

Temporal Information Retrieval and Question Answering using Retrieval Augmented Generation

[Group 11]: Sonith Bingi, Manoj Arulmurugan, Prasad Jawale, Gauri Patki

Brief Recap

- Traditional RAG systems ignore temporal context, leading to outdated or temporally mismatched evidence retrieval
- Temporal RAG integrates time-aware retrieval to ensure evidence aligns with query timeframes and reflects when facts are valid
- MRAG framework addresses this through modular architecture combining dense retrieval with semantic-temporal hybrid re-ranking

Baseline Methods

- Standard RAG model without temporal features and compare it with a temporal RAG
- MRAG benchmarks on the TempRAGEval dataset

Experiment Results

Metric	K=1	K=5	K=10	K=20
Hit@K	0.2069	0.5268	0.7211	0.8199
MRR@K	0.2069	0.3109	0.3383	0.3451
Recall@K	0.0422	0.1228	0.1927	0.2236

Table 1: MRAG Re-ranking performance on TempRAGEval (fixed pool).

Experiment setup

- Used ATLAS 2021 Wikipedia corpus, **mine positives** by matching TempRAGEval gold evidence sentences into passages (multi-pattern Aho–Corasick), then add BM25 **hard negatives** (lexically similar but non-evidence) to keep the pool.
- Embed passages and queries with **Contriever** and search the corpus with **FAISS** (inner-product over L2-normalized vectors) to get top-100 candidates per question. This is the baseline candidate generator MRAG will re-rank.
- From the Contriever top-100: build **question-focused signals** (keyword/date cues & LLM brief summaries), score evidence at the **sentence level**, aggregate to a **passage score** (max over sentences), and **re-rank** all candidates; this implements the semantic-temporal hybrid re-ranking described in MRAG and evaluate on the test set (TempRAGEval).

Findings

- Achieved Hit@20 of 82%, MRR@20 of 0.35, and Recall@20 of 22% on TempRAGEval
- Hybrid approach particularly benefited questions with evolving entities or recurrent events where semantic retrieval struggles.

Future Plan

Future Extension

- 1 - Fine-tune Contriever using a temporal hard mining on the corpus using year-based negatives using a triplet loss + train on ms marco triplets to avoid catastrophic forgetting
- 2 - Implement metadata-aware retrieval using publication timestamps and validity windows
- Integrate enhanced temporal retriever into MRAG and evaluate on temporal QA benchmarks like ChroniclingAmericaQA

Contribution Statement

- Sonith Bingi (25%) - Modular-RAG corpora processing and temporal hard mining
- Manoj Arulmurugan (25%) - Modular RAG re-ranking with hybrid method
- Prasad Jawale (25%) - Dense Passage Retrieval\Contriever Retrieval
- Gauri Patki (25%) - Metadata-aware retrieval\Contriever Retrieval

Metric	Base Model	T5-Split [Finetuned]
Hit@1	0.817	0.857
Hit@5	0.917	0.944
Hit@10	0.941	0.958
Hit@20	0.957	0.976