Project Title : Automated Book Publication Workflow

Objective:

Create a system to fetch content from a web URL, apply an AI-driven "spin" to chapters, allow multiple human-in-the-loop iterations. Store the spun versions in a vector Database.

Workflow:

1. **Web Scrapping**: Playwright

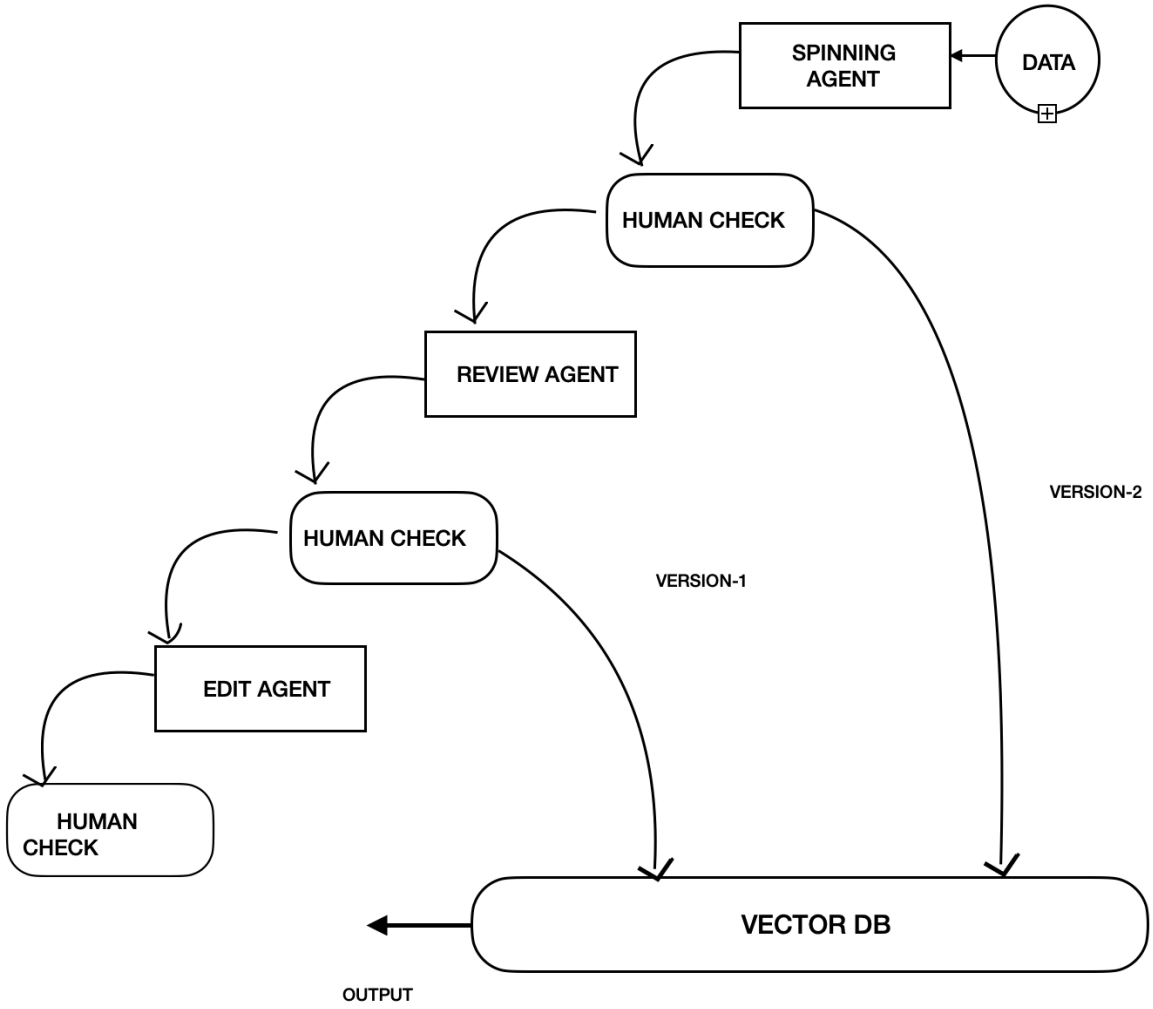
    • Launches chromium in headless mode.
    • Navigates to the given URL.
    • Scrapes the <p> tags inside the in which the content of the book are present (div.mw-parser-output)
    • Process and store the content inside a variable
    • close the browser

2. **Langchain Worflows and Chains with storage in Chroma**DB: Langchain, Langgraph,Groq (LLM), Human-in-the-loop, ChromaDB

    • initialized the llm (groq) with the respective API Key.
    • created prompt templates for each of the following workflows:
        1. Spinning Agent
        2. Reviewing Agent
        3. Editor Agent
    • created a state for the whole graphical chain
    • defined the nodes:
        1. Spin node
        2. Review node
        3. Edit node
    • Added human-in-the-loop using edit_with_user() function.
    • Added theese nodes to the builder.
    • invoked the graph to run the workflows.

3. **Function to view the collections and data of the chromadb.**

Flow Diagram:



Technologies Used:

1. **Primary Developement Language**: Python
2. **Web Scrapping**: Playwright
3. **Agentic workflows and Chains**: Langchain, Langgraph
4. **LLM :** Groq
5. **Database:** ChromaDB