2025-2 졸업프로젝트

멀티모달 *RAG*와 성과 기반 분석을 활용한 식물 성장 분석 연구

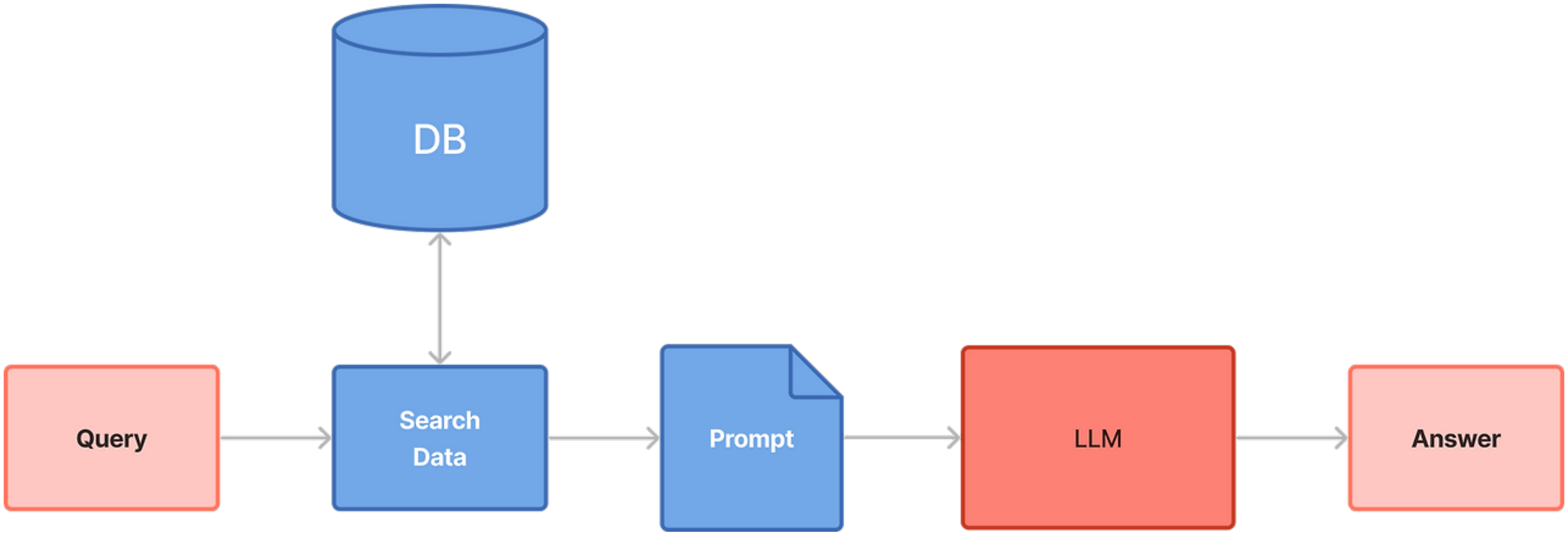
2025.09.24.

2021104417 우다현

멀티모달 *RAG*와 성과 기반 분석을 활용한 식물 성장 분석 연구

-> 결국은 RAG의 원리를 도입한 챗봇 응용

- RAG를 도입한 챗봇은 어떤 원리?



기술적 선택이 필요한 구간

★ 1 임베딩 생성

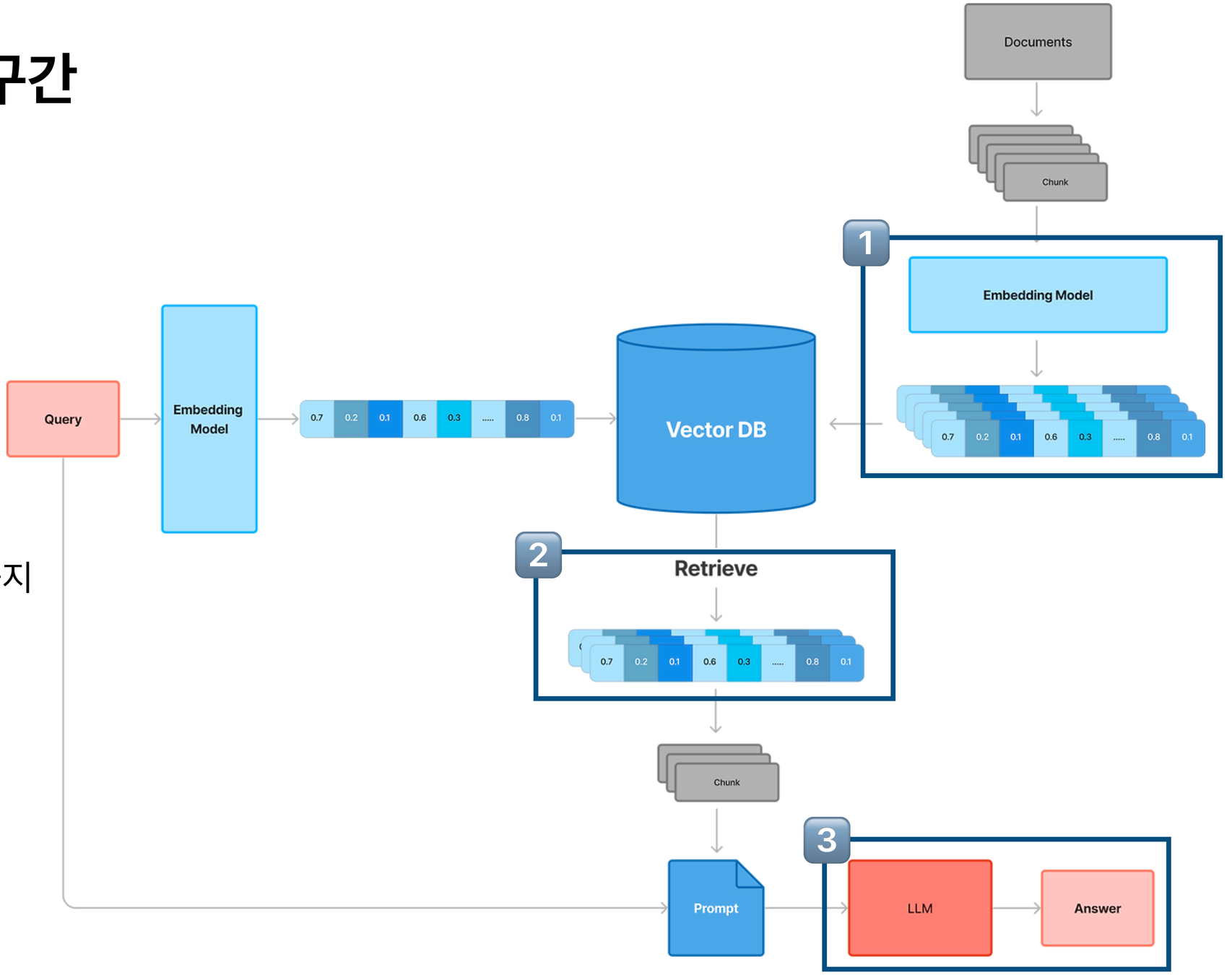
- 멀티모달 정보를 어떻게 임베딩할지

★ 2 retrieval 단계

- 검색할 때 어떤 evidence를 가져올지

3 답변 생성 단계

- 어떤 생성 모델을 사용할지

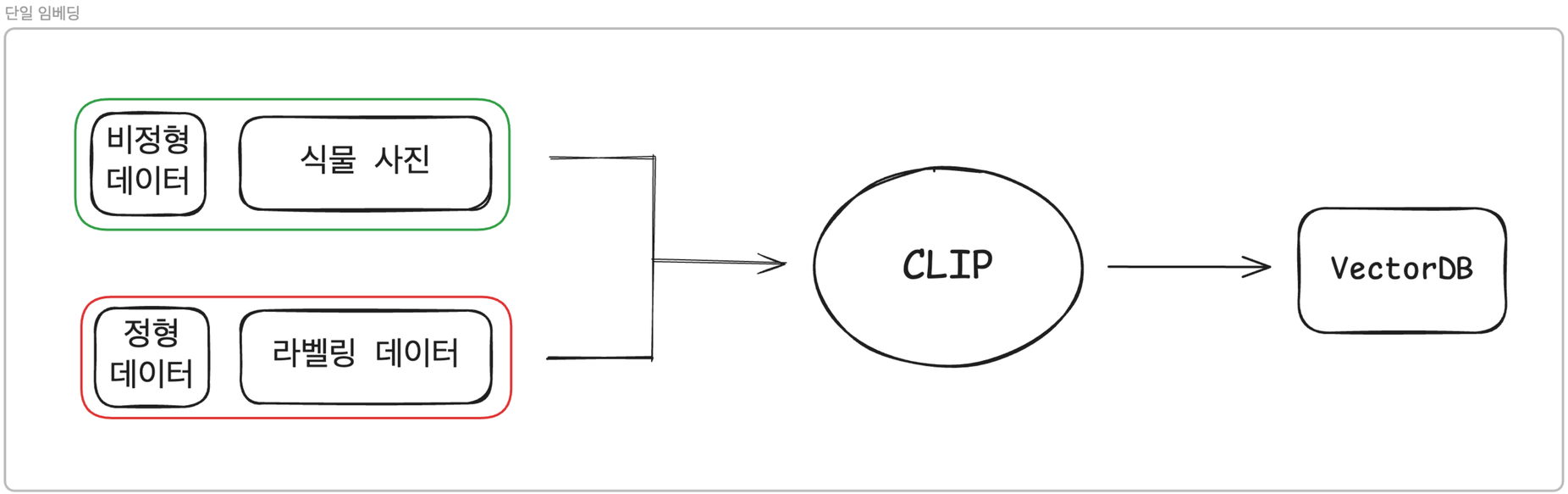


1️⃣ 임베딩 방법

근본적으로 다른 유형의 데이터인 **비정형 데이터**와 **정형 데이터**를 어떻게 효과적으로 통합하고 활용할 것인가

1. 단일 벡터 임베딩

- 하나의 모델을 사용하여 모든 종류의 데이터를 하나의 공유된 벡터 공간으로 변환하는 방식
- OpenAI의 CLIP(Contrastive Language-Image Pre-training)은 단일 임베딩 패러다임의 대표적인 사례



Google DeepMind 28 Aug 2025

1 단일 벡터 임베딩 모델의 한계

On the Theoretical Limitations of Embedding-Based Retrieval

Orion Weller*,1,2, Michael Boratko[1], Iftekhar Naim[1] and Jinhyuk Lee[1]
[1]Google DeepMind, [2]Johns Hopkins University

연구 배경

- 대부분의 RAG 시스템은 문서/이미지를 하나의 벡터로 임베딩 후 검색
- 하지만 “복잡한 정보 관계를 한 벡터에 담는 것이 가능한가?”에 대한 근본적 의문 존재

주요 결과

→ 벡터 차원 한계로 인해 모든 관계를 표현할 수 없음
→ 단순 질의(예: “누가 사과를 좋아하는가?”)에서도 최신 임베딩 모델들이 실패
→ 하나의 벡터에 모든 의미를 압축 → 정보 손실 불가피

대안

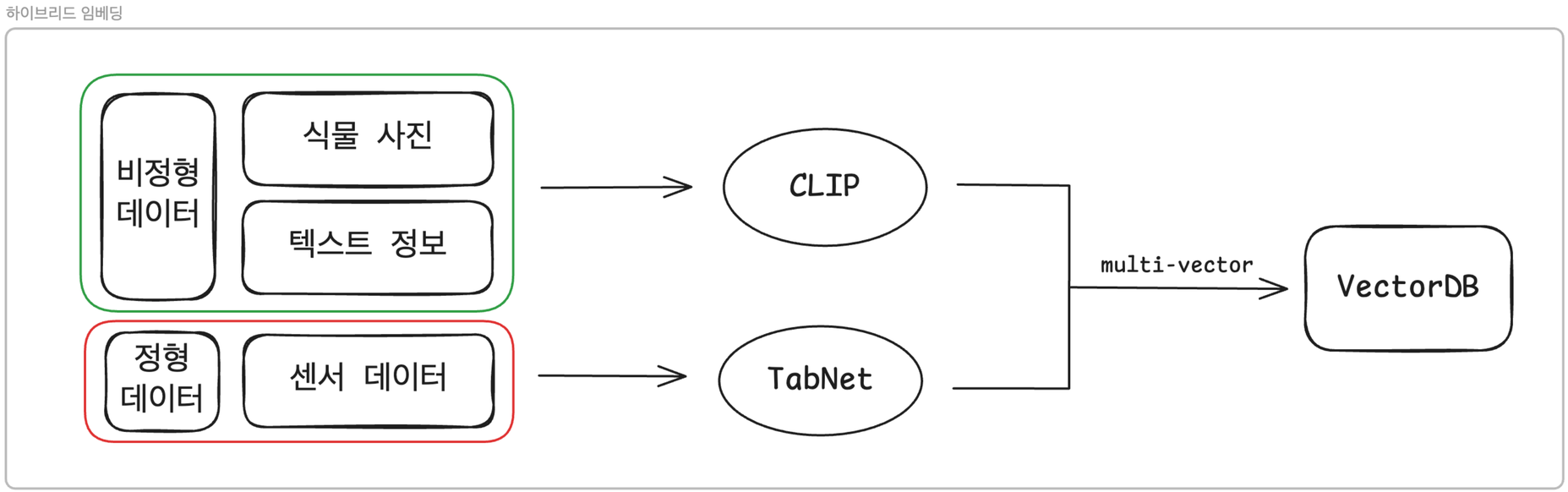
- **Multi-vector**: 문서/질의를 여러 벡터로 분해 (예: ColBERT) → 세밀한 매칭 가능
- **Sparse + Dense Hybrid**: 키워드 기반(BM25) + 임베딩 기반 결합

1️⃣ 임베딩 방법

2. 하이브리드 임베딩

: 여러 임베딩 기술이나 데이터 소스를 결합하여 각 방법의 보완적인 강점을 활용하는 접근법

: [이미지 & 텍스트는 CLIP 모델] + [센서 데이터는 정형 데이터에 강점 있는 TabNet 모델] multi-vector 방법



2️⃣ 멀티모달 임베딩 처리 방법

Beyond Text: Optimizing RAG with Multimodal Inputs for Industrial Applications

: 임베딩 전략이 RAG 전체 성능에 어떤 영향을 주는지

(1) Text-Only RAG

- OpenAI text-embedding-3-small 로 임베딩 후 벡터DB 검색

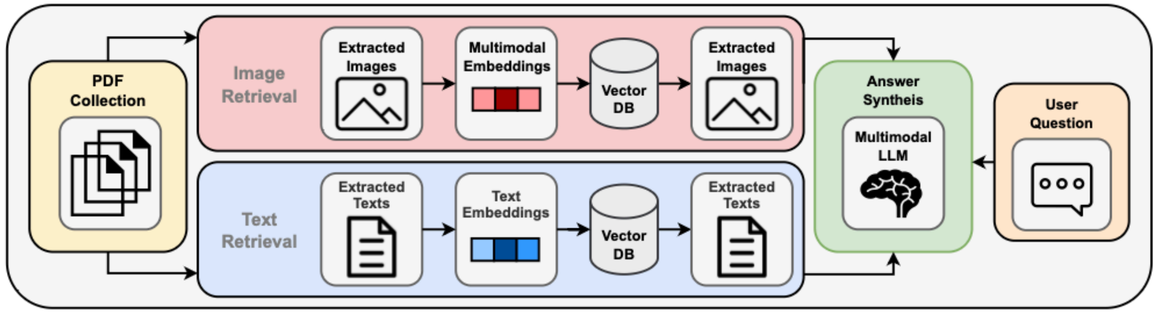
(2) Image-Only RAG

- 방법 A: 멀티모달 임베딩 (CLIP)
  - 이미지와 질의를 같은 임베딩 공간에 투영해 검색
- 방법 B: 이미지 요약 후 텍스트 임베딩
  - GPT-4V/LLaVA로 이미지 요약 → 텍스트로 변환 후 임베딩 → 검색
  - LangChain의 Multi-Vector Retriever 사용해 요약과 원본 이미지 연결

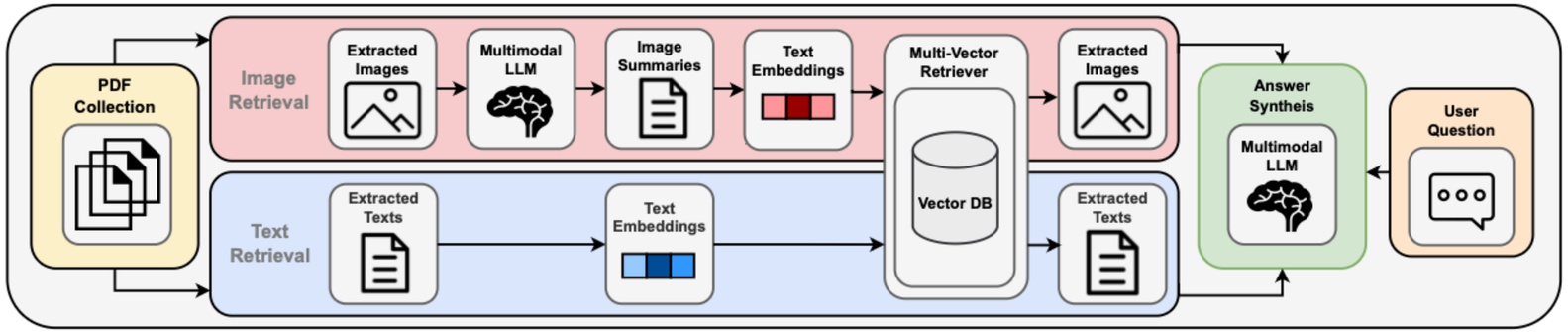
멀티모달 임베딩 처리 방법

(3) Multimodal RAG

- 구성 A: Separate Vector Store
  - 텍스트와 이미지를 따로 임베딩 → 별도 DB 검색 → 결과 합침



- 구성 B: Combined Vector Store
  - 이미지를 텍스트 요약으로 변환 → 텍스트와 같이 임베딩 → 하나의 DB에서 검색



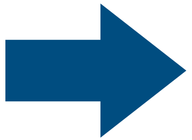
멀티모달 임베딩 처리 방법

결과 정리

1. 텍스트 기반 RAG → 안정적, LLM만 사용한 것보다 확실히 좋음

2. 이미지 단독 RAG
   - A - CLIP 기반: 성능 낮음 (이미지 의미 부족).
   - B - 요약 기반: CLIP보다 훨씬 낫고, 텍스트 검색과 잘 결합됨

3. 멀티모달 RAG
   - A - Separate Store: 이론적으로 깔끔하지만, 이미지 검색 성능이 발목을 잡음
   - B - Combined Store: 이미지 → 텍스트 요약 후 통합 DB에 넣는 방식이 가장 실용적이고 성능 좋음

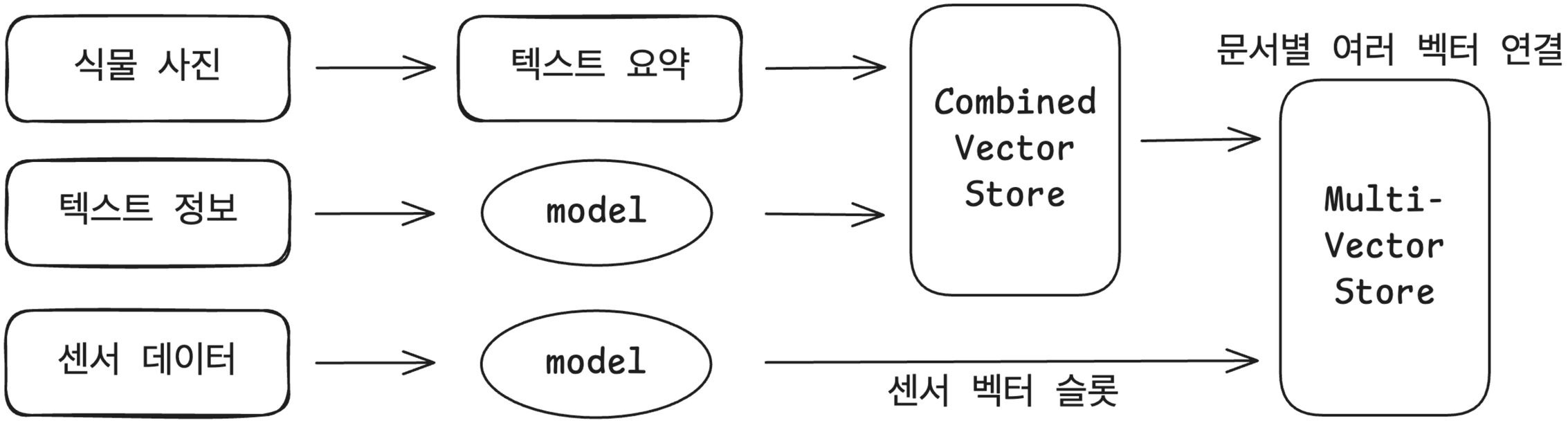


- **CLIP 단일 벡터 대신, 이미지를 텍스트 요약 (caption) + 라벨로 변환 후 텍스트 임베딩**
- **텍스트와 동일한 공간에 저장하는 (Combined Vector Store)**

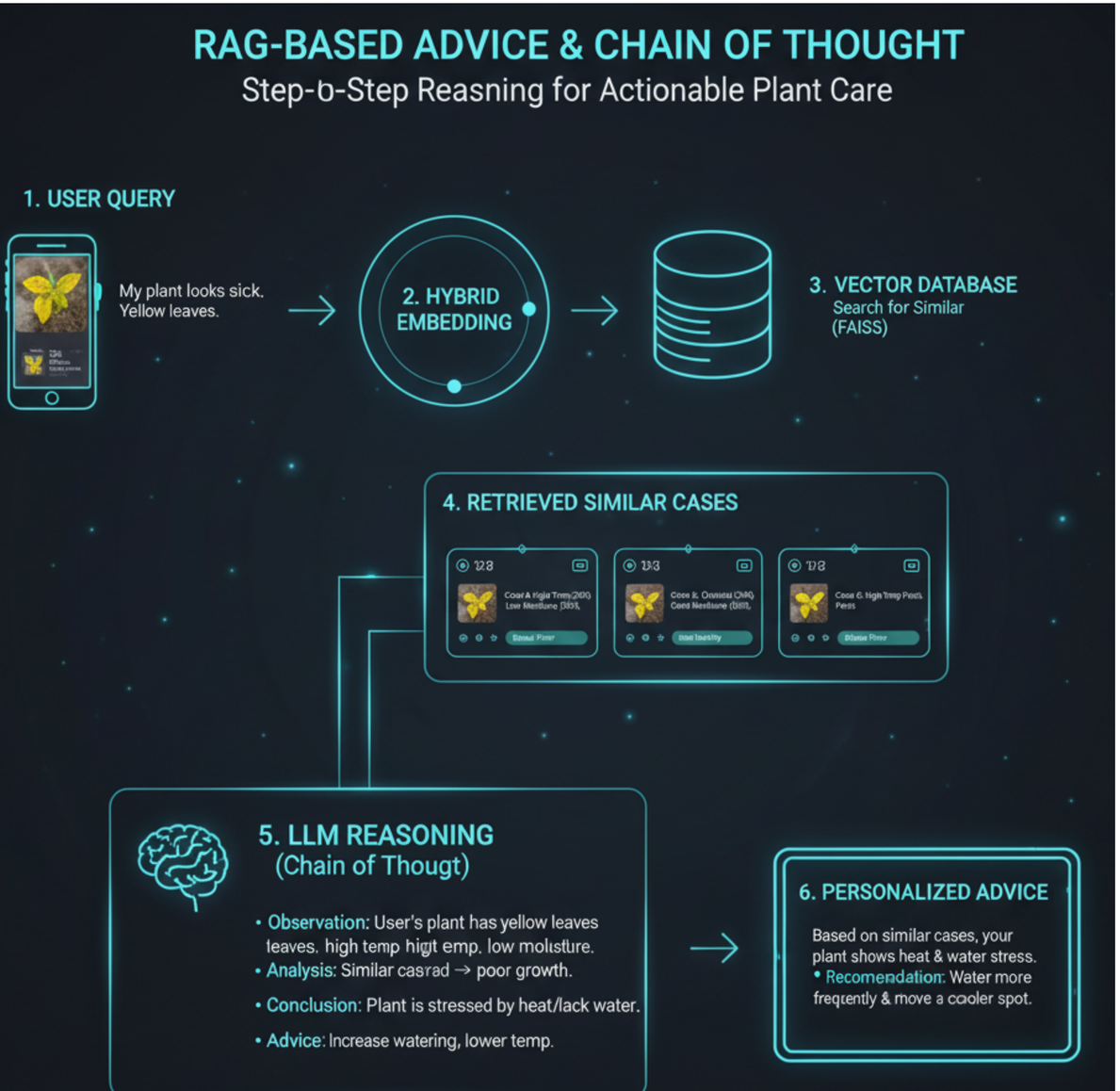
1️⃣, 2️⃣ 선택한 방법

앞선 두 연구의 방법을 참고하여 새로운 방식 설계

=> 텍스트+이미지 요약은 Combined Vector Store, 센서 데이터는 Multi-Vector 구조로 채택



예상 input 과 output



- **Input**
  - 식물 사진
  - 텍스트 설명 “스투키를 키우고 있는데 요즘 성장이 더딘 것 같아요”
  - (선택적으로) 센서 데이터를 입력

- **Output**

[간단한 진단/설명]

- “최근 온도가 높아 성장 속도가 둔화될 수 있습니다.”

[행동 조언]

- “물을 2~3일마다 조금씩 주는 것을 권장합니다.”
- “온도가 30도를 넘지 않도록 환기를 해주세요.”

[유사 사례 기반 추천 (벡터 검색 활용)]

- “비슷한 조건에서 잘 자란 사례에서는 토양 수분이 35% 이상이었습니다. 참고하세요.”

[상태 평가 요약]

- “현재 성장 상태: 보통. 주요 이슈: 수분 부족. 개선 권장: 물 공급 주기 단축.”

감사합니다.