One-Stop Recommender Sytem using Vector Similarity Search

Aman Sirohi*, Vikhyat Bansal†, R Sriviswa‡, Souvik Gorain§
Department of Artificial Intelligence Engineering
Amrita Vishwa Vidyapeetham
{cb.en.u4aie21003, cb.en.u4aie21076, cb.en.u4aie21046, cb.en.u4aie21065}@cb.students.amrita.edu,

***Abstract*—The increasing volume of digital content in diverse entertainment domains such as movies, anime, and music poses significant challenges for personalized recommendation systems. This paper introduces ”One Stop Recommender System,” a unified platform designed to deliver tailored recommendations across these domains. Leveraging a Retrieval-Augmented pipeline, our approach combines the strengths of traditional retrieval-based methods with the creativity of generative models to enhance recommendation accuracy and user satisfaction. The system utilizes datasets from TMDB for movies, a Kaggle-sourced anime dataset, and a music dataset from Kaggle to ensure comprehensive coverage and relevance. A dedicated website complements the system, featuring an intuitive landing page, home page, and category-specific recommendation pages to streamline user interactions. This work demonstrates the potential of Retrieval-augmented Vector Similarity Search in building scalable, multi-domain recommendation systems and its application in creating seamless, engaging user experiences.**

***Index Terms*—Retrieval-Augmented Generation, Vector Similarity Search, recommendation, scalable, movies , anime, music**

I. Introduction

The exponential growth of digital content across streaming platforms has revolutionized the way users access entertainment. With an ever-expanding array of movies, anime, and music available at their fingertips, users often find themselves overwhelmed by the sheer volume of choices. This phenomenon, known as ”choice overload,” necessitates the development of advanced recommendation systems to provide personalized suggestions, enabling users to discover content that aligns with their unique tastes and preferences efficiently. Recommendation systems have become integral to enhancing user satisfaction and engagement, making them indispensable tools in the entertainment industry. Traditional recommendation systems have focused on single-domain applications, such as movie or music recommendations, often leveraging collaborative filtering, content-based filtering, or hybrid methods. While effective in isolated domains, these systems lack the flexibility and scalability needed to cater to diverse user preferences across multiple entertainment categories. The increasing convergence of user interests across movies, anime, and music highlights the need for a unified platform capable of delivering cross-domain recommendations in a seamless and personalized manner. To address this gap, we present the ”One Stop Recommender System,” a comprehensive platform designed to deliver tailored recommendations across three key entertainment domains: movies, anime, and music. Our system employs a Retrieval-Augmented Generation (RAG) pipeline, which combines retrieval-based techniques for precision with the generative capabilities of large language models to enhance contextual relevance and diversity in recommendations. This hybrid approach enables the system to capture nuanced user preferences and adapt dynamically to varied input queries. The system is built upon robust datasets that provide rich and diverse information for generating high-quality recommendations. For movies, we use the TMDB dataset, renowned for its detailed metadata and popularity in recommendation research. The anime recommendations are powered by a Kaggle-sourced dataset compiled from multiple online sources, ensuring comprehensive coverage of the anime domain. Music recommendations are based on a Kaggle dataset, offering a wide variety of tracks across genres and styles. To deliver an intuitive user experience, we have developed a feature-rich website that serves as the interface for the One Stop Recommender System. The website includes a landing page introducing the platform, a home page summarizing key features, and category-specific pages dedicated to movies, anime, and music recommendations. This design ensures that users can seamlessly navigate between domains and access personalized suggestions tailored to their preferences. The ”One Stop Recommender System” not only demonstrates the potential of RAG in building scalable and versatile recommendation systems but also addresses the limitations of traditional single-domain engines. By unifying multiple entertainment categories into a single platform, this work contributes to the advancement of AI-driven personalization, offering an innovative solution for addressing the complexities of modern content discovery in the entertainment industry.

II. Methodology

The implementation of the ”One Stop Recommender System” involved the development of a unified platform that delivers tailored recommendations for movies, anime, and music using a Retrieval-Augmented Generation (RAG) pipeline. The methodology followed a structured approach comprising data collection, preprocessing, model design, system integration, and website development to create an end-to-end solution.

A. Data Collection

To ensure comprehensive and diverse recommendations, three datasets were utilized:

- Movies: The TMDB dataset, known for its extensive metadata on movies, including genres, ratings, cast, and crew information.
- Anime: A Kaggle-sourced dataset compiled from multiple online sources, providing information on anime titles, genres, ratings, and user reviews.
- Music: A Kaggle dataset containing a variety of tracks, including metadata such as genres, artist names, and popularity scores.

These datasets were chosen for their richness and relevance to their respective domains.

B. Data Preprocessing

Each dataset was preprocessed to ensure compatibility with the recommendation pipeline. Steps included:

- Cleaning: Handling missing values, removing duplicates, and normalizing text-based fields.
- Feature Engineering: Extracting and standardizing key features such as genres, popularity scores, and user ratings.
- Tokenization: For text fields like movie overviews, anime descriptions, and song lyrics, tokenization and embedding techniques were applied to prepare the data for downstream tasks.

C. RAG-Based Recommendation System

The core of the ”One Stop Recommender System” is a Retrieval-Augmented Generation (RAG) pipeline, which combines retrieval-based methods and generative models to deliver highly relevant and context-aware recommendations.

- Retrieval Component: The retrieval component is responsible for fetching relevant items from the datasets based on user input or historical preferences.
  - Dataset Vectorization: Each dataset (movies, anime, music) was preprocessed to generate vector representations of its items. For this, embeddings were created using pre-trained or fine-tuned models specific to each domain:
    - Movies: Contextual embeddings generated using models like BERT or Sentence-BERT on metadata fields such as titles, overviews, and genres.
    - Anime: Similar embeddings were created using the Kaggle anime dataset, focusing on fields like descriptions, genres, and user reviews.
    - Music: Metadata such as track names, artist information, and genre tags were vectorized for music recommendations.
  - Indexing: A vector index was created for efficient similarity searches. Tools like FAISS (Facebook AI Similarity Search) were used to enable rapid retrieval of the top-N most relevant items from each dataset.
  - Query Representation: When a user provides input (e.g., a movie they like, keywords, or preferences), the query is converted into a vector representation using the same embedding model. This ensures consistency between query vectors and the item vectors in the index.
  - Similarity Matching: The system computes the similarity between the query vector and item vectors using cosine similarity. The top-N matches are passed to the generation stage.
- Generation Component (although this is not necessary for our project, instead we are using vector similarity search): The generation component enhances the recommendations by contextualizing them based on user input and preferences.
  - Model Architecture: A pre-trained language model, such as OpenAI’s GPT or similar large language models, was fine-tuned on domain-specific data. The model is trained to generate coherent and personalized recommendations based on prompts and the retrieved items.
  - Prompt Design: Carefully designed prompts guide the generative model to produce recommendations. The prompts include:
    - Contextual information from the retrieval stage (e.g., the titles and descriptions of top-N retrieved items).
    - User preferences or additional input (e.g., genres, keywords).
    - Task-specific instructions to generate concise and engaging outputs.
  - Personalization: The generative model adapts to user preferences by learning patterns from interaction data. This allows it to suggest recommendations that align closely with individual tastes, even when explicit queries are not provided.
  - Filtering and Post-Processing: After generation, the recommendations are filtered to avoid redundancy and ensure diversity in suggestions. Post-processing ensures that only high-quality and relevant items are presented to the user.
- Integration of Retrieval and Generation: The RAG pipeline operates in a tightly integrated loop:
  - Retrieval-to-Generation Transition: The retrieved items serve as the input context for the generative model. This ensures that the recommendations are grounded in high-relevance items from the dataset.
  - Generation Feedback Loop: The output of the generative model is used to refine future retrievals by dynamically updating user profiles or embeddings.

D. System Integration

The recommendation engine has been integrated into a centralized backend system, ensuring seamless interaction between the retrieval and generation modules with the preprocessed datasets. A robust API layer has been developed to handle user queries, retrieve recommendations, and serve them efficiently to the front-end interface. Additionally, the backend incorporates a SQLite3 database, streamlining the storage and access of datasets, thus enhancing the overall system's performance and scalability.

III. Website Development

A user-friendly website was created to serve as the interface for the recommender system. The website comprises:

- Landing Page: Introducing users to the platform and its features.
- Home Page: Providing an overview of the system and quick access to category-specific recommendations
- Category-Specific Pages: Dedicated pages for movies, anime, and music, each displaying personalized recommendations. The pages feature intuitive layouts and filters for refining suggestions based on user preferences.

IV. Testing and Optimization

The system underwent extensive testing to ensure reliability and performance:

- A/B Testing: Conducted to compare different configurations of the RAG pipeline and refine the model's parameters.
- User Feedback: Iterative feedback from test users was incorporated to improve the website's usability and recommendation quality.

V. Deployment

The final system was deployed as a fully functional web-based application, accessible to users seeking personalized recommendations across movies, anime, and music. The architecture is designed for scalability, enabling the addition of new domains or features in the future. This structured methodology ensures that the ”One Stop Recommender System” delivers high-quality, context-aware recommendations while providing a seamless user experience across all entertainment domains.

VI. Analysis of Results

The results of the project are demonstrated through snapshots showcasing various functionalities of the ”One Stop Recommender System” website. These images provide a visual representation of the system's interface and its key features :

VII. Conclusion

The ”One Stop Recommender System” demonstrates the potential of combining Retrieval-Augmented Generation (RAG) techniques with robust dataset integration to create a unified recommendation platform spanning movies, anime, and music. By leveraging the precision of retrieval models and the contextual creativity of generative models, the system provides highly relevant, personalized, and engaging recommendations, addressing the growing need for efficient content discovery in the entertainment industry.

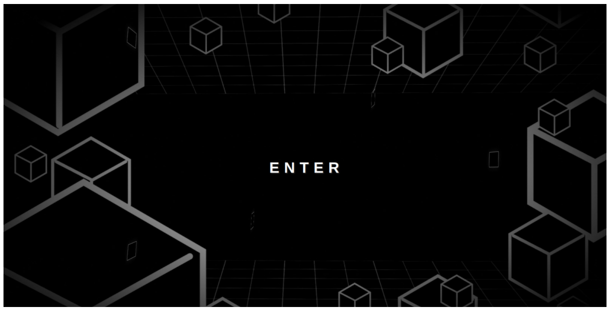


Fig. 1: The landing page of the website, providing an intuitive and visually appealing introduction to the system. Press ENTER to explore.

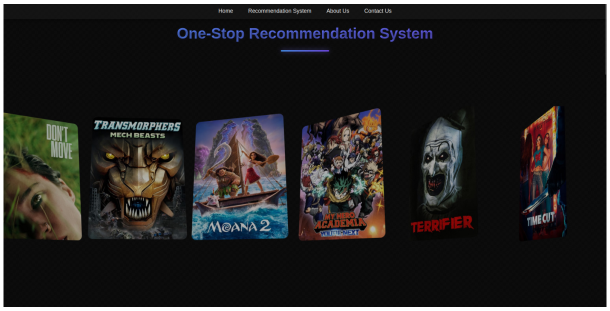
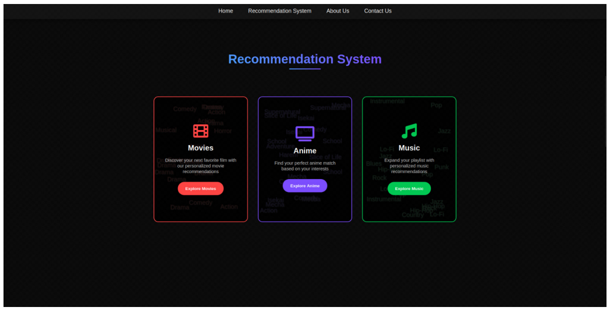


Fig. 2: The home page of the recommender system, featuring a dynamic gallery of movie posters with a parallax effect for aesthetic interactivity.

The implementation of a user-friendly website further enhances accessibility, enabling seamless navigation across categories and intuitive interaction with the recommendations. The modular design of the system ensures scalability, allowing for the inclusion of additional domains and features in the future.

This work highlights the strengths of RAG pipelines in building cross-domain recommendation systems and demonstrates their practical application in real-world scenarios. It contributes to advancing the state of AI-driven personalization, offering a versatile and innovative approach to solving the challenges of choice overload in digital entertainment. Future directions include expanding the system to incorporate user feedback for iterative learning, exploring multimodal datasets for richer recommendations, and enhancing computational efficiency for real-time performance.



Recommendation System

Movies

Discover your next favorite film with our personalized movie recommendations

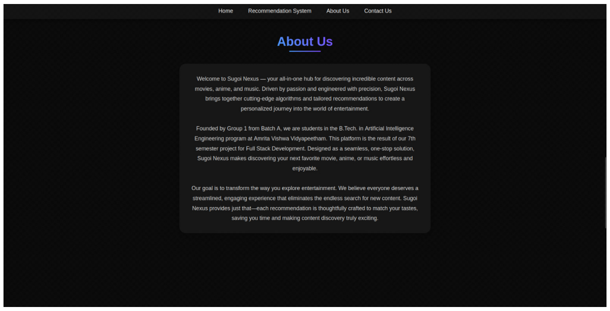
Anime

Find your perfect anime match based on your interests

Music

Expand your playlist with personalized music recommendations

Fig. 3: Different categories where the recommender system provides recommendations. Each category tile has subtle genre keywords floating in the background.



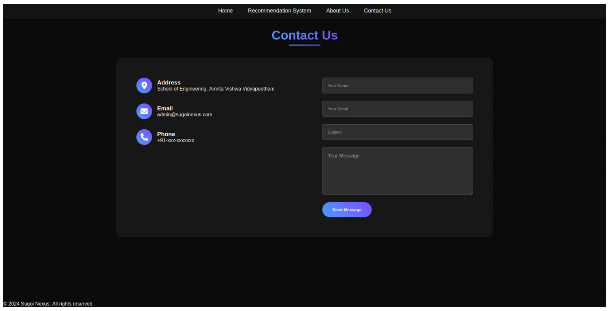
About Us

Welcome to Sugoi Nexus — your all-in-one hub for discovering incredible content across movies, anime, and music. Driven by passion and engineered with precision, Sugoi Nexus brings together cutting-edge algorithms and tailored recommendations to create a personalized journey into the world of entertainment.

Founded by Group 1 from Batch A, we are students in the B.Tech. in Artificial Intelligence Engineering program at Amrita Vishwa Vidyapeetham. This platform is the result of our 7th semester project for Full Stack Development. Designed as a seamless, one-stop solution, Sugoi Nexus makes discovering your next favorite movie, anime, or music effortless and enjoyable.

Our goal is to transform the way you explore entertainment. We believe everyone deserves a streamlined, engaging experience that eliminates the endless search for new content. Sugoi Nexus provides just that—each recommendation is thoughtfully crafted to match your tastes, saving you time and making content discovery truly exciting.

Fig. 4: About Us Section



Contact Us

Address
School of Engineering, Amrita Vishwa Vidyapeetham

Email
admin@sugoinexus.com

Phone
+91-xxx-xxxxxxx

Your Name

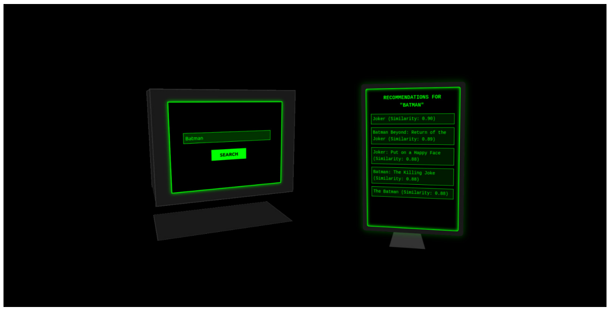
Your Email

Subject

Your Message

© 2024 Sugoi Nexus. All rights reserved.

Fig. 5: Contact Us Section


Batman

RECOMMENDATIONS FOR "BATMAN"

Joker (Similarity: 0.90)

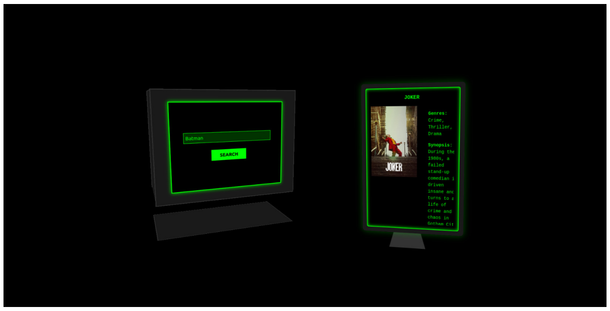
Batman Beyond: Return of the Joker (Similarity: 0.89)

Joker: Put on a Happy Face (Similarity: 0.88)

Batman: The Killing Joke (Similarity: 0.88)

The Batman (Similarity: 0.88)

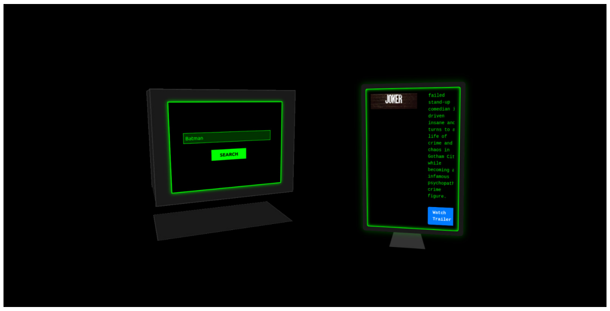
Fig. 6: Figure denotes recommendations being made along with similarity values .


Batman

Genres:
Crime,
Thriller,
Drama

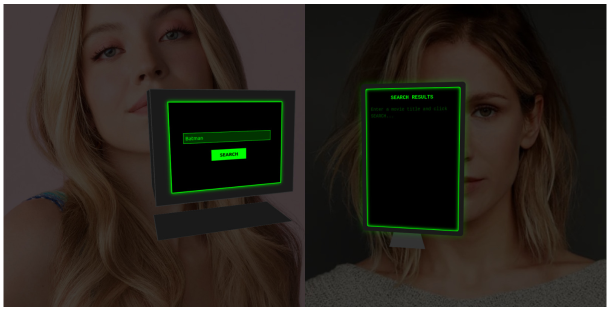
Synopsis:
During the
1980s, a
failed
stand-up
comedian i
driven
insane anc
turns to a
life of
crime and
chaos in
Gotham Cit

Fig. 7: Extra information about the recommended movie selected by the user is being rendered such as a title, poster, genres and a brief synopsis.


Batman

failed
stand-up
comedian i
driven
insane anc
turns to a
life of
crime and
chaos in
Gotham Cit
while
becoming a
infamous
psychopath
crime
figure.

Fig. 8: Figure shows an option to watch trailer on YouTube.


Batman

SEARCH RESULTS

Enter a movie title and click SEARCH...

Fig. 9: Figure shows a non functional feature where a dynamic collage animation of 2 actors will be rendered.