

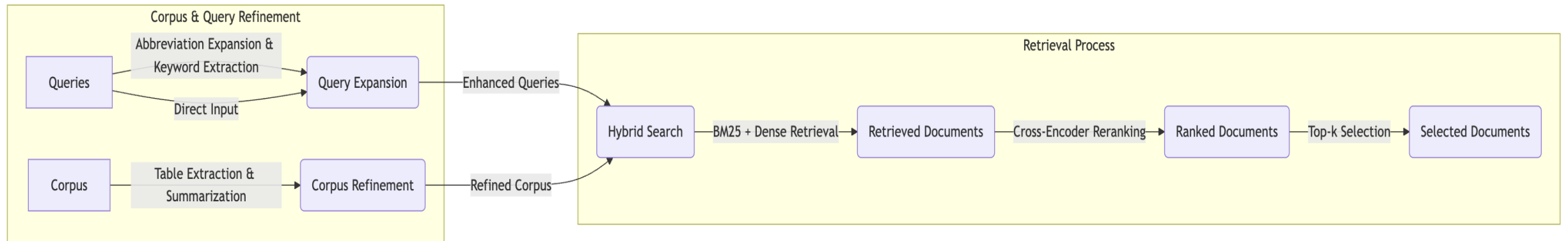
[ISE261] FinanceRAG Challenge

Main Strategy: Hybrid Search & Query Expansions

Team 이박터(KAIST 극예술연구회) | 2024.12.14
정재웅(ISE,20220627), 이민준(CEE), 김동언(Freshman)

Overview: (1) Architecture

- Objective: **Develop a RAG system for financial question answering.**
- Our Strategy:
 - Query Expansion(Abbreviation, Keyword)
 - Hybrid Search (BM25 + Dense)
 - Prompt Engineering(CoT, Persona-based approach)



Overview: (2) Evaluation metrics by models

type	Task	Method	expansion	Embed/Dense	Reranking	Eval
Passage	FinDER	hybrid	4o-mini/keyword/company accord	text-embedding-3-small	gemma2-lightweight	0.62575
Passage	FinDER	hybrid	4o-mini/keyword	text-embedding-3-small	gemma2	-
Passage	FinDER	hybrid	4o-mini/keyword	rag-matryoshka	gemma2-lightweight	-
Passage	FinQABench	hybrid	4o-mini/keyword	rag-matryoshka	gemma2-lightweight	0.9254
Passage	FinanceBench	hybrid	4o-mini/keyword/company accord	rag-matryoshka	gemma2-lightweight	0.93226
Tabular	TATQA	dense (L2)	4o-mini/keyword, replace table	voyage-3	voyageai/rerank-2	0.62031
Tabular	TATQA	dense (L2)	4o-mini/keyword, replace table	text-embedding-3-small	voyageai/rerank-2	0.62251
Tabular	TATQA	dense (L2)	4o-mini/keyword, replace table	voyage-3	colbert-ir/colbertv2.0	0.46127
Tabular	TATQA	dense (L2)	4o-mini/keyword, replace table	text-embedding-3-small	colbert-ir/colbertv2.0	-
Tabular	FinQA	dense (L2)	4o-mini/keyword, replace table	voyage-3	voyageai/rerank-2	0.76698
Tabular	FinQA	dense (L2)	4o-mini/keyword, replace table	text-embedding-3-small	voyageai/rerank-2	0.6015
Tabular	FinQA	dense (L2)	4o-mini/keyword, replace table	voyage-3	colbert-ir/colbertv2.0	0.50837
Tabular	FinQA	dense (L2)	4o-mini/keyword, replace table	text-embedding-3-small	colbert-ir/colbertv2.0	-
Tabular	ConvFinQA	dense (L2)	4o-mini/keyword, replace table	voyage-3	voyageai/rerank-2	0.76345
Tabular	ConvFinQA	dense (L2)	4o-mini/keyword, replace table	text-embedding-3-small	voyageai/rerank-2	-
Tabular	ConvFinQA	dense (L2)	4o-mini/keyword, replace table	voyage-3	colbert-ir/colbertv2.0	0.54802
Tabular	ConvFinQA	dense (L2)	4o-mini/keyword, replace table	text-embedding-3-small	colbert-ir/colbertv2.0	-
Tabular	MultiHiertt	hybrid	4o-mini/keyword, replace table	text-embedding-3-large	colbert-ir/colbertv2.0	0.5012
Tabular	MultiHiertt	hybrid	4o-mini/keyword, replace table	text-embedding-3-large	voyageai/rerank-2	0.3231

Problem Statement/Dataset

- **Challenges:**

- Complex financial language and terminologies
- Integration of numerical and textual data
- Scalability with large datasets

- **Dataset Description**

- Passage Tasks: FinDER, FinQABench, FinanceBench
- Tabular Tasks: TATQA, FinQA, ConvFinQA, **MultiHiertt**
- Data Format: corpus.jsonl, queries.jsonl, and TSV labels



Hardest Task!!

Dataset Analysis

Task	#-of-queries	Percentage
FinDER	216	4.6%
FinQABench	100	2.1%
FinanceBench	150	3.2%
TATQA	1663	35.6%
FinQA	1147	24.6%
ConvFinQA	421	9.0%
MultiHiertt	974	20.9%
Total	4671	100.0%

- Although the task is divided into 7 datasets, approximately 80% of the queries come from **tabular datasets**.
- For **passage tasks**, we adopted a simple strategy.
- For **tabular datasets**, we experimented with various methods to explore optimization directions.

Baseline Approach

- Analyzed the provided base code to identify its strengths and limitations.
- **Base Code Setup:**
 - Embeddings generated via Hugging Face models.
 - Sparse retrieval performed using BM25.
 - Cross-encoder-based reranking for improved accuracy.
- **Problems & Challenges**
 - Certain models did not run due to RAM limitations both on local machines and Colab A100 GPUs.
 - > Hugging face leaderboard의 상위 5개 model은 아예 안 돌아감.

Baseline Approach

- **Challenges:**

- Certain models did not run due to RAM limitations both on local machines and Colab A100 GPUs.

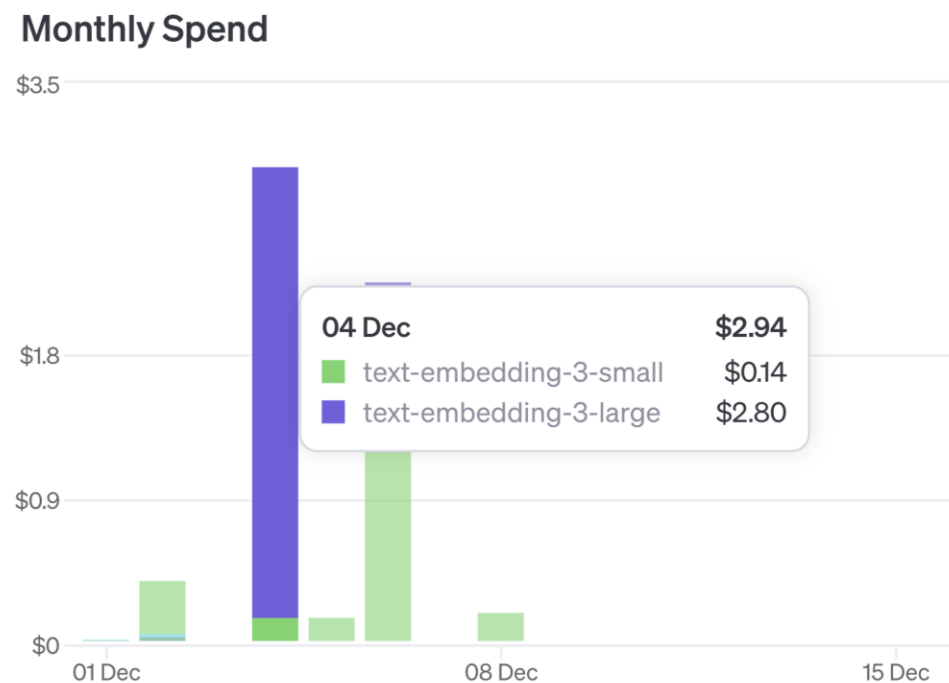
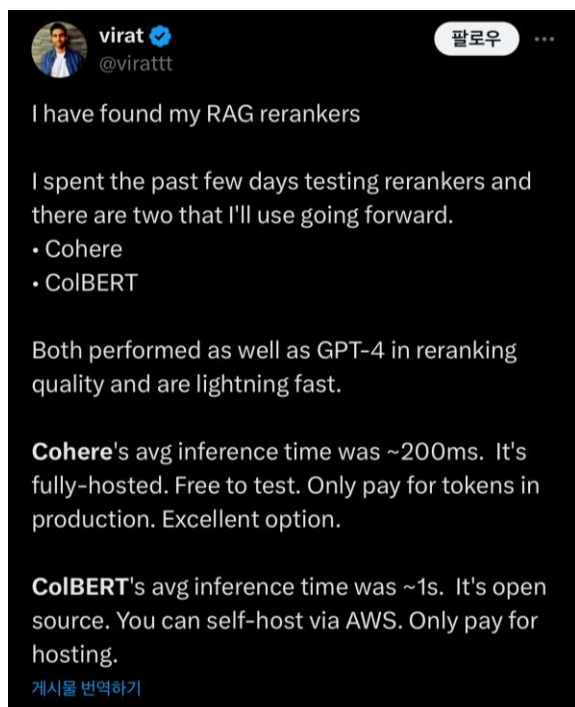
- **Alternative Approach:**

- Due to model limitations, switched to **OpenAI embeddings(small/large both)**
- Stored embeddings in a Vector Database (VectorDB) for efficient similarity search.
- Attempted to use state-of-the-art rerankers such as ColBERT and Cohere models

Baseline Approach

• Lesson Learned

- OpenAI's embeddings were more affordable than expected.
- We used **Cohere** and **ColBERT**, which were popular on platforms like Reddit and X, and they indeed performed well.



Key Processes and Goals

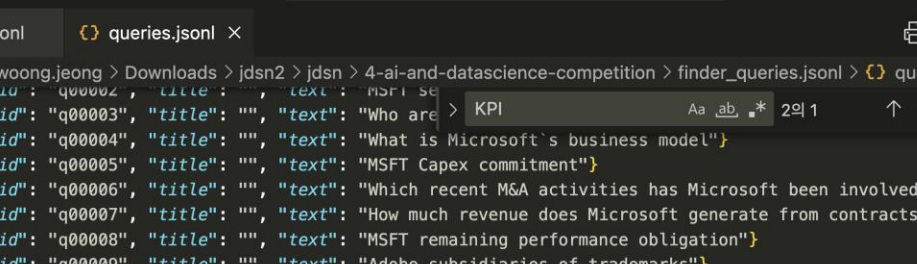
- **Key Processes:**

- **Query Expansion:** Enhanced queries using keywords, abbreviations, and corporate identifiers with ChatGPT API.
- **Corpus Refinement:** Preprocessed documents via summarization and table extraction to focus on key data.
- **Hybrid Search:** Combined sparse (BM25) and dense retrieval, balancing keyword and semantic relevance.
- **Reranking:** Refined top-10 results with cross-encoder models for better ranking accuracy.

- **Goals:**

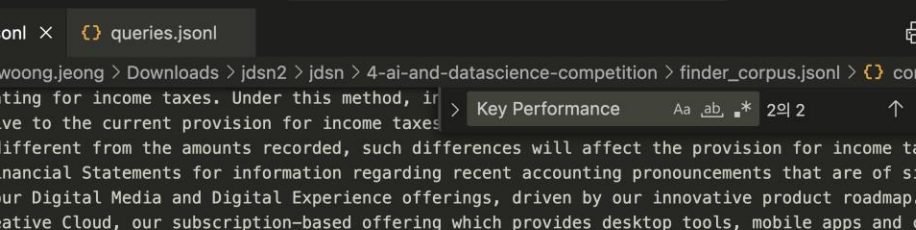
- Achieve high retrieval accuracy while maintaining computational efficiency.
- Optimize resource usage for both passage and tabular datasets.

Query Expansion - Data Eyeballing (1)



```
Users > jaewoong.jeong > Downloads > jdsn2 > jdsn > 4-ai-and-datascience-competition > finder_queries.jsonl > queries.jsonl
2  {"_id": "q00002", "title": "", "text": "MSFT SE"}
3  {"_id": "q00003", "title": "", "text": "Who are > KPI"}
4  {"_id": "q00004", "title": "", "text": "What is Microsoft's business model?"}
5  {"_id": "q00005", "title": "", "text": "MSFT Capex commitment"}
6  {"_id": "q00006", "title": "", "text": "Which recent M&A activities has Microsoft been involved in?"}
7  {"_id": "q00007", "title": "", "text": "How much revenue does Microsoft generate from contracts with customer"}
8  {"_id": "q00008", "title": "", "text": "MSFT remaining performance obligation?"}
9  {"_id": "q00009", "title": "", "text": "Adobe subsidiaries of trademarks?"}
10 {"_id": "q00010", "title": "", "text": "ADBE share repurchase?"}
11 {"_id": "q00011", "title": "", "text": "fully diluted shares outstanding ADBE?"}
12 {"_id": "q00012", "title": "", "text": "Who are the members of Adobe's management team?"}
13 {"_id": "q00013", "title": "", "text": "ADBE RPO?"}
14 {"_id": "q00014", "title": "", "text": "ADBE KPI?"}
15 {"_id": "q00015", "title": "", "text": "How are Coupang's KPIs?"}
16 {"_id": "q00016", "title": "", "text": "Coupang segment margin?"}
17 {"_id": "q00017", "title": "", "text": "CPNG capital expenditure?"}
18 {"_id": "q00018", "title": "", "text": "CPNG any recent M&A activities?"}
19 {"_id": "q00019", "title": "", "text": "When did Coupang's Farfetch consolidation start?"}
20 {"_id": "q00020", "title": "", "text": "What is FLC, and how is its revenue recognized by Coupang?"}
21 {"_id": "q00021", "title": "", "text": "When did new FLC contract begin CPNG?"}
22 {"_id": "q00022", "title": "", "text": "CPNG free cash flow?"}
23 {"_id": "q00023", "title": "", "text": "Class of Shares CPNG?"}
24 {"_id": "q00024", "title": "", "text": "any highlights from Linde's 2023 earnings result?"}
25 {"_id": "q00025", "title": "", "text": "Linde Segment breakdown?"}
```

Searching for "KPI" in Finder queries.



```
{
  "text": "nting for income taxes. Under this method, in",
  "start": 219,
  "end": 220,
  "context": "ive to the current provision for income taxes",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "different from the amounts recorded, such differences will affect the provision for income taxes and the",
  "start": 221,
  "end": 222,
  "context": "financial Statements for information regarding recent accounting pronouncements that are of significance",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "our Digital Media and Digital Experience offerings, driven by our innovative product roadmap. As we execu",
  "start": 223,
  "end": 224,
  "context": "eative Cloud, our subscription-based offering which provides desktop tools, mobile apps and cloud-based se",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "enterprises. We expect Creative Cloud will drive sustained long-term revenue growth through a continued e",
  "start": 225,
  "end": 226,
  "context": "levate awareness, consideration and purchase of subscriptions to our Creative Cloud offerings. These stra",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "s built around our Adobe Acrobat family of products, with a set of integrated mobile apps and cloud-based",
  "start": 227,
  "end": 228,
  "context": "we utilize a data-driven operating model (\u201cdata-driven\u201c) and our Adobe Experience Cloud solutions to",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "the key performance metric",
  "start": 229,
  "end": 230,
  "context": "our management uses to assess the health and trajectory of our overall Digital",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "eginning of the year and held constant throughout the year for measurement purposes. We calculate ARR as fo",
  "start": 231,
  "end": 232,
  "context": "d Services + Annual Creative ETLA Contract Value"}",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "ons and Services + Annual Document Cloud ETLA Contract Value"}",
  "start": 233,
  "end": 234,
  "context": "iscal 2023 was $12.37 billion, up from $10.98 billion at the end of fiscal 2022. Document Cloud ARR exiting",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "growth in ARR has positively affected our revenue growth. Creative revenue in fiscal 2023 was $11.52 bill",
  "start": 235,
  "end": 236,
  "context": "ressed by our Digital Experience segment. The Adobe Experience Cloud apps and services are designed to ma",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "mers across the following strategic growth pillars: \u201cData insights and audiences. Our products, inclu",
  "start": 237,
  "end": 238,
  "context": "officers and digital marketers, users of our Digital Experience solutions include advertisers, campaign ma",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "ur Digital Experience solutions, as well as an extensive ecosystem of partners, including marketing agenci",
  "start": 239,
  "end": 240,
  "context": "2023, up from $4.42 billion in fiscal 2022 which represents 11% year-over-year growth. Driving this growth",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": "nt, we are subject to risks and exposures from the evolving macroeconomic environment, including the effec",
  "start": 241,
  "end": 242,
  "context": "dictable as a result of our subscription-based business model, the broader implications of these macroecon",
  "category": "Key Performance",
  "score": 2.0,
  "source": "Adobe Acrobat",
  "document": "Adobe Acrobat",
  "page": 2,
  "text": ""
```

When searching for "KPI" in the Finder corpus, no results appeared, as it was listed under a different name, "Key Performance Metric."

Query Expansion - Data Eyeballing (2)

- **Findings**

- Queries contain financial abbreviations (e.g., KPI, M&A).
- In the corpus, these abbreviations are written in their expanded forms.
- Queries are often associated with specific companies, but even with hybrid search, the retrieved corpus may not match the company in the query.

- **Solution**

- **Expanded queries** and added additional **keywords** to improve alignment.
- **Inserted company tickers** into both corpus and queries to ensure retrieval is limited to the relevant company's corpus.
- Try: Applied logical expansion to enhance query-corpus matching accuracy.

Query Expansion: Code Implementation

- Rewrite the queries using gpt-4o-mini

```
client = OpenAI(api_key=API_KEY)

response = client.chat.completions.create(
    model="gpt-4o-mini",
    messages=[
        {"role": "system", "content": system_prompt},
        {"role": "user", "content": user_prompt}
    ],
    temperature=0
)
```

```
# Instructions for API calls
rewrite_instruction = (
    "Rewrite the following financial query for clarity and extract the key terms. "
    "Provide the result as:\n"
    "Rewritten Query: [Your output here]\n"
    "Keywords: [Comma-separated keywords]\n"
)

logical_expression_instruction = (
    "Convert the following financial query into a logical search expression using AND/OR operators. "
    "Use parentheses for grouping and focus on clarity.\n"
)

ticker_instruction = (
    "Extract the company name and ticker symbol from the following query. "
    "Provide the result as:\n"
    "Company Name: [Full company name]\n"
    "Ticker: [Ticker symbol].\n"
)
```

> Prompt used in our tasks

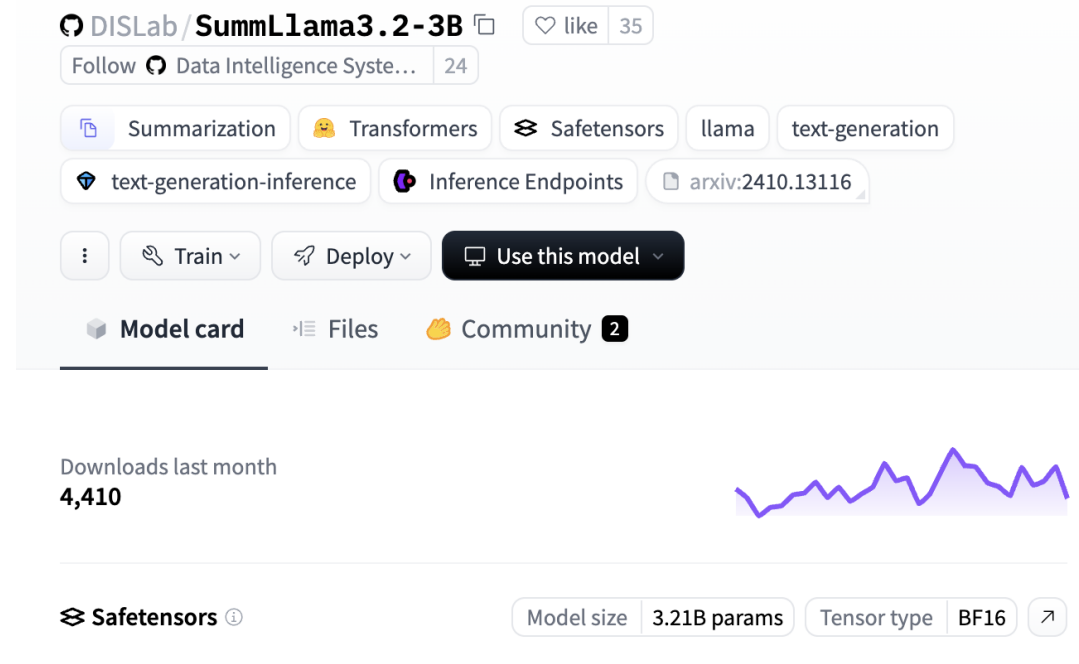
Query Expansion: Results

ID	Query	Modified Query + Keywords	Logical Expression	Name & Ticker
q00001	What are the service and product offerings from Microsoft	What products and services does Microsoft offer? Microsoft, products, services, offerings	(Microsoft AND (products OR services))	Microsoft(MSFT)
q00002	MSFT segment breakdown	Can you provide a breakdown of the different segments of Microsoft (MSFT)? MSFT, segment breakdown, Microsoft	(Microsoft OR MSFT) AND (breakdown OR segments OR divisions OR categories)	Microsoft(MSFT)
q00003	Who are Microsoft's key customers?	Who are the main customers of Microsoft? Microsoft, key customers, main customers	(main customers AND Microsoft)	Microsoft
q00004	What is Microsoft's business model	What is the business model of Microsoft? Microsoft, business model	(Business AND model AND Microsoft)	Microsoft (MSFT)
q00005	MSFT Capex commitment	What is Microsoft's capital expenditure (Capex) commitment? Microsoft, Capex, capital expenditure, commitment	(Microsoft AND "capital expenditure" OR Capex OR "Capex commitment")	Microsoft(MSFT)
q00006	Which recent M&A activities has Microsoft been involved in	What recent mergers and acquisitions (M&A) has Microsoft participated in? Microsoft, mergers and acquisitions, M&A, recent activities	(Microsoft AND (mergers OR acquisitions) AND (recent OR latest) AND (2023))	Microsoft (MSFT)
q00007	How much revenue does Microsoft generate from contracts with customers?	What is the revenue generated by Microsoft from customer contracts? Microsoft, revenue, customer contracts	(revenue AND generated AND Microsoft AND customer contracts)	Microsoft (MSFT)
q00008	MSFT remaining performance obligation	What is the remaining performance obligation for Microsoft (MSFT)? MSFT, remaining performance obligation, Microsoft	(remaining performance obligation) AND (Microsoft OR MSFT) AND (data up to October 2023)	Microsoft Corporation(MSFT)
q00009	Adobe subsidiaries of trademarks	What are the trademarks associated with Adobe's subsidiaries? Adobe, subsidiaries, trademarks	(Adobe AND trademarks) AND (subsidiaries)	Adobe
q00010	ADBE share repurchase	What are the details regarding Adobe's share repurchase program? ADBE, share repurchase, Adobe, program, details	(Adobe AND "share repurchase program" AND details)	Adobe Inc.(ADBE)

- Utilized 1) Query, 2) **Modified Query (with additional keywords)**, and 3) Logical Expressions.
- When using the Modified Query, evaluation scores showed a slight **improvement**.
- Logical Expressions, however, resulted in a **decline** in performance.

Corpus Refinement

- **Challenge:** Efficiently retrieve key information from large-scale documents.
- **Method:**
 - **Summarization:** Simplified text using models like 4o-mini.
 - **Table Extraction:** Applied to datasets such as MultiHiertt.
- **Result:** Reduced data size and improved efficiency.
- Also tried with summarization techniques developed by DISlab.



Hybrid Search Approach

- **Methodology**

- Sparse Retrieval:
Performed keyword-based search using BM25.
- Dense Retrieval:
Leveraged embedding models for semantic search.

- **Combination Strategy:**

- $S_h = \alpha S_s + (1 - \alpha) S_d$
 - Where S_h is the hybrid score, S_s is the sparse retrieval score, and S_d is the dense retrieval score, with α as the weighting parameter.
- Effectively captured both keyword relevance and semantic context of the text.

Reranking Strategy

- **Techniques**
- **Reranking with SOTA Models:**
Applied state-of-the-art models for reranking.
- **Results**
 - Enhanced precision in top-10 results.
 - Observed cases where performance improved by more than **2x** after applying reranking.
 - Reranking showed greater effectiveness in **tabular datasets** compared to passage datasets.

Additional: Prompt Engineering Inspiration



KyungHyun Cho · 팔로우 중

A mediocre combination of a mediocre scientist/physicist, a mediocre m...

2개월 · 🌐

here's my prompt of the day:

""""

You are an extremely thoughtful and verbose assistant who can help user extremely well with complicated problems that require sophisticated reasoning. You are mathematically enlightened and rely heavily on mathematical and statistical reasoning. You think aloud generally. You use tags strictly as instructed. You do not express your own views or beliefs beyond what's strictly necessary to follow the instruction. Your life depends on strictly following the user instruction.

Follow the user instruction below in a comprehensive and detailed way. You will try up to five attempts at following the user instruction. Each attempt must be marked with "<A_number>".

At each attempt, first write down your internal thoughts. Each step of your internal thoughts must be marked with "<T_number>" and must start with a short title describing the internal thought. This must include your draft response and its evaluation. After this, write your final response after "<R>". If your draft is not satisfactory, make sure to go back to an earlier internal thought step (mark it clearly with "<T_number>") where the mistake was made and retry from there on, as often as necessary.

- **Inspiration by KyungHyun Cho**

- **Chain of Thought (CoT):**

Applied step-by-step reasoning to break down complex financial queries for more structured responses.

- **Iterative Refinement:**

Used repeated attempts and self-reflection to refine answers, ensuring higher accuracy and relevance.

- **Persona-based Prompting:**

Assigned a specialized role (e.g., analytical assistant) to guide responses with clarity, consistency, and domain expertise.

- **Application to Our Objectives:**

Enhanced query understanding and retrieval accuracy by combining these techniques for financial datasets.

Prompt Engineering Technique

- Innovation: Persona-based approaches for prompt design.
- Key Technique: Chain-of-Thought (CoT) prompts.

```
prompt = """
```

```
You are an extremely thoughtful and detail-oriented financial expert specializing in logical search optimization. Your role is to transform financial queries into precise and clear search expressions. You rely on systematic reasoning and adhere strictly to the user's instructions.
```

```
Follow the user instruction below in a comprehensive and detailed way. You will attempt up to five iterations to follow the user instruction. Each attempt must be marked with `<A_number>`.
```

```
At each attempt:
```

- ```
1. Write down your **internal thoughts** step by step. Each step must be marked with `<T_number>` and start with a brief title describing the thought. This must include your draft response and its evaluation after `<R>`.
```
- ```
2. If your draft is not satisfactory, identify the point of error (mark it with `<T_number>`) and retry from there as often as necessary.
```
- ```
3. All attempts must be unique and not duplicate previous ones.
```

```
After producing up to five responses (`<R>`), compare them carefully based on their accuracy and clarity. Choose the best as the final response. Repeat the selected response verbatim under `<Final Response>`.
```

```
Instruction:
```

```
Convert the following financial query into a logical search expression using AND/OR operators. Use parentheses for grouping and focus on clarity.
```

```
"""
```

# Future Work & Limitation

- **Explore Advanced Models:**

Test more state-of-the-art models with optimized hardware to improve performance.

- **Multi-stage Reranking:**

We could implement both **initial** and **final reranking** stages to refine results more than once.

- **Additional Experiments:**

- Introduce **domain-specific embeddings** tailored for financial datasets.
- Apply **adaptive hybrid search weighting** to dynamically adjust alpha values for sparse and dense retrieval balance.