Progress Report

[Free Topic] Search Engine for Movie Adaptations of Books

1. Which tasks have been completed?
    a. Filter unrelated information and crawl useful data from IMDB and Good Reads:
        for movies: url, title, writers, short description, top cast, storyline
        for books: url, title, writers, description, characters
    b. Install Elasticsearch with Docker, set up container, store movie and book documents in database and create index
    c. Writing testing queries using Elasticsearch Query DSL, especially "query string" for retrieving search results and "more like this query" for retrieving similar documents, use Kibana dev tool to view query result

2. Which tasks are pending?
    a. Build the front end, plan to use Flask templates for simplicity
    b. Enhance the accuracy of recommendation system by customizing ranking of "more like this query"
    c. Evaluate retrieving results against a test set with known book-to-movie relationship

3. Are you facing any challenges?
    a. Docker and ElasticSearch are new to me. So, I spent some time learning about configurations and Elasticsearch DSL.
    b. There's no publicly available book-to-movie adaptation dataset. I have to crawl data from either user-created IMDB lists or IMDB search results with keyword "adaptation". Data cleaning and manual labelling is required.