Calypso

A semantic book recommender that uncovers the "vibe" of your next favorite read with Vector Embeddings and AI.

» submit a suggestion · » report a bug · » contact

Forks repo not found Stars repo not found

CI issues repo not found last commit repo not found license repo not found contributions welcome

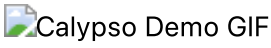

Table of Content

About

A "vibe matcher" for books, integrating modern AI/ML workflows (Python/FastAPI) with consumer-facing web applications (Next.js). Calypso uses semantic search to understand the emotional context of a user's request. By leveraging vector embeddings (`all-MiniLM-L6-v2`) and the `Hardcover.app` API, it allows users to search for natural language queries like *"A sci-fi about loneliness in space"* or *"A mystery that feels like a rainy day in London"* and retrieve statistically relevant matches.

:hammer_and_wrench: Tech Stack

Development Tools

- ☑ **Visual Studio Code** : IDE
- ☑ **Git & GitHub** : Version control
- ☑ **Postman** : API endpoint testing
- ☑ **Swagger UI** : API documentation & interactive testing

Backend (AI & API)

- ☑ **FastAPI/Python** : High-performance web framework
- ☑ **LangChain** : Orchestration for AI models
- ☑ **Pinecone** : Vector Database for semantic indexing
- ☑ **Hardcover API** : Live metadata fetch (GraphQL)

- ☑ **Sentence Transformers** : Local embedding generation

➕ Frontend

- ☑ **Next.js 14+** : React
- ☑ **TypeScript** : Static type checking & interfaces
- ☑ **Tailwind CSS** : Utility-first styling system
- ☑ **Axios/Fetch** : HTTP client for API requests

➕ DevOps

- ☑ **Vercel** : Frontend Deployment
- ☑ **Render / Railway** : Backend Deployment
- ☑ **Kaggle** : Dataset sourcing

🍄 Features

➕ Semantic Search Engine

- ☑ **Natural Language Querying:** Users can type sentences describing a plot or feeling, not just keywords.
- ☑ **Vector-Based Retrieval:** Uses cosine similarity to find books with matching "vibes" in the vector space.

📊 Hybrid Data Pipeline

➕ Static Ingestion (Current Phase)

Our core search functionality relies on a pre-built vector index derived from a curated dataset.

- ☑ **Dataset:** Sourced from Kaggle (`7k-books-with-metadata`), containing titles, authors, descriptions, and thumbnails.
- ☑ **Preprocessing:** - Data cleaning via `pandas` to remove null values and standardize category tags.
  - "The Deep Clean" logic ensures no broken records enter the database.
- ☑ **Vectorization:** - We use **Sentence Transformers** (`all-MiniLM-L6-v2`) to convert book descriptions + titles into 384-dimensional dense vectors.
- ☑ **Indexing:** - Vectors are upserted to a **Pinecone Serverless Index** (`calypso-books`) using Cosine Similarity.
  - Metadata (Title, Author, Description) is stored alongside vectors for single-shot retrieval (avoiding a secondary database lookup).

➕ Live Enrichment (Future Phase)

To overcome the limitations of a static CSV (low-res images, outdated ratings), we are building a hybrid pipeline.

🌵 Documentation

Setup

1. **Clone the repository**

```
 git clone [https://github.com/librariium/calypso.git]
(https://github.com/librariium/calypso.git)
 cd calypso
```

2. **Backend**

```
cd backend
python3 -m venv venv
source venv/bin/activate
pip install -r requirements.txt
```

3. **Frontend**

```
cd ../frontend
npm install
```

Development

- Run Backend: cd backend && uvicorn main:app --reload
- Run Frontend: cd frontend && npm run dev
- API Docs: Access Swagger UI at http://localhost:8000/docs

Contributing

> 1. Fork the Project
> 2. Create your Branch (`git checkout -b my-branch`)
> 3. Commit your Changes (`git commit -m 'add my contribution'`)
> 4. Push to the Branch (`git push --set-upstream origin my-branch`)
> 5. Open a Pull Request

License

This project is licensed under LICENSE.