Book Recommender using NLP

Carsten Lydeking

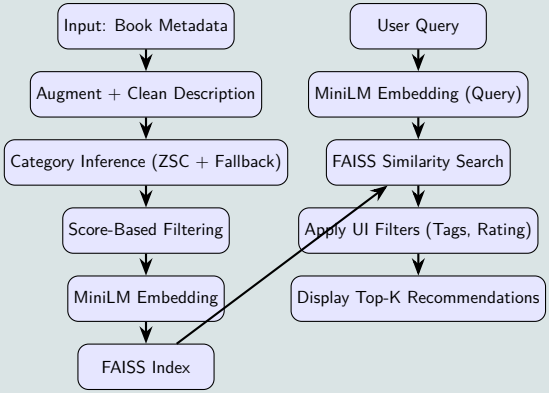
Zealand Business College

Oral Exam – AI and ML, 4th Semester

- Increasing demand for privacy-preserving, local-first ML applications.
- Typical recommender systems rely on cloud APIs and user profiles.
- Goal: explore feasibility of a fully offline, content-based book recommender system.
- Research question:
  How can a local ML model be used to recommend books based on natural language descriptions?

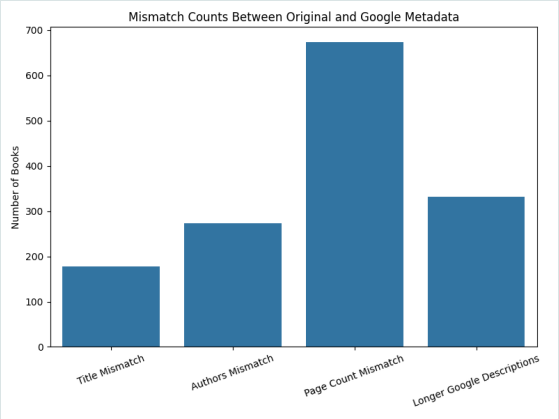
Modular, fully local processing pipeline:

- Data cleaning and augmentation
- Category inference via zero-shot classification + fallback
- Sentence embedding with MiniLM
- Fast vector similarity search with FAISS
- Offline UI built with Streamlit



Original dataset $\sim$ 6800 books

- Missing or inconsistent fields (authors, categories, descriptions)
- Very short or low-quality descriptions
- Category noise across sources
- OpenLibrary and Google Books API used to enrich metadata
- Rows with < 9 words in description removed
- Final dataset: 5160 high-confidence books



- Zero-shot classification with BART-MNLI
- 13 candidate categories defined
- Fallback keyword rules added for weak predictions
- Per-category metrics calculated:
  - Precision
  - Recall
  - F1-score
- Final filtering based on confidence thresholds:
  - description_length $\geq$ 200 chars
  - avg_score $\geq$ 0.2
  - max_score $\geq$ 0.4