

## Overview

- Collect.py : Scrapes the website <https://books.toscrape.com> and stores it in a MongoDB database hosted on MongoDB Atlas
- Prepare.py: Runs queries on the collected data and generates an XML (books\_data.xml) which is verified against an XSD (books\_data.xsd).
- Access.py: Applies XSLT to the XML to generate a HTML file

Steps to run the project:

1. `python -m venv venv`
2. For MacOS: `source venv/bin/activate`  
For Windows: `venv\Scripts\activate`
3. `pip install -r requirements.txt`
4. Add your MongoDB connection string in the collect.py file before running it  
**Note: If you run the script with the already existing MongoDB connection string then duplicate data might get stored in the database since the current database already contains the scraped data.**
5. Then run the prepare.py file to generate XML
6. Run access.py file to apply XSLT to XML which generates a HTML file (books.html)
7. Open books.html in a browser to view the results