This repository contains the source code for a book recommendation system based on Scrapy and Word2Vec. The system collects book data from Amazon and uses Word2Vec to recommend similar books to users based on their queries. The system is implemented in Python and uses the Gensim library for Word2Vec.

## Files

The repository contains three files:

\* \*\*bookscraper.py:\*\* This file contains the code for scraping book data from Amazon. The scraper uses a rotating list of user agents to avoid being blocked by Amazon, and it delays between requests to avoid overloading the Amazon servers. The scraper also handles errors gracefully, such as when a book page or review page is not found.

\* \*\*database.py:\*\* This file contains the code for creating the database schema for the system. The database schema stores the book data scraped by the `bookscraper.py` file, as well as the Word2Vec model trained on the book data.

\* \*\*nlp\_book\_rec.py:\*\* This file contains the code for the book recommendation system. The system uses the Word2Vec model to find books that are similar to the user's query. It then ranks the similar books based on their similarity to the user's query and their rating on Amazon. Finally, it returns the top ranked books to the user as recommendations.

## Article explaining how the project works

I have written an article where I explain how the book recommendation system works in more detail. You can read the article here: [Link to article]

## Feedback and suggestions

If you have any feedback or suggestions, please feel free to leave a comment on the GitHub repository or contact me directly.

## License

This repository is licensed under the MIT License.

## Additional note

\* The `booklens.py` file can be configured to use different Word2Vec model parameters by changing the `VECTOR\_SIZE`, `WINDOW`, `MIN\_COUNT`, and `SG` variables.