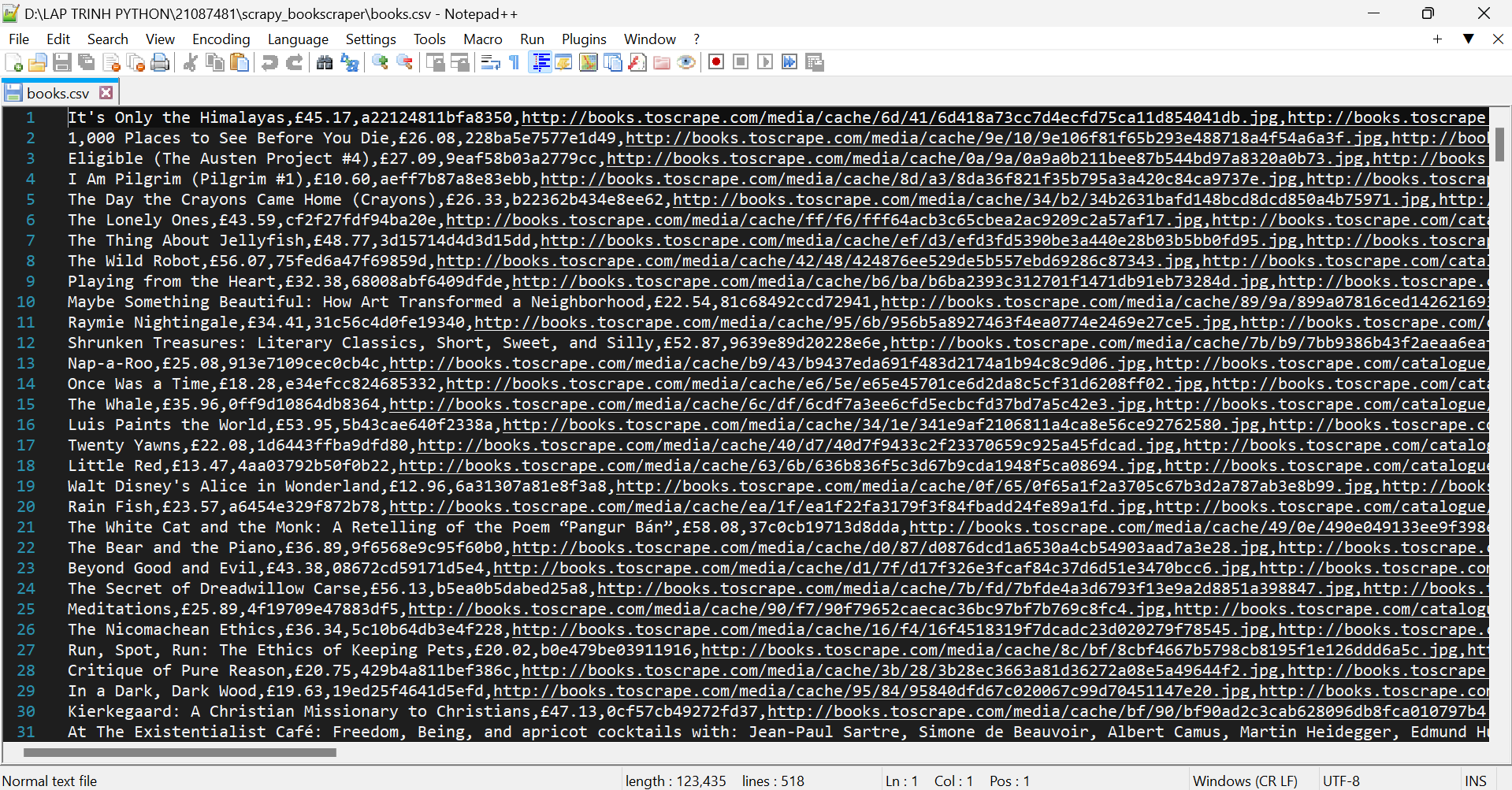
Website cụ thể mà em sử dụng để crawl data: <https://books.toscrape.com/>

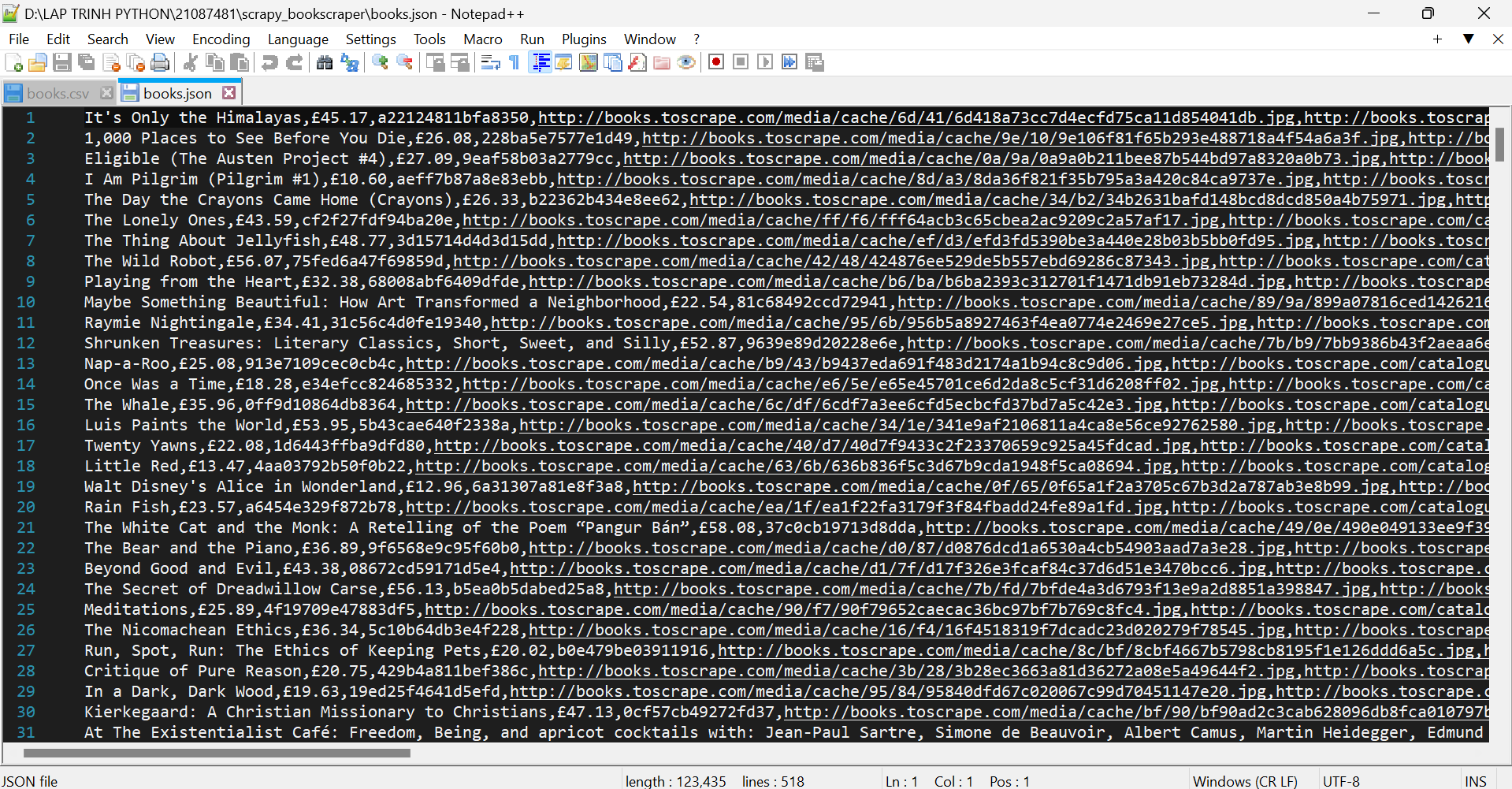
- Công nghệ sử dụng: Python, Scrapy, Docker

- Các sample data mà em thu thập được như sau: file csv, file json, database mariadb, database mongodb

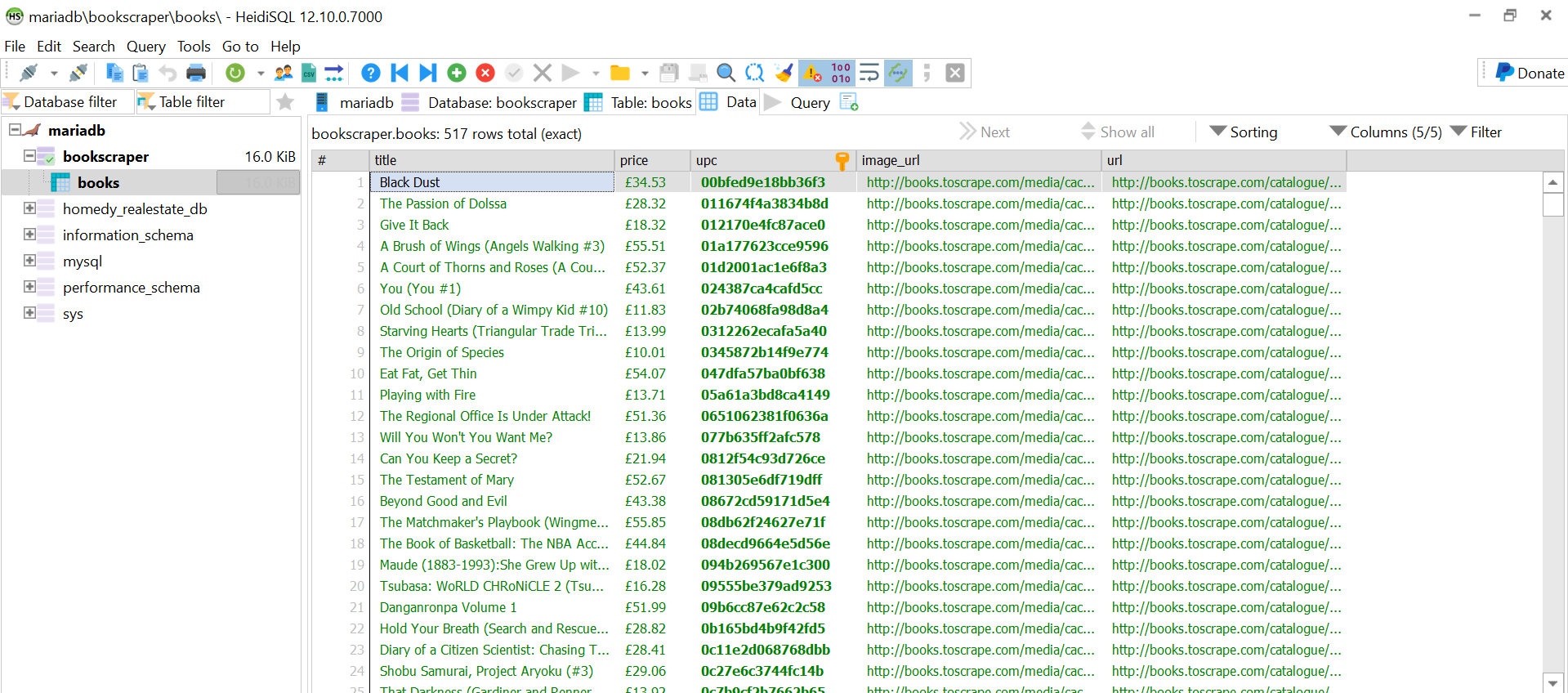
Chạy lệnh spider trên Terminal và lưu trữ dữ liệu dưới dạng JSON, CSV, MARIADB:



CSV

JSON

MARIADB



- Triển khai trên Docker:

