

Leitmotif:
Keyword-based Associated Stock and
Document Search Service

Ngseo Kim

Table of contents

1. What

1. What is it
2. What is keyword
3. What are some "Keyword-Company Relevance"
4. What kind of data sources are included
5. What are some comparisons or competitions

2. Why

1. Why is it needed
2. Why is it hard

3. Who

1. Who needs

4. How

1. How to evaluate

What is it?

- With simple keyword, system recommends the top-k related stock for it

What is keyword?

- "Keyword" in the service refers to word clusters representing a single concept when nouns or multiple words are combined.
- Single words could be common nouns like "치매", "음성인식" or proper nouns like "안철수", "이재명" or even company names like "쿠팡" "삼성전자"
- For multiple words like "플렉서블 디스플레이," both must be together to represent a specific concept as a keyword.
- Combinations like "황사 미세먼지" or "메르스 코로나" may not directly indicate a specific concept but can be keywords under a higher-level concept.

What are some "Keyword-Company Relevance"?

1. Company makes money with selected keyword
2. Company is related to keyword in a social issue perspective
3. Keyword is related to method or key idea of how company makes money
4. Some renowned analyst declared it is relevant

What kind of Data Sources are included?

1. Diverse news
2. Telegram messages
3. Corporate disclosure (DART)
4. Other communities for related stock analysis

What are some comparisons or competitions?

1. Toss securities
2. Fin-up
3. Judal

Why is it needed?

- Investment with related stock is very usual method in Korean stock market
- Need for finding stock with unknown keyword can help investors to cope with dynamic stock market

Why is it hard?

1. Diverse data sources
2. Real time streaming
3. Low system resource + Highly efficient search engine
4. Vague evaluation
5. Lack of Korean tokenizer support
-> Since not using Elasticsearch for resource efficiency

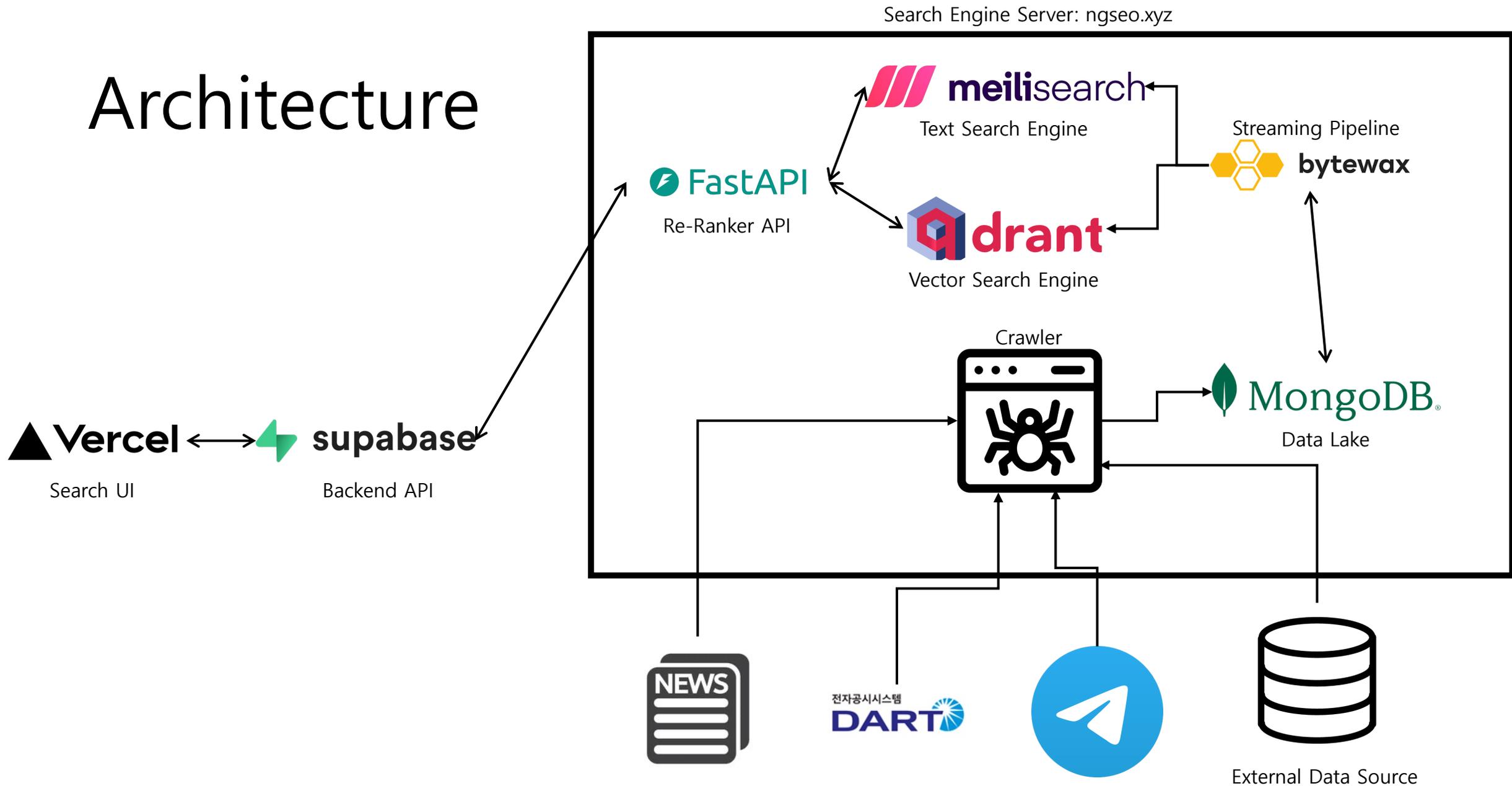
Who needs?

- Investors who needs information for related stock from unknown keyword

How to evaluate?

- Collect evaluation dataset from analyst report or opinion column
- The main goal of this system is to provide competitively high-quality recommendations for stocks related to keywords like 'analyst report'.

Architecture



Hybrid Search

<https://qdrant.tech/articles/hybrid-search/>

https://cookbook.openai.com/examples/search_reranking_with_cross-encoders