Project Proposal: Planner for Family Activities

Patiharn & Liangkobkit
1012871867, patiharn.liang@gmail.com

https://colab.research.google.com/drive/1vh_D2th_JABNn6x3Yv3K8aDz92ni-TSg?usp=sharing

September 25, 2025

Introduction

This project develops a Google-style hybrid search planner that retrieves and ranks age-appropriate family activities. Combining BM25, Sentence-BERT embeddings, FAISS, and a neural re-ranker, the system personalizes results, respects calendar constraints, and exports weekly timetables, adapting modern deep learning retrieval methods to practical family scheduling.

Illustration

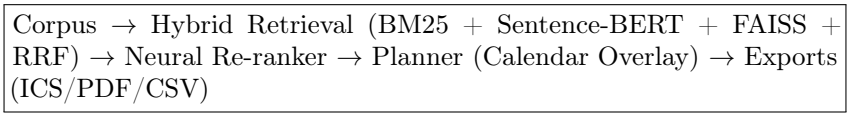

Figure 1: System overview diagram.

Background & Related Work

We build on foundational methods in information retrieval and modern semantic search. **BM25** provides strong keyword-based baselines *[Robertson & Zaragoza, 2009]*. **Sentence-BERT** enables dense embeddings for semantic similarity *[Reimers & Gurevych, 2019]*. **FAISS** supports efficient approximate nearest-neighbor search at scale *[Johnson et al., 2017]*. **MMR** introduces diversity-aware ranking *[Carbonell & Goldstein, 1998]*. For content, we collect activity data from trusted family and education resources: *Raising Children Network*, *Active for Life*, *Oxford Owl*, *Positive Action*, *Escape Room Geeks* and more.

Data Processing

Corpus. The dataset with more than 100 activities is gathered with the following columns: `title, age_min, age_max, duration_mins, tags, cost,`

`indoor_outdoor, season, materials_needed, how_to_play, players, parent_caution`.
Data Cleaning. The text columns are turned to be list columns in the dataset, such as `materials_needed` and `how_to_play`.

Architecture

Hybrid Search (Google-style, small scale). The system combines *BM25* (sparse) and *Sentence-BERT* (dense) with *FAISS* ANN; results are merged via *Reciprocal Rank Fusion* (RRF) and finalized by a *neural re-ranker*.

Sparse Retrieval (BM25). Keyword-based scoring over the corpus to capture exact intent and rare-term importance.

Dense Retrieval (Sentence-BERT). Sentence-level embeddings for semantic similarity so paraphrases/synonyms are retrieved even without exact term overlap.

Approximate Nearest Neighbor (FAISS). Efficient large-scale vector search for dense retrieval, enabling fast top-K candidate generation.

Fusion (RRF). Combines BM25 and dense candidate lists into a single slate, trading off precision/recall robustly across queries.

Neural Re-ranker (feed-forward / cross-encoder–style). Reorders the fused top-K using features such as *AgeFit*, *PreferenceMatch*, *LoadBalance*, *ContextFit*, *Practicality*, plus diversity via *MMR*.

Planner & Visualization. Greedy slotter that respects calendar busy blocks, visualized on a calendar overlay and integrated with read-only Google Calendar data.

Frontend. Next.js interface with toggles (*Exclude on Export*) and filters.

Exports. ICS (planned activities only, or including events from Google Calendar data), CSV, and PDF weekly grid.

Baseline Model

The baseline is a **BM25-only ranker** with simple metadata filters (age, duration). It does not use semantic embeddings, personalization, or diversity constraints.

Ethical Considerations

Privacy. Calendar integration is read-only; exports can exclude personal events.
Child Safety. Activities are age-filtered, and flagged for supervision.
Bias. We monitor cost and age balance; MMR ensures diversity and fair exposure.

References

[1] S. E. Robertson and H. Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in IR*, 3(4):333–389, 2009.

[2] N. Reimers and I. Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *EMNLP*, 2019.

[3] J. Johnson, M. Douze, H. Jégou. Billion-scale similarity search with GPUs (FAISS). *arXiv:1702.08734*, 2017.

[4] J. Carbonell and J. Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, 1998.

[5] Active for Life. 57 fun physical activities to do with kids aged 2 to 4. `https://activeforlife.com/57-fun-physical-activities-to-do-with-kids-aged-2-to-4/`, 2024.

[6] Active for Life. 50 indoor physical activities for kids. `https://activeforlife.com/50-indoor-physical-activities-for-kids/`, 2024.

[7] Oxford Owl. Fun learning activities for 6–7 year olds. `https://home.oxfordowl.co.uk/kids-activities/learning-activities-age-6-7/`, 2025.

[8] Positive Action. 20 evidence-based social skills activities and games for kids. `https://www.positiveaction.net/blog/social-skills-activities-and-games-for-kids`, 2024.

[9] Escape Room Geeks. 30 indoor games for kids to burn out energy at home. `https://escaperoomgeeks.com/indoor-games-for-kids/`, 2024.