| 발간등록번호 |
| --- |
| 11-1371689-100001-01 |


# 국립중앙도서관 지능형 검색 서비스 적용방안 연구



국립중앙도서관

# 제출문

**국립중앙도서관 귀하**

본 보고서를 국립중앙도서관에서 의뢰한 [국립중앙도서관 지능형 검색 서비스 적용방안 연구] 용역 최종보고서로 제출합니다.

2025. 07


- 연구기관: ㈜렛서
- 연구책임자: 김태희 부대표
- 공동연구자: 김지환 연구원
- 연구참여자 : 양승빈 연구원, 김강열 연구원, 고동휘 연구원

이 연구는 2025년도 국립중앙도서관 연구개발비로 수행되었으며,
본 연구에서 제시된 정책 제안이나 의견 등은 국립중앙도서관의 공식 의견이 아닌
본 연구진의 견해임을 밝혀 둡니다.

# 목 차


## 제 1장. 서 론

1. 연구배경 (7p)
2. 연구대상 및 문제점 (8p)
3. 연구목적과 연구범위 (10p)
4. 연구방법 (12p)

## 제 2장. 활용 및 적용 기술

1. Retrieval Augmented Generation (RAG) (15p)
2. LLM-based Agent (16p)
3. LLM as Judge (17p)

## 제 3장. 연구 수행 내용

1. RAG 기반 검색 도우미 (18p)
   1.1.개요 (18p)
   1.2.데이터 분석 및 가공 (19p)
   1.3.검색 모듈 구축 (21p)
   1.4.생성 모듈 구축 (23p)
   1.5.검색 도우미 구축 (24p)
2. LLM as Judge (27p)
   2.1.평가 시스템 구축 (27p)
   2.2.Human Alignment 측정 (36p)
3. 본문 활용 가능성 탐색 (38p)
   3.1.배경 (38p)
   3.2.분석 결과 (39p)
4. 한계점 (41p)

## 제 4장. 결 론

1. 연구요약 (43p)
2. 연구결과 (44p)
3. 기대 효과 (46p)
4. 추후 연구 및 필요작업 (48p)
5. 제언 (50p)

# 그림차례

그림 1.1 연구 진행 프로세스 (8p)

그림 2.1 Retrieval-Augmented Generation(RAG) 구조개요 (12p)

그림 3.1. 검색 도우미 구조 (15p)

그림 3.2 초록 & 가상리뷰 합성 예시 (16p)

그림 3.3 검색 도우미 예시 (Vector-based Retrieval) (22p)

그림 3.4 검색 도우미 활용 프롬프트 예시 (25p)

그림 3.5 LLM as Judge에 활용된 프롬프트 예시 (27p)

그림 3.6 데이터 활용 옵션에 따른 비교 평가 평균 승률 (30p)

그림 3.7 다른 데이터 활용 옵션끼리 측정한 비교 평가 평균 승률 (30p)

그림 3.8 시나리오 별 비교 평가 평균 승률 (31p)

그림 3.9  평가자 별 평균 효용성 점수 (32p)

그림 3.10  데이터 활용 옵션에 따른 평균 효용성 점수 (32p)

그림 3.11 시나리오 별로 구분한 데이터 활용 옵션에 따른 평균 효용성 점수 (33p)

그림 3.12 Human - Pair-wise Evaluation (비교평가) 예시 (34p)

그림 3.13 Human - Point-wise  Evaluation (정성평가) 예시 (34p)

그림 4.1 지능형검색 반영된 사용자화면  예시 (50p)

# 초 록

1. 연구 목적

본 연구는 국립중앙도서관 도서 데이터를 기반으로 문맥 이해 및 의미 기반 검색을 지원하는 지능형 검색 도우미를 구축하고, 그 성능과 효용성을 RAG(Retrieval-Augmented Generation) Framework와 LLM 기반 평가 기법을 통해 실증적으로 검증하는 것을 목적으로 한다. 기존 키워드 중심 검색은 자연어 질의 처리와 복합적 정보 요구 대응에 한계가 있으므로, 본 연구는 이를 보완하기 위한 대안을 제시하고자 한다.

2. 연구 방법

연구에서는 BM25 알고리즘 기반 키워드 검색과 Dense Embedding 기반 의미 검색을 결합한 하이브리드 검색 모듈과, 검색 결과를 바탕으로 답변을 생성하는 생성 모듈을 설계하였다.
데이터베이스는 국립중앙도서관이 제공한 6,873건의 메타데이터와 711건의 본문 데이터를 전처리하여 구축하였으며, GPT-4.1을 활용해 가상의 초록과 리뷰를 합성하여 검색 품질 향상에 활용하였다.
검색 모듈은 데이터 조합에 따라 네 가지 옵션(①제목+키워드, ② 제목+키워드+목차, ③ 제목+키워드+목차+초록, ④ 제목+키워드+목차+초록+리뷰)으로 구성되었고, 사용자의 질의 유형에 따라 동적으로 검색 전략을 선택할 수 있는 구조를 적용하였다. 성능 검증은 LLM as Judge 평가 시스템을 활용하였으며, 평가 기준은 정확성, 관련성, 완전성, 명확성·간결성으로 설정하였다. 또한 Human 평가와 비교하여 신뢰성을 검증하였다.

3. 연구 결과

LLM 기반 성능 평가는 데이터 조합이 풍부할수록 응답 품질이 유의하게 향상됨을 보여주었다. ELO 점수 기준으로 가장 단순한 조합(제목+키워드)은 747점, 가장 많은 데이터를 활용한 조합(제목+키워드+목차+초록+리뷰)은 1224점을 기록하였다. 특히 후자는 전자 대비 93%의 우세 승률을 보였다.
시나리오별 분석에서도 동일한 경향이 나타났으며, 특히 2개 이상 주제 기반 탐색(시나리오 2)에서 성능 격차가 크게 확인되었다. Human 평가에서는 본문 데이터를 포함한 응답이 가장 높은 점수를 기록하였고, 개인적 상황·정서적 요구(시나리오 3)에서 효과가 두드러졌다.
또한 Human Alignment 검증 결과, 전체 88개 평가 샘플 중 72%가 인간과 동일한 판단을 보였으며, Cohen’s Kappa 계수는 0.419로 중간 수준의 일치도를 나타냈다. 이는 LLM Judge 방식이 일정 수준에서 인간 평가를 대체할 수 있음을 시사한다.

## 4. 결론

본 연구는 국립중앙도서관 데이터를 활용한 지능형 검색 도우미 구축을 통해 다음과 같은 결론을 도출하였다.

1. 복합 데이터 조합(목차, 초록, 리뷰 포함)은 검색 및 응답 품질 향상에 기여를 한다.
2. 하이브리드 검색 전략(BM25+Dense Embedding)은 다양한 질의 유형에서 효과적으로 작동하며, 자연스러운 응답 생성을 지원한다.
3. 본문 데이터 활용은 개인화 추천과 문맥 기반 탐색에서 핵심적 자원으로, 향후 도서관 정보검색 시스템의 경쟁력 제고에 중요한 역할을 할 수 있다.

따라서 본 연구는 기존 메타정보 중심 검색 시스템의 한계를 보완하고, 사용자 중심의 의미 기반 검색 환경으로의 전환 가능성을 실증적으로 입증하였다는 점에서 학술적·실무적 의의를 지닌다.

# 제 1장. 서 론

## 1. 연구배경

국립중앙도서관은 국내에서 발행된 각종 출판물과 지식 정보를 수집·보존하고 이를 체계적으로 정리하여 국민에게 제공하는 중추적인 기관이다. 그러나 최근 인공지능(AI)을 중심으로 한 정보서비스 환경이 급속히 발전함에 따라, 도서관 검색 시스템 또한 사용자 중심의 접근성과 편의성을 강화할 필요성이 대두되고 있다. 특히, 단순한 키워드 검색을 넘어, 사용자의 의도와 문맥을 이해하는 지능형 검색서비스에 대한 요구가 증대되고 있다. 이에 따라, 국립중앙도서관이 보유한 방대한 데이터의 품질과 검색 활용 가능성을 검증하고, 차세대 검색서비스의 방향성을 모색하는 것이 중요한 과제로 떠오르고 있다.

### 1.1 국립중앙도서관의 공공 인프라 역할

국립중앙도서관은 국내에서 발간된 출판물과 다양한 지식 정보를 총괄적으로 수집·보존하고, 이를 체계적으로 정리하여 국민에게 제공하는 중추적인 공공 인프라이다.

- 도서관 주요현황(2024년 기준)에 따르면 국립중앙도서관 대표홈페이지 연 방문자는 약 530만명, 디지털 콘텐츠 연간 페이지뷰는2,200만건의 이용 기록이 존재한다. 이는 도서관이 제공하는 물리적 자료뿐 아니라 디지털 콘텐츠에 대한 꾸준한 수요를 반영한다.

### 1.2 AI 기반 정보서비스 환경의 가속화

인공지능 기술의 비약적 발전과 함께, 전 세계 도서관 및 정보서비스 기관들은 검색, 추천, 요약 등 다양한 분야에 AI를 적극적으로 도입하고 있다.

- 국외 사례: 예를 들어, 미국 캘리포니아 주립대학교 롱비치(CSULB9.1)의 도서관은 스마트 검색, 자동 요약, 대화형 챗봇 기능 등을 통해 정보 탐색의 정확성과 효율성을 높이고 있다 . 이는 도서관

서비스가 이용자의 문맥과 요구를 파악하여 맞춤형 대응을 수행하는 방안을 제시한다. (https://www.csulb.edu/artificial-intelligence-tools-inside-library-resouces)

### 1.3 국립중앙도서관이 보유한 데이터의 품질 및 검색 활용성

지능형 검색서비스 구현을 위해서는 데이터의 품질과 다양성이 핵심적 자원으로 작용한다.

- 소장 규모: 국립중앙도서관은 오프라인 도서 약 14백만건과 온라인 자료 약 33백만건을 보유하고 있다. 이러한 대규모 소장자료는 AI 기반 검색 시스템의 풍부한 훈련 자료로 활용할 수 있다.
- 데이터 종류: 메타데이터 관점에서 국립중앙도서관이 보유한 데이터는 다음과 같은 항목들로 구성되어 있다.

  1. 기본 메타정보: 도서 제목, 저자, 출판사, 발행연도, ISBN
  2. 구조화 정보: 목차, 키워드, 분류체계(주제어)
  3. 텍스트 자료: 본문 전문, 초록(요약)
  4. 이용 및 접근 데이터: 대출 이력, 디지털 열람 통계, 사서 추천 기록 등

  이들 데이터는 문맥 기반 검색, 개인화 추천, 관련도 판단 등 지능형 검색의 다양한 기능 구현을 위한 핵심 자원으로 활용 가능하다.

## 2. 연구대상 및 문제점

### 2.1 연구대상

본 연구는 국립중앙도서관이 보유한 도서 정보 데이터를 기반으로, 지능형 검색서비스 구현 가능성을 검토하는 것을 주요 대상으로 삼는다. 연구에 사용된 데이터는 다음의 조건을 모두 충족하는 도서를 중심으로 선정되었다.

- 기본 정보 포함: 책제목, 키워드, 목차, 본문 등 사용자의 질의에 응답하기 위한 핵심 텍스트 요소를 포함

- 언어 조건 충족: 본문이 한글로 작성되어 있으며, 한국어 기반 대규모언어모델 LLM(Large Language Model)에서 처리 가능
- 저작권 제한 없음: PDF 원문이 존재하고, 저작권이 없는 도서
- 메타데이터 정합성 확보: 동양서 분류기호 중 중분류 값이 명확하게 존재하고, 교과서 등 실험 목적에 적합하지 않은 자료는 제외

최종적으로 연구대상으로 확정된 데이터는 총 6,873건이며, 이 데이터셋을 기반으로 다양한 검색 시나리오 실험과 검색 품질 검증을 수행하였다.

## 2.2 문제점

① 키워드 중심 검색의 한계

대부분의 도서관 검색 시스템은 도서 제목, 저자명, 출판사 등 구조화된 키워드를 기반으로 동작한다. 그러나 실제 이용자는 해당 정보를 정확히 기억하지 못하는 경우가 많으며, 책의 내용 일부나 기억에 남는 키워드만을 가지고 검색하는 사례가 빈번하다.

- *예시: 한 사용자가 어린 시절 읽었던 그림책을 다시 찾고자 하며, 책의 정확한 제목이나 저자는 기억나지 않지만 "달빛 아래에서 고양이가 춤추던 장면"이 인상 깊었다고 회상한다. 이 사용자는 '고양이', '달빛', '그림책' 등의 키워드를 입력하여 검색을 시도하지만, 기존 검색 시스템은 제목이나 저자, 출판사 등 키워드가 명시적으로 포함되지 않은 경우 해당 도서를 검색 결과에 포함시키지 못한다.*

- 문제점: 해당 도서가 실제로 존재하더라도 메타정보가 제한적일 경우, 텍스트 본문이나 이미지 캡션에만 관련 키워드가 있는 경우에는 검색 결과로 노출되지 않음. 특히 아동 도서나 창작 도서처럼 감성 중심 묘사가 많은 경우 키워드만으로는 내용 유추가 어렵다.

② 문맥 기반 검색 기능 부재로 인한 주제 일치 실패

학술 자료나 특정 주제를 기반으로 자료를 탐색하려는 이용자 수요는 증가하고 있으나, 기존 검색 시스템은 키워드 매칭에 의존하고 있어 문맥을 충분히 이해하지 못하는 한계를 보이고 있다. 이는 특히 복합 개념이나 문맥을 고려해야 하는 질의에서 심각한 검색 정확도 저하를 초래한다.

- *예시: 사용자가 "전남지역 청년정책 현황"이라는 주제로 도서를 검색할 경우, "전남", "청년", "정책"이라는 키워드가 모두 포함되지 않으면 시스템은 연관성이 낮은 결과를 반환하거나 아예 검색 결과를 제공하지 못한다.*

- 문제점: 키워드 단편 정보에 의존하는 구조는 문장 구조나 맥락을 이해하지 못해, 유사한 주제임에도 검색 매칭률이 낮아지고, 관련성이 높은 자료를 간과하게 만든다.

③ 개인화 추천 기능의 미비

최근 정보 이용 환경은 이용자의 연령, 정서, 관심사, 이용 맥락 등을 반영한 개인화된 정보 제공 기능의 필요성이 확대되고 있다. 그러나 국립중앙도서관을 포함한 대부분의 기존 검색 시스템은 사용자 프로필이나 감성 태그 등을 기반으로 한 추천 기능이 거의 없거나 제한적이다.

- *예시: 한 중학생 이용자가 "슬프지만 위로가 되는 청소년 소설"을 찾고자 할 경우, 현재 시스템은 '슬픔'이나 '위로' 등의 감성 요소를 키워드에 반영하지 않기 때문에 정확한 추천이 어렵다.*

- 문제점: 검색 시스템이 연령대나 정서적 니즈를 고려하지 않으면, 콘텐츠의 성격과 사용자의 상태 간의 불일치가 발생하며, 이는 추천 품질과 만족도 하락으로 이어진다.

# 3. 연구목적과 연구범위

## 3.1 연구목적

본 연구의 궁극적인 목적은 국립중앙도서관이 보유한 도서 데이터를 기반으로, 기존의 키워드 중심 검색 시스템의 한계를 극복할 수 있는 지능형 검색서비스 구현 가능성을 실증적으로 검토하는 데 있다. 특히, 사용자의 문맥적 질의에 대응 가능한 검색 시스템을 구축하기 위해 다음과 같은 세부 목적을 설정하였다.

- 문맥 이해 기반 검색을 위한 최적의 데이터 조합 구성
  기존 메타정보(도서 제목, 키워드 등)만으로는 복잡한 사용자 질의에 효과적으로 대응하기 어려운

한계를 해결하고자, 다양한 정보 항목(목차, 초록, 가상리뷰, 본문 등)의 조합을 설계하고 검색 효율성을 비교한다.

- 개인화 및 정서 기반 검색 대응을 위한 신규 데이터 활용 가능성 탐색
  사용자의 연령, 감성, 상황 등을 반영할 수 있는 데이터 항목의 필요성을 규명하고, 기존 데이터 기반의 가상 초록 및 리뷰 등 생성 가능한 컬럼을 정의함으로써 사용자 맞춤형 검색 수요에 대응 가능한 구조를 마련한다.

- 자연어 기반 검색 시스템의 실현 가능성 검토
  대규모 언어모델(LLM)과 벡터 데이터베이스를 결합하여 자연어 질의에 대한 정확하고 관련성 높은 응답이 가능한 검색 도우미 구조를 설계하고 실험을 통해 그 구현 가능성을 평가한다.

## 3.2 연구범위

본 연구의 범위는 국립중앙도서관이 보유한 도서 관련 데이터에 기반하여, 지능형 검색서비스 구축을 위한 전 과정(데이터 분석 → 시스템 설계 → 실험 및 평가)을 포함한다. 구체적인 연구범위는 다음과 같다.

- 국립중앙도서관 서지·목차· 본문 데이터의 품질 및 검색 활용성 평가
  - 도서 제목, 키워드, 목차, 본문 등 주요 항목을 중심으로 실질적 검색 응답에 활용 가능한 데이터 항목을 식별하고, 해당 항목들의 결합 조합을 통해 검색 성능의 변화를 정량적으로 분석한다.
  - 이러한 데이터를 모두 보유한 도서 샘플셋을 구성하여 연구 대상 데이터로 확정한다.

- 기존 메타 정보 외에 신규 데이터 항목 추가 활용 가능성 검토
  - 연령대, 감성, 상황 기반의 추천 수요를 수용하기 위해 기존 데이터에서 생성 가능한 초록 및 리뷰 데이터를 합성하고, 이를 검색에 활용하였을 때의 성능 변화를 평가한다.
  - 본 합성 데이터는 문맥 이해 및 추천 근거 설명력 향상 측면에서의 기여도를 검토한다.

- 벡터 데이터베이스 및 LLM 기반 검색 시스템 구축 실험 및 평가

- 데이터 전처리, 임베딩, 벡터화 등 일련의 벡터DB 구축 과정을 통해 문맥 기반 검색이 가능한 구조를 구현한다.

- 검색 시나리오(내용 기반 탐색, 주제 기반 추천, 세그먼트 기반 추천)를 설정하고, 각 질의 유형에 대해 질의 응답 시스템의 정합성과 활용 가능성을 평가한다.

- LLM을 활용하여 사용자 질의에 대한 자연어 응답을 생성하고, 정확성·관련성·완전성 기준에 따라 응답 품질을 측정한다.

## 4. 연구방법

본 연구는 총 8주에 걸쳐 수행되었으며, 데이터 분석 및 선정부터 시스템 설계, 구현, 검증, 최종 보고서 작성에 이르기까지 체계적인 절차를 기반으로 진행되었다. 각 주차별 수행내용은 다음과 같다.

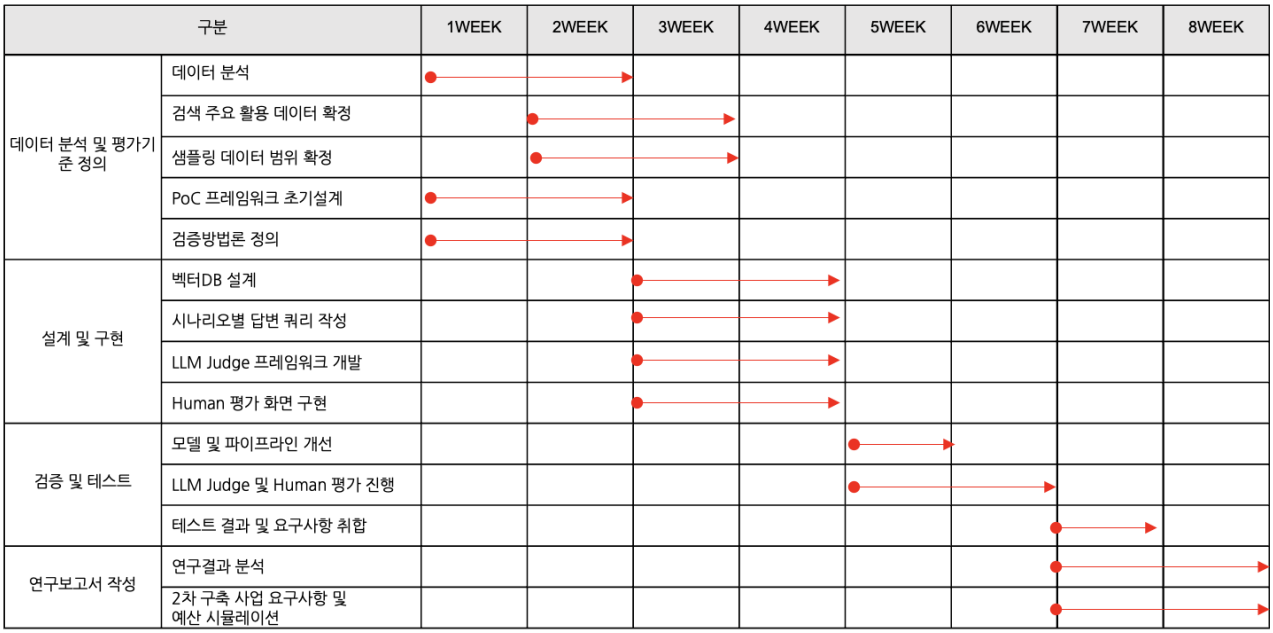



그림 1.1 연구 진행 프로세스

## 4.1 [1~3 주차] : 데이터 분석 및 연구대상 데이터범위 확정

연구 초기 단계에서는 국립중앙도서관이 보유한 방대한 도서 데이터의 구조와 항목을 분석하고, 지능형 검색서비스 구현에 필수적인 데이터를 식별하는 작업이 수행되었다.

- 데이터 수급 및 컬럼 검토: 국립중앙도서관의 통합 데이터베이스를 대상으로, 도서 제목, 키워드, 목차, 본문 등 주요 텍스트 정보의 확보 여부를 중심으로 컬럼 구조를 검토하였다.
- 연구대상 데이터 확정: 검색 시스템 개발에 필요한 모든 핵심 정보를 포함한 도서를 기준으로 연구 대상 데이터셋을 구성하였으며, 이 과정을 통해 전체 도서 중 실험 가능성이 높은 데이터로 추출되었다.
- 테스트 질의 데이터셋 정의: 연구 평가를 위한 시나리오 기반 질의셋 정의를 위해, 실제 사용자 검색 행태를 반영할 수 있도록 '통합검색 질의'과 '사서에게 물어보세요' 데이터를 분석하였다. 이로써 사용자 질의 유형(내용 기반 검색, 다양한 주제 기반 추천, 연령,상황별 맞춤 기반 추천)에 따른 질의 템플릿을 구성하였다.

## 4.2 [3~4 주차] : 설계 및 구현

중간 단계에서는 검색 도우미 Framework를 설계하고, 데이터 조합에 기반한 벡터 데이터베이스 구축, 질의-응답 흐름 구현, 평가 시스템 구성 등의 실질적인 개발이 진행되었다.

- 검색 도우미 Framework 설계: 사용자가 자연어 형태로 질의를 입력하면, 대규모 언어모델(LLM)이 이를 이해하고 키워드 기반 질의와 벡터 기반 질의를 동시에 수행한다. 검색 방식은 국립중앙도서관의 기존 검색용 데이터베이스와 본 연구에서 구축하는 벡터 데이터베이스를 각각 활용하여 병렬적으로 진행되며, 최종 검색 결과를 종합하여 사용자 맞춤형 응답을 생성한다.
- 데이터 조합별 벡터DB 구축: 데이터 항목(책제목, 키워드, 목차, 본문, 초록, 가상리뷰)을 다양하게 조합한 후, 각 조합에 대해 텍스트 임베딩 과정을 거쳐 벡터DB를 구축하였다. 초록과 가상리뷰는 내부 데이터를 기반으로 생성하여 포함시켰다.
- 질의 시나리오 작성: 앞서 정의된 질의셋을 바탕으로, LLM이 처리할 수 있는 형태의 질의 질의를 작성하였다.

- LLM 기반 평가 Framework 및 UI 구현: 평가 시스템에서는 GPT 기반 LLM을 이용한 자동 평가 구조를 구현하고, Human 평가자가 평가에 참여할 수 있도록 전용 사용자화면을 설계하였다. 동시에 지능형 검색서비스의 결과를 확인할 수 있는 사용자 예시화면도 함께 구현되었다.

## 4.3 [5~7 주차] : 검증 및 테스트

이 단계에서는 구축된 시스템에 대한 품질 검증과 성능 개선 작업이 집중적으로 수행되었다.

- 내부 검증 및 품질 개선: 개발된 시스템을 대상으로 품질검수를 위해 내부 QA(Quality Assurance)를 통해 지능형 검색서비스의 PoC와 평가 화면의 오류를 점검하고, 모델 파이프라인의 성능 개선을 진행하였다.

- LLM Judge 평가: 자동 평가 Framework 를 통해 질의 시나리오에 대한 다양한 응답의 정확성, 관련성, 완전성 등을 Pairwise 방식으로 평가하였다.

- 국립중앙도서관 협력 기반의 Human 평가: 실제 현장 전문가들과 협력하여 Human 평가를 수행하였다. 평가 결과는 자동 평가 결과와의 일치도를 분석하고, 지표 정제 및 향후 개선 방향 도출에 활용되었다.

## 4.4 [7~8 주차] : 연구보고서 작성

최종 단계에서는 연구 전 과정에 대한 종합 분석과 함께 보고서 작성을 수행하였다.

- 결과 데이터 정제 및 분석: 각 테스트 시나리오의 성능 결과를 정량적, 정성적으로 분석하여 주요 인사이트를 도출하였다.

- 요구사항 취합 및 개선안 도출: 평가 과정에서 제시된 사용자 피드백 및 기술적 한계를 바탕으로, 향후 시스템 고도화를 위한 개선사항을 정리하였다.

- 최종 보고서 작성: 연구의 목적, 방법, 실험 내용 및 결과를 체계적으로 정리한 종합 보고서를 작성함으로써 연구를 마무리하였다.

# 제 2장. 활용 및 적용 기술

## 1. Retrieval Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) 는 대규모 언어 모델의 출력을 응답을 생성하기 전에 학습 데이터 소스 외부의 신뢰할 수 있는 지식 베이스 (예시. 회사 문서 의료 기록)를 참조하도록 하는 프로세스입니다.

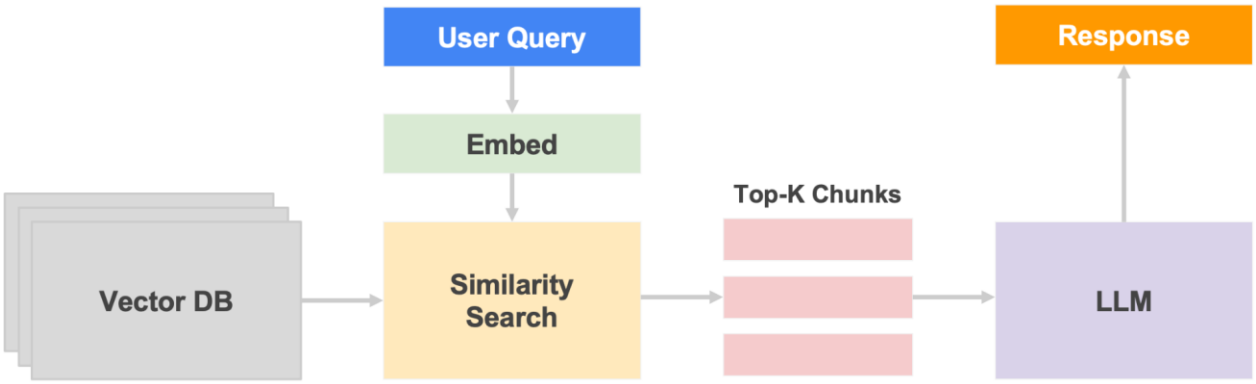



그림 2.1. Retrieval-Augmented Generation(RAG) 구조개요

- RAG는 검색 (Retrieval)과 생성 (Generation)을 결합한 방식으로, 사전에 구축된 데이터베이스로부터 관련 정보를 검색한 후 이를 바탕으로 자연어 답변을 생성하는 기술이다 [1, 2, 3].
- 생성 모델은 문맥을 잘 이해하여 자연스러운 답변을 생성하지만, 사실 관계가 부정확하거나 사전에 학습하지 못한 정보를 답변 생성에 반영하지 못하는 단점이 있다. RAG는 이러한 단점을 보완하기 위해 검색을 기반으로 필요한 정보를 보충하여 답변 생성에 활용하는 기법으로, 크게는 검색 단계와 생성 단계로 구성되고, 구체적인 절차는 다음과 같다.

### 1.1. 검색 단계

- 사용자 질의가 주어지면, 데이터베이스에서 질의와 관련된 문서들을 검색한다.
- 문서 검색 방식으로는 ① 키워드 기반 검색 ② 의미 기반 검색 ③ 이를 혼합한 하이브리드 검색이 존재한다.

- 키워드 기반 검색은 주로 BM25 기반의 방법론들이 활용되며, 질의 내의 키워드와 문서 간의 용어 빈도, 역문서 빈도와 같은 통계적 정보를 이용하여 검색 결과를 도출한다. 이는 키워드가 명확한 질의나 문서의 핵심 주제가 분명한 경우 효과적으로 활용될 수 있다 [4].
- 의미 기반의 검색은 임베딩 모델을 활용한 Dense Vector 기반의 방법론이 사용되며, 질의와 문서를 각각 연속적인 벡터 공간에 투영하여 이들 간의 의미적 유사도를 계산하여 관련 문서를 검색한다. 이는 키워드가 명시적이지 않거나 문장 간의 의미 유사성이 중요한 복잡한 질의에서 강점을 보인다 [5, 6].
- 하이브리드 검색은 키워드 기반 검색과 의미 기반 검색을 결합한 방식으로, 각 방식의 장점을 모두 활용하여 보다 정확하고 포괄적인 검색 결과를 얻는 데 목적이 있다 [7].

### 1.2. 생성 단계

- 검색된 문서들을 맥락 정보로 활용하여 LLM이 답변을 생성한다.
- 이 때 LLM은 검색된 문서 내 정보를 참조하여 정확하고 관련성 높은 답변을 생성한다.
- 본 연구에서는 국립중앙도서관의 도서 데이터를 기반으로, 키워드 기반과 의미 기반의 검색을 결합한 하이브리드 검색을 수행한 후, 검색 결과를 바탕으로 RAG 기술을 적용하여 사용자의 질의에 대해 적절한 도서 정보와 함께 정확하고 관련성 높은 답변을 생성하도록 한다.

## 2. LLM-based Agent

- LLM 기반의 에이전트는 자연어 처리 기술을 활용하여 사용자의 요청이나 질의에 자율적으로 판단하고, 적절한 작업 수행이나 답변 생성을 할 수 있는 인공지능 시스템이다. 특정 과업을 부여받으면, 자율적인 계획 수립과 환경과의 상호작용을 통해 사전에 제공된 도구를 효과적으로 활용하여 문제를 해결하며, 필요 시 여러 차례 LLM을 호출하기도 한다 [8, 9, 10].
- 다양한 상황과 질의에 맞춰 유연한 대응이 가능하며, 목표에 최적화된 도구를 이용해 작업을 수행하므로 보다 정확하고 신속한 결과를 제공할 수 있다.

- 본 연구에서는 LLM-based Agent를 활용하여 도서 검색 도우미를 구현한다. 사용자의 다양한 질의를 이해하고, 적절한 도서 검색 도구를 자율적으로 호출하여 사용자에게 가장 적합한 도서 정보를 제공하도록 한다.

## 3. LLM as Judge

- LLM as Judge는 LLM의 자연어 이해 능력을 활용하여 시스템이나 모델이 생성한 답변의 품질을 평가하는 방법이다 [11, 12].
- 일반적으로 평가 작업은 사람이 직접 수행하는 것이 가장 신뢰할 수 있지만, 인적 평가에는 시간과 비용이 많이 소요되며, 평가자의 주관성에 따라 평가 결과가 일관되지 않을 수 있다. LLM as Judge는 사람 평가의 대안으로서 떠오르는 평가 방법이다.
- LLM as Judge는 생성된 답변 간의 상대적 우수성을 비교하거나, 주어진 답변이 특정 평가 기준에 얼마나 부합하는지를 판단하는데 사용된다. 일반적인 사용 방식은 다음과 같다.
    - 평가 기준 정의: 평가 시 LLM에 제공할 명확한 평가 기준과 가이드라인을 미리 정의한다. 보편적으로 사용되는 기준으로는 정확성, 관련성, 완전성, 명확성 등이 있다 [13].
    - 평가 대상 입력: 평가하고자 하는 답변 또는 문장 쌍을 LLM 평가자에게 입력으로 제공하며, 평가 대상은 단일 문장의 품질 평가 또는 두 개 이상의 답변을 상대적으로 비교 평가하는 방식으로 이루어진다.
    - 평가 결과 산출: 입력된 평가 기준과 대상을 바탕으로 LLM은 정량적 점수, 상대적 순위, 또는 자유로운 텍스트 형태의 평가 근거를 생성하여 출력한다.
- 최근 연구에서는 GPT-4와 같은 최신의 LLM이 사람 평가자와 유사한 수준의 평가 신뢰성과 일관성을 보여주는 사례들이 보고되고 있으며, 특히 평가 기준과 평가 대상에 대한 구체적인 가이드라인 및 예시를 함께 제공하면 평가의 신뢰성을 더욱 높일 수 있다고 알려져 있다 [14].
- 본 연구에서는 LLM as Judge를 활용하여, 서로 다른 데이터 활용 방식으로 생성된 답변 간 품질을 객관적으로 비교 평가하며, 이를 통해 데이터의 효용성을 정량적으로 검증한다.

# 제 3장. 연구 수행 내용

## 1. RAG 기반 검색 도우미

### 1.1. 개요

- 키워드 기반 검색의 한계를 해소하기 위해 주어진 사용자 질의의 문맥을 이해하고 답하는 검색 도우미를 구축한다.

- Retrieval Augmented Generation (RAG) 기반의 Framework를 구축하여,
  국립중앙도서관에서 제공한 데이터를 기반으로 사전에 구축한 데이터베이스에서 사용자 질의와 연관된 도서 정보를 검색하고 해당 정보를 바탕으로 답변을 생성하도록 한다.

  - 검색 도우미는 사용자 질의와 관련된 정보를 데이터베이스로부터 검색하는 검색 모듈과 검색된 정보를 토대로 답변하는 생성 모듈로 구성된다.

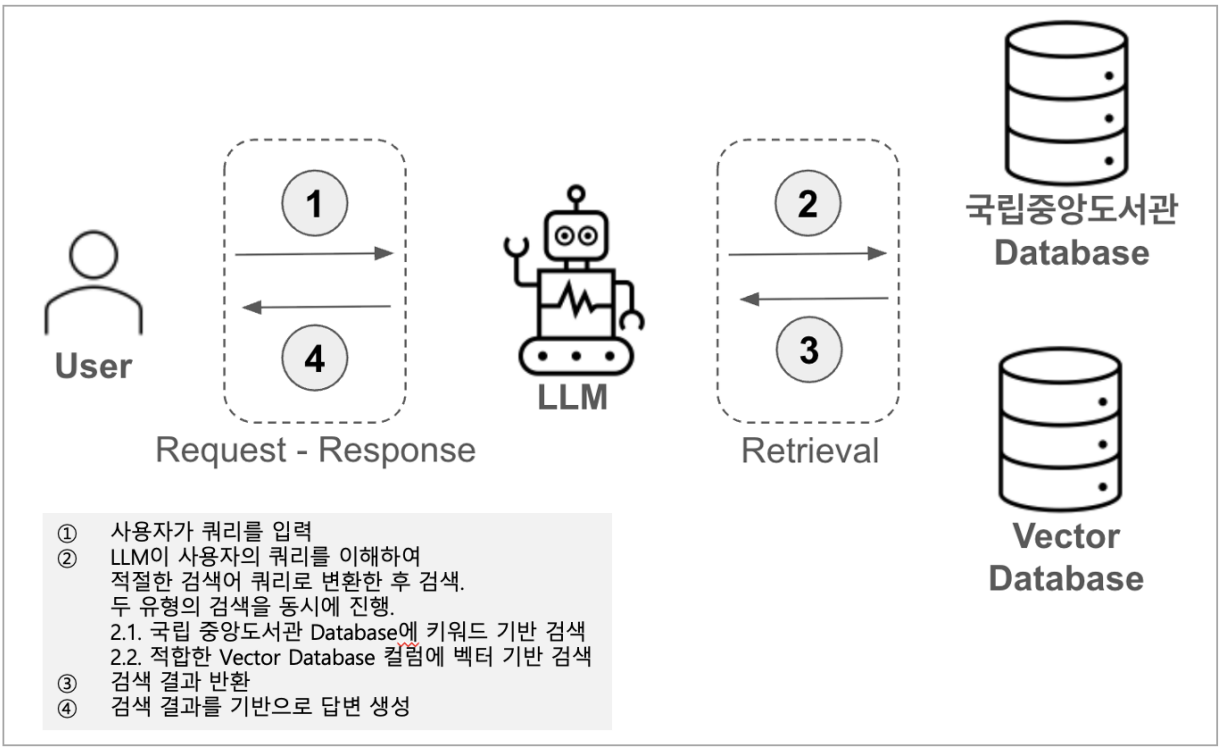


그림 3.1. 검색 도우미 구조

## 1.2 데이터 분석 및 가공

### 1.2.1. 데이터 분석

- 최종 도서 데이터는 6,873건으로, 아래 조건을 만족하는 도서만을 활용한다.
  - PDF 본문 파일이 있으면서 저작권 없는 도서
  - 도서 제목과 키워드, 목차 데이터를 모두 보유한 도서
  - 본문 언어가 한글인 도서
  - 자료 유형 중 교과서와, 동양서 분류기호 중 분류값이 없는 도서는 제외
- 본문 데이터는 6,873건 중 711건의 도서를 활용한다. 도서 선정근거는 다음과 같다.
  - 제공된 6,873건의 본문 데이터는 800GB를 상회하여 많은 시간과 자원이 필요
  - 파일 크기가 가장 작은 1,000건 중 텍스트 추출이 가능한 문서만 활용

### 1.2.2. 검색 활용 정보 식별 및 생성

- 주어진 도서 데이터에는 검색에 활용되기 어려운 정보들이 다수 존재하며, 해당 데이터를 검색에 모두 활용하게 되면 검색 정확도가 매우 저하된다.
- 따라서, 주어진 도서 데이터에서 실제로 검색에 활용할 정보를 식별하여, 실제 검색에 활용될 데이터만 추출한다.
- 본 연구에서는 제목, 키워드, 목차 와 더불어, LLM을 기반으로 합성한 가상의 초록 및 가상의 리뷰를 검색에 활용한다.
- 가상의 초록과 가상의 리뷰는 제목, 키워드, 그리고 목차를 활용하여 합성한 것으로 각각 다음과 같은 정의 및 절차에 따라 생성한다.

#### 1.2.2.1. 초록 합성

- 초록을 “주요 주제와 하위 주제 기반으로 도서 내용 전개의 흐름을 알 수 있는 설명문”으로 정의하여 합성을 수행한다.

- 유사한 형식으로 생성하도록 유도하기 위해 In-context Learning 을 활용하여 2개의 입출력 예시를 제공하며, 각 예시는 도서에 대한 제목, 키워드, 목차로 구성된 입력 예시와 그에 대응되는 실제 초록을 위의 정의에 맞게 가공한 출력 예시로 구성된다.

- 초록 합성에 대한 가이드라인, 2개의 입출력 예시, 초록 합성의 대상이 되는 도서의 제목, 키워드, 목차 데이터를 OpenAI의 GPT-4.1에 입력하여 초록을 합성 한다 [15].

#### 1.2.2.2. 리뷰 합성

- 리뷰를 “감성 리뷰, 비판적/분석적 리뷰, 독자 맞춤 리뷰, 상황기반 리뷰, 실용/정보 중심 리뷰”로 정의하여 합성을 수행한다.

- 초록 합성과 동일하게 In-context Learning을 활용하여 2개의 입출력 예시를 제공한다. 각 예시는 실제 도서에 제목, 키워드, 목차로 구성된 입력 예시와 위의 정의에 맞게 사람이 직접 생성한 리뷰로 구성된 출력 예시로 이루어진다.

- 리뷰 합성에 대한 가이드라인, 2개의 입출력 예시, 리뷰 합성의 대상이 되는 도서의 제목, 키워드, 목차 데이터를 GPT-4.1 모델에 입력하여 리뷰를 합성한다.

| ITITLE | KEYWORD | TOC_KEYWORD | ABSTRACT | REVIEW |
|---|---|---|---|---|
| 결혼심리 | 결혼심리 인생 연애 결혼 | 목차<br>第一部<br>結婚의 目的 = 1<br>一. 結婚의 重大性 = 1<br>二. 結婚의 生理的 目的 = 2<br>三. 結婚의 社會的 目的 = 4<br>四. 結婚의 心理的 目的 = 6<br>第二部<br>戀愛와 結婚 = 9<br>一. 戀愛란 무엇이냐 = 9<br>(1)<br>스탕달의 愛經 = 9<br>(2)<br>쇼오펜하우엘의 戀愛 = 10<br>(3)<br>톨스토이의 肉慾憎惡 = 11<br>(4) | 『결혼심리』는 인생의 중요한 전환점인 결혼을 심리, 사회, 생리 등 다각적인 시각에서 분석하고 그 깊은 의미를 탐구하는 책이다. 제1부에서는 결혼의 생리적·사회적·심리적 목적과 중대성을 살펴 결혼이 인간 삶에 미치는 본질적 영향을 밝힌다. 제2부에서는 고전 사상가 및 각종 이론을 토대로 연애의 본질을 조명하고, 다양한 연애 유형 및 결혼과의 관계를 심도 있게 논의한다. 제3부에서는 결혼의 해부라는 제목 아래 결혼의 적령기, 성격과 성별 문제, 사회적 지위, 그리고 결혼 이전에 경험할 수 있는 비극적 현실까지 다루며 현실적인 조언을 제시한다. 제4부에서는 신혼의 달콤함부터 배우자의 부정, 그리고 행복한 결혼생활을 위한 열쇠까지 결혼 이후 마주치는 다양한 심리적 변화와 문제를 이야기한다. 부록에는 사랑에 관한 시, 서한, 금언을 담아 인간의 사랑과 결혼에 관한 문학적 감수성까지 아우른다. 이 책은 결혼을 앞두었거나 결혼생활 중인 이들이 자신의 삶과 관계를 더 깊이 이해하고자 할 때 길잡이가 되어 준다. | 1. 감성 리뷰<br>이 책을 읽으며 결혼이라는 인생의 큰 파도 앞에 설렘과 불안을 동시에 느꼈습니다. 저자의 깊은 통찰은 마음 한 켠에 평온한 등불을 켜 주었고, 사랑과 결혼에 대한 나의 생각을 부드럽게 어루만져 주었어요. 특히 사랑의 다양한 형태와 변화하는 심리를 살펴볼 때 내 안의 미묘한 감정들이 생생하게 떠올라 가슴이 따뜻해졌습니다.<br><br>2. 비평적/분석적 리뷰<br>『결혼심리』는 결혼의 목적부터 연애와 심리적 측면, 그리고 결혼 후 현실까지 체계적으로 분석한 점이 돋보입니다. 다양한 철학자의 사랑관을 인용해 이론적 깊이를 더했으며, 연애의 다양한 왜곡 형태도 세밀히 다루어 균형 잡힌 시각을 제시했습니다. 다만 일부 문헌 중심의 서술이 현대적 사례나 심리학 최신 연구와 연결이 약해 보이는 점은 아쉽습니다. 전반적으로 심리학과 철학적 접근의 조화가 잘 이뤄진 읽을거리입니다.<br><br>3. 독자별 맞춤 리뷰<br>- 청소년에게: 고등학생 친구들에게 추천해요. 결혼과 사랑이 단순한 감정이 아니라 여러 목적과 심리가 얽혀있는 복잡한 현상임을 이 책이 쉽게 알려줍니다. 연애에 대해 깊게 생각해 볼 수 있는 좋은 기회가 될 거예요.<br>- 중장년층에게: 은퇴 후 인생의 새로운 국면을 맞는 분들에게 특히 권합니다. 결혼의 의미를 다시 한 번 돌아보고, 오랜 연애와 삶을 성찰하면서 마음의 평안을 찾는 데 도움이 됩니다.<br>- 감성 독자에게: 마음이 지치거나 사랑에 대해 혼란스러울 때, 이 책 속 따뜻한 시와 편지들이 잔잔한 위로를 건네줍니다. 고통과 행복이 공존하는 사랑의 진솔한 모습을 통해 감정을 치유할 수 있습니다.<br><br>4. 상황별 추천 리뷰<br>일상에서 사랑과 결혼에 대한 고민이 깊어지는 순간, 이 책은 훌륭한 동반자가 되어 줍니다. 예를 들어 결혼을 앞둔 준비기, 부부관계에 갈등이 있을 때, 또는 중요한 인생 전환기를 맞아 내면을 점검하고 싶을 때 이 책을 펼쳐보세요. 짧은 휴식 시간이나 주말 느긋한 오후에도 적절한 독서가 될 것입니다.<br><br>5. 실용적/정보적 리뷰<br>총 180여 쪽 분량으로 구성된 『결혼심리』는 결혼의 이론적 목적부터 다양한 사랑의 양상, 결혼 전후의 현실 문제까지 폭넓게 다룹니다. 내용은 고전 철학과 심리학 이론이 혼합되어 전문 용어가 일부 있으나, 전반적으로 중급 이상의 독서 난이도를 가졌습니다. 본문 외에도 사랑의 시와 편지, 명언들이 부록으로 실려 감성적인 깊이를 더합니다. 평소 사랑과 결혼에 대한 철학적·심리적 탐구에 관심 있는 독자라면 하루 20~30분씩 나누어 읽기에 적합합니다. |

그림 3.2 초록 & 가상리뷰 합성 예시

- 제목, 키워드, 목차, 가상의 초록, 가상의 리뷰는 각각 데이터베이스화 되어 검색에 활용되며, 구체적인 방법은 검색 모듈 구축 섹션에서 서술한다.

#### 1.2.3. 답변 활용 정보 식별

- 주어진 도서 데이터에는 사용자에 대한 답변 생성 측면에서 활용되기 어려운 정보가 다수 존재하며, 해당 데이터를 답변 생성 모듈에 입력할 시 불필요한 정보를 출력하거나 원치 않는 답변을 생성할 수 있다. 따라서, 주어진 도서 데이터에서 실제로 답변에 활용할 정보를 식별하여 답변 생성에 해당 정보만 참고할 수 있도록 한다.
- 사용자가 도서 검색 관련 질의를 입력했을 때 생성할 정보는 도서 이름과 위치 정보, 추천 이유이다. 이와 관련된 데이터는 도서 검색에 활용한 정보들과 도서의 위치 정보로 국한되므로, 주어진 해당 값만 추출하여 답변 생성에 활용한다.

## 1.3 검색 모듈 구축

- 사용자의 질의가 주어지면 연결된 데이터베이스에 존재하는 데이터와 유사도 비교를 수행하여 사전에 정의된 개수 만큼 가장 유사한 데이터를 검색한다.
- 질의와 도서 데이터 간의 유사도 계산은 BM25를 기반으로 한 키워드 기반 유사도 측정과 Dense Embedding 모델을 기반으로 한 Dense Vector 기반 유사도 측정으로 이루어진다.
- 두 방식으로의 유사도 측정을 위해, 주어진 도서 데이터를 적합한 형태의 데이터베이스로 구축한다.

#### 1.3.1. 키워드 데이터베이스 구축

- 검색에 활용할 도서 데이터를 텍스트화 한 후, 각 텍스트에 대해 토큰화를 수행한다. 토큰은 주어진 텍스트를 구분하는 단위로, BM25에서는 질의와 주어진 문서의 동일한 토큰이 등장하는 빈도를 기반으로 유사도를 측정한다. 일반적으로 영어, 일본어, 중국어 등의 문장은 띄어쓰기를 기반으로 토큰화를 하지만, 한국어는 띄어쓰기 만으로 단어를 정확히 나누기 어렵다. 형태소를 기준으로 나누어야 하며 이를 위해, mecab-ko라이브러리를 활용하여 주어진 텍스트의 명사를 추출하는 방식을 통해 토큰화 한다. [16].

- BM25 알고리즘을 기반으로 문서와 토큰 간의 가중치를 계산한다. BM25는 단어의 빈도와 문서 빈도의 역수를 기반으로 점수를 부여하는 알고리즘으로, 각 데이터를 이루는 모든 텍스트에 BM25 알고리즘을 적용하여 각 데이터마다 토큰의 BM25 점수를 계산하고 BM25 인덱스를 생성한다.
- 사용자 질의가 들어오면 해당 질의에 대해 동일한 토큰화를 적용한 후, BM25 인덱스를 통해 각 도서와 키워드 기반 유사도를 측정하여 가장 유사하다고 판단되는 도서를 사전에 정의한 개수만큼 검색한다.

### 1.3.2. Dense Vector 데이터베이스 구축

- 키워드 데이터베이스 구축과 동일하게, 검색에 활용되는 데이터에 대해 추출하여 텍스트화 한다.
- 각 텍스트에 대해 임베딩을 생성한다. 임베딩은 텍스트를 고정된 차원의 Dense Vector로 변환하여 의미적 유사성을 벡터 공간에서 표현하는 방법이다. 본 연구에서는 OpenAI의 'text-embedding-3-large' 모델을 사용하여 각 텍스트를 Dense Vector로 변환한다. 생성된 벡터는 문서의 의미를 효과적으로 반영하여 의미 기반의 검색에 활용될 수 있도록 한다 [17].
- 이 때, 비용 및 자원의 효율화를 위해 벡터화 하는 최대 텍스트의 길이를 4096 토큰으로 제한한다.
- 각 도서에 대해 생성된 Dense Vector를 벡터 인덱스로 저장한다. 벡터 검색 엔진 구축을 위해 오픈소스 라이브러리인 FAISS를 활용한다 [18].
- 사용자 질의가 입력되면 질의에 대해 동일한 임베딩을 적용하여 Dense Vector로 변환한 후, 구축된 벡터 인덱스를 통해 각 도서 벡터와 유사도를 측정한다. 코사인 유사도를 활용하여 계산하며, 가장 유사도가 높은 도서를 사전에 정의한 개수만큼 검색 결과로 반환한다.

### 1.3.3. Hybrid Search

- 키워드 기반 검색은 단어의 빈도수와 같은 통계적 정보를 기반으로 검색하기에 정확한 텍스트 매칭에는 강력하지만, 문맥적 의미를 반영하기 어렵다. 이와 반대로, Dense Vector 기반의 검색은 의미적 유사성을 잘 포착하지만, 특정 용어나 정확한 단어가 중요할 경우 원치 않은 결과가 나오지 않을 수 있다.

- 본 연구에서는 키워드 기반과 Dense Vector 기반의 검색 기법을 결합하여 각 방식의 장점을 극대화하고 한계를 보완하는 하이브리드 검색 기술을 활용하며, 두 검색 결과를 Convex Combination 방식으로 결합하여 최종 유사도 점수를 산출한다.

- Convex Combination은 키워드 기반의 유사도 점수와 Dense Vector 기반 유사도 점수를 가중치 합 형태로 결합하는 방식이다 [19].

$$f_{Convx}(q,d) = \alpha \times f_{LEX}(q,d) + (1-\alpha) \times f_{SEM}(q,d)$$

각 변수는 다음과 같은 의미를 가진다.

$\boldsymbol{q}$ : 검색 시스템에 입력된 사용자의 질의(query)를 나타낸다.

$\boldsymbol{d}$ : 검색 시스템이 보유한 문서(document)를 나타낸다.

$\boldsymbol{f_{LEX}(q,d)}$ : 질의와 문서 간의 키워드 기반 유사도 점수를 나타내며, 일반적으로 BM25와 같은 정확한 키워드 매칭 알고리즘을 사용하여 계산된다.

$\boldsymbol{f_{SEM}(q,d)}$ : 질의와 문서 간의 의미적 유사도 점수를 나타내며, 일반적으로 벡터 임베딩 Dense Vector Embedding을 이용하여 계산된다.

$\boldsymbol{\alpha}$ : 키워드 기반 점수와 의미적 유사도 점수의 중요도를 조정하는 가중치 파라미터이다. 이 값은 0에서 1 사이의 범위를 가지며, 검색 시스템의 목적이나 특정 태스크의 특성에 따라 키워드 정확성($f_{LEX}$)과 의미적 유사성($f_{SEM}$) 중 어느 쪽에 더 비중을 둘지 결정한다.

- 이러한 Convex Combination 접근법을 통해, 키워드 정확성을 중시하는 검색과 의미적 유사성을 중시하는 검색을 유연하게 처리할 수 있다.

- 다만, BM25의 점수 분포는 0에서 1사이에 위치하며, Dense Vector 기반의 Cosine 점수 유사도는 -1에서 1사이에 위치하기 때문에, 각 점수를 유사한 범위로 맞추기 위한 정규화를 수행한다. 본 연구에서는 Min-max 정규화를 활용하며, 실제 도출된 점수의 최댓값과 최솟값을 기준으로 0과 1사이로 정규화한다.

- 최종적으로 하이브리드 기반의 유사도 점수가 산출되면, 가장 높은 점수를 가진 도서 데이터를 사전에 정의된 개수만큼 사용자에게 반환한다.

- 반환되는 도서 정보는 답변 활용 정보 섹션에서 작성한 내용과 동일하게, 검색에 활용된 도서 데이터와 도서가 위치한 장소 정보로 구성된다.

## 1.4 생성 모듈 구축

- 생성 모듈 에서는 검색 모듈에서 반환된 도서 정보를 바탕으로 사용자 질의에 적절한 답변을 제공한다.
- RAG 기술을 활용하며, 답변을 생성하는 LLM에게 제공되는 정보는 사용자 질의, 검색 모듈로부터 반환된 도서 정보, 사전에 정의된 프롬프트로 구성된다.
- 프롬프트는 LLM의 페르소나와 행동 양식을 규정하는 명령어이다. 주어진 검색 모듈을 활용해 적절한 도서에 대해 검색한 후, 책 제목과 위치, 추천 이유를 포함하여 답변을 생성하도록 한다. 구체적인 명령어 관련 내용은 아래에서 후술한다.

## 1.5 검색 도우미 구축

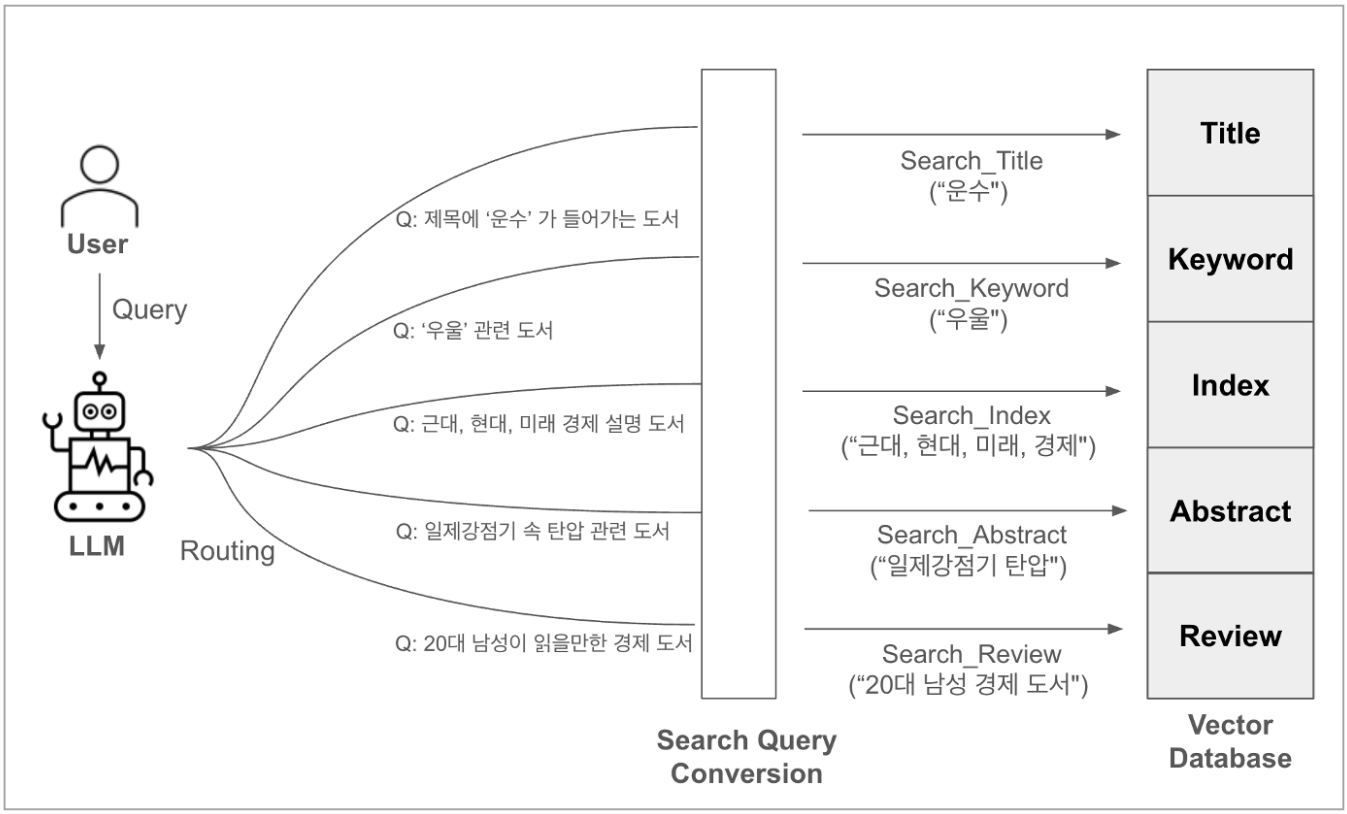



그림 3.3 검색 도우미 예시 (Vector-based Retrieval)

- 일반적인 RAG 파이프라인을 실제 검색 서비스에 적용했을 때 발생 가능한 문제는 크게 두 가지가 존재한다.

- 사용자가 도서 검색과 관련된 질의를 하지 않았음에도 불구하고, 사용자 의도와는 상관없이 도서 검색을 수행하여 어색한 답변이 생성될 수 있다.

  *예를 들어, 사용자가 "너는 무엇을 해주는 서비스야?" 라고 질의했을 때, RAG 파이프라인에 따라 해당 질의와 관련된 도서를 검색한 후 이를 답변에 활용하여 "저는 도서 검색 도우미입니다. 다음과 같은 서비스와 관련된 도서를 추천 드립니다." 와 같은 답변을 생성할 수 있다.*

- 사용자 질의에는 불필요한 정보가 포함되어 있어, 검색에 직접 활용하면 원치 않은 검색 결과가 반환될 수 있다.

  *예를 들어, 사용자가 "나는 중학생 아들을 둔 40대 학부모야. 읽을만한 도서 추천해줘" 라고 요청했을 때, RAG 파이프라인은 해당 질의 전체를 도서와의 유사도 측정을 위해 활용하기 때문에 "중학생"과 "아들" 관련된 도서가 검색될 수 있다.*

- 이것은 일반적인 RAG 파이프라인이 질의의 의도와 무관하게 모든 상황에서 도서를 검색하고 답변을 생성하도록 고정되어 있기 때문이다. 따라서 이러한 한계점을 극복하기 위해 사용자의 질의 의도를 파악하고 답변을 생성하는데 최적화된 검색 도구를 호출하는 검색 도우미를 구축한다.
- 검색 도우미는 LLM이 사용자 질의에 따라 필요한 검색 도구를 선택하여 호출한다. 본 연구에서는 검색 도우미의 동적인 호출을 다음과 같은 세 가지 측면에서 지원한다.

### 1.5.1. 검색 도구 호출 여부

- 검색 도우미에게는 "사용자에게 적절한 응답을 제공하는 검색 도우미" 라는 역할이 부여되며, 도서에 대한 검색을 수행할 수 있는 도구가 제공된다.
- 도서 검색 도구는 검색 질의를 입력하여 호출했을 때 관련된 도서 정보를 받아볼 수 있는 도구로, *book_search(query: string)* 와 같이 함수 호출 형식으로 정의된다.
- 검색 도우미는 사용자 질의에 따라 적합한 검색 도구를 호출할 수 있다.

### 1.5.2. 검색 질의 작성

- 검색 도우미는 일반적인 RAG 파이프라인과 다르게, LLM이 사용자 질의를 이해하고 검색에 적합한 형식으로 가공하여 검색 도구의 입력으로 활용할 수 있다. 즉, “나는 중학생 아들을 둔 40대 학부모야. 읽을만한 도서 추천해줘” 라는 검색 질의에 대해, *book_search(query = “학부모 권장 도서”)* 와 같은 도구 호출이 가능하다.

### 1.5.3. 검색 데이터베이스 선택

- 사용자의 질의 의도에 적합한 검색 결과를 생성하기 위해, 상황에 따라 검색에 최적화된 데이터 유형은 다를 수 있다. 예를 들어, 제목에 파이썬이 들어가는 도서 추천” 이라는 질의는 검색에 활용되는 정보 중 제목 데이터만 필요하며, “우울할 때 읽을 도서” 와 같은 질의는 도서의 키워드나 초록, 리뷰 데이터가 필요할 것이다.
- 주어진 모든 데이터를 활용해 질의와의 유사도를 측정하면 불필요한 정보로 인해 유사도가 낮게 측정되어 원치 않은 결과가 반환될 수 있다.

*예를 들어, “제목에 파이썬이 들어가는 도서 추천”이라는 질의에 대해 “초보자를 위한 파이썬” 이라는 제목을 가진 도서가 실제로는 매우 적합한 도서이다. 하지만 해당 도서의 리뷰가 “초보자에게 안성 맞춤 이에요!” 라면, 사용자 질의와 리뷰의 유사도가 매우 낮게 측정되어 최종 유사도가 낮게 계산됨으로써 검색 결과로 반환되지 못할 수 있다.*

- 따라서, 상황별로 검색에 활용할 데이터 유형을 각각 데이터베이스로 구축하고 검색 도우미가 원하는 데이터베이스를 선택하여 검색을 수행할 수 있도록 구축한다. 즉, 도서의 제목, 키워드, 목차, 가상의 초록, 가상의 리뷰 별로 키워드 및 Dense Vector 데이터베이스를 구축한 후, book_search(query: string, target database: string) 형식으로 호출하도록 하여 원하는 데이터를 기반으로 검색할 수 있도록 지원한다.
- “우울할 때 읽을 도서”와 같은 질의는 단일 데이터베이스로 검색하는 것 보다, 도서의 키워드, 초록, 리뷰 데이터를 모두 활용하여 검색하는 것이 효과적일 것이다. 이처럼, 여러 데이터베이스에 대해 검색하는 유용한 상황을 지원하기 위해 2개 이상의 데이터베이스에 대한 검색을 지원한다.

*예를 들어, 검색 도우미는 book_search((query = ”위로”, target database = ”키워드”), (query = “위안을 주는”, target database = “초록”), (query = “우울할 때 읽을 도서”, target database = “리뷰”)) 와 같이 호출할 수 있다. 각 데이터베이스에서 산출된 유사도를 모두 더하여 최종 유사도를 산출한 후, 최종 유사도가 가장 높은 도서들이 반환된다.*

- 검색 도우미가 사용자 질의에 따라 적합한 데이터베이스를 선택할 수 있도록 In-context Learning을 활용하여 각 데이터베이스마다 2개의 입력 예시와 여러 데이터베이스에 검색을 수행하는 예시를 제공한다.
- 검색 도우미는 자율적인 판단에 따라 검색 도구를 적절하게 활용해 사용자 질의와 관련된 도서를 검색하며, 검색 도구로부터 반환된 도서 정보를 참고하여 도서 제목, 위치, 추천 이유를 기반으로 답변한다.
- 검색 도우미에 활용된 프롬프트는 다음과 같다.

```
You are an excellent Korean-language search assistant.
When a user asks a question related to book searches, you must use the provided search tool to answer the request.

The search tool accepts input in the format [("target database","search query")],
where the possible values for "target database" and their descriptions are as follows:

"title": Database containing book titles. Use this when users explicitly mention the title or request specific book titles.
"keyword": Database containing book keywords. Ideal for thematic searches, general topics, or subject areas broadly related to the query.
"index": Database containing book tables of contents. Select this when users request information about specific content topics within the table of contents or detailed subject areas.
"abstract": Database containing book abstracts or summaries. Prefer this when the user requests books summarized by content themes, readability, or targeted usage scenarios.
"review": Database containing reader reviews of books. Best used for queries seeking subjective recommendations based on emotional states, reader demographics, or user experiences.
"content": Database containing partial text content from books. You must perform a search on this database at least once, regardless of whether you use other databases for searching.

If you want to query the 'title' database, you MUST set the 'query_items' argument as follows:

query_items [("title",<YOUR_SEARCH_QUERY>)]

Select the most suitable database and construct a relevant query based on the type of user's question.
Then, generate a relevant response for the user based on the search tool results.

Below are practical examples of Korean-language queries and corresponding query_items calls:

1. Question: "도서 '파이썬 프로그래밍 입문' 있어요?"
   Call: query_items: [("title","파이썬 프로그래밍 입문")]

... (생략)

12. Question: "파이썬에서 데이터 전처리 기법을 자세히 다룬 책이 있나요?"
   Call: query_items: [("content","파이썬 데이터 전처리")]

You may query multiple databases simultaneously if the user's query requires information that could span across different types of databases.
In such cases, you can query multiple databases simultaneously by specifying multiple pairs, each corresponding to a separate database, in a single tool invocation. NEVER invoke the tool multiple times.
For example, you can query the 'title' and 'keyword' databases simultaneously as follows:
query_items: [("title",<YOUR_SEARCH_QUERY_FOR_TITLE>), ("keyword",<YOUR_SEARCH_QUERY_FOR_KEYWORD>)]

Multiple databases (When the user's query potentially relates to more than one database) examples:

1. Question: "'블록체인'이라는 제목을 가진 책 중, 금융 분야에서 참고할 만한 책이 있나요?"
   Call: query_items: [("title","블록체인"), ("keyword","금융")]

... (생략)

7. Question: "'파이썬' 제목의 책 중에서 빅데이터 분야를 다루고 실제 프로젝트 사례가 포함된 책이 있나요?"
   Call: query_items: [("title","파이썬"), ("keyword","빅데이터"), ("content","프로젝트 사례")]

After retrieving the search tool results, be sure to include the book title, location, and a 1-2 sentence recommendation explaining when responding to the user.
If the retrieved books are deemed irrelevant to the user's query, exclude them from your response.
```

그림 3.4 검색 도우미 활용 프롬프트 예시

# 2. LLM as Judge

## 2.1 평가 시스템 구축

### 2.1.1. 목표

- 평가 시스템은 크게 두 가지를 평가하고 검증하는 것을 목표로 한다.
- 첫 번째는 국립중앙도서관 지능형 RAG 기반 검색 시스템에서 다양한 도서 관련 데이터를 활용했을 때 얻을 수 있는 효용성을 평가하는 것이다.

*예를 들어, "목차 데이터를 검색에 활용하면 특정 질의 유형에 대해 성능이 향상되는 경향성을 보인다." 와 같은 결론을 도출하는 것이 목표이다.*

- 두 번째는 검색 도우미가 생성한 답변의 효용성을 평가하는 것이다. 실제 사람이 검색 시스템을 활용했을 때 느끼는 유용성을 정량적인 지표로 측정하는 것이 목표이다.

### 2.1.2. 평가 방법

#### 2.1.2.1. 다양한 도서 데이터의 효용성 평가

- 검색에 활용할 수 있는 다양한 도서 데이터를 검색 시스템에 활용했을 때와 활용하지 않았을 때 생성되는 답변을 비교하여 해당 데이터에 대한 효용성을 검증한다.
- 답변에 대한 비교 평가를 위해 LLM as Judge를 활용하며, 평가자로서의 LLM은 사용자 질의와 도서 정보, 평가 기준을 가지고 답변 A와 답변 B를 비교하여 어떤 답변이 더욱 우수한지를 평가한다.
- 평가 기준 및 절차는 다음과 같다.

1. 평가 기준

| 구분 | 내용 |
|---|---|
| 정확성(Accuracy) | 응답이 사실에 근거한 올바른 정보를 제공하는가?<br>제공된 도서 정보와의 일관성과 정확성을 확인한다. |

| | |
|---|---|
| 관련성 (Relevance) | 응답이 사용자의 질의를 얼마나 잘 다루는가?<br>질의의 핵심 내용, 맥락, 원하는 형식을 반영하고 있는가? |
| 완전성 (Completeness) | 응답이 필요한 세부사항이나 추천을 충분히 포함하고 있는가?<br>사용자의 질의에 대한 중요한 요소를 빠짐없이 다루고, 명확한 설명 또는 요약과 함께 합리적인 추천을 제공하는가? |
| 명확성 및 간결성 (Clarity and Conciseness) | 정보가 명확하게 표현되어 있어 사용자가 이해하기 쉬운가?<br>불필요하거나 중복되는 정보를 피하고 있는가? |
| 도서 검색에 대한 추가 고려사항 | 사용자가 요청한 경우, 추천 자료의 적절성.<br>제공된 도서 정보를 명확하게 참조하고 있는가? |

2. 비교 절차

위의 각 평가 기준에 따라 옵션 A와 옵션 B를 각각 독립적으로 평가한다.
우수한 응답(Option A 또는 Option B)을 선정하거나, 동점(tie)으로 판단한다.
단, 동점은 '정말로 유의미한 차이가 없는 경우'에만 선택한다.

모든 기준에서의 전반적인 성과를 고려하여 최종 결정을 내린다.

- 답변 A와 답변 B는 같은 질의에 대한 검색 도우미의 답변이지만, 활용한 데이터가 다르다.

  *예를 들어, A는 도서의 제목과 키워드를 기반으로 검색을 수행하고 검색 모듈로부터 반환된 제목과 키워드 정보를 토대로 답변을 생성하지만, B는 제목과 키워드, 목차를 기반으로 검색 및 생성을 수행한다.*

- 활용 데이터에 대한 옵션은 전체 도서에 대해 각 데이터가 존재하는 비율을 기준으로 구분하였으며 (제목과 키워드는 도서의 99% 이상이 가지고 있는 데이터이지만, 그 외의 데이터는 적거나 없는 데이터이다), 각 옵션은 다음과 같다.

  (1) 제목 + 키워드
  (2) 제목 + 키워드 + 목차

(3) 제목 + 키워드 + 목차 + 가상의 초록
(4) 제목 + 키워드 + 목차 + 가상의 초록 + 가상의 리뷰

- 비교 평가는 하나의 질의에 대해 두 개의 답변을 서로 비교하는 방식으로 진행한다. 이때, LLM as Judge가 먼저 나온 답변을 더 선호하는 편향을 줄이기 위해, 답변의 순서를 바꾸어 다양한 비교 쌍을 만든다. 따라서, 하나의 평가 질의마다 총12개(4x3)의 비교 샘플을 생성한다.

- 모든 평가 데이터에 대해 옵션 간의 우위 결과가 도출되면 해당 결과를 종합하여 ELO Rating 측정을 통한 순위 파악, 질의 시나리오 별 승률 비교 등 분석을 수행한다.

- LLM as Judge에 활용된 프롬프트는 다음과 같다.

```
As an expert librarian, your task is to evaluate the responses of two AI assistants to a user's query regarding book searches.
You will compare two options, Option A and Option B, provided in response to the user's query, along with specific book information.
Your goal is to determine which assistant produced a more accurate, relevant, and comprehensive response based strictly on the user's query and the provided book information.

1. Evaluation Criteria:
- Accuracy:
- Does the response provide factually correct information?
- Check consistency and correctness against the given book information.

- Relevance:
- How well does the response address the user's query?
- Does it reflect the main points of the user's question, context, and desired format?

- Completeness:
* Does the response sufficiently cover necessary details or recommendations?
* Ensure all critical aspects of the user's query are thoroughly addressed, with clear explanations or summaries and justified recommendations.

- Clarity and Conciseness:
- Is the information clearly articulated and easy for the user to understand?
- Evaluate if the response avoids unnecessary or redundant information.

- Additional Considerations for Book Searches:
- Appropriateness of recommendations if requested by the user.
- Clear reference to the provided book information.

2. Comparison Process:
- Individually evaluate both Option A and Option B based on each criterion listed above.
- Determine a winner (Option A, Option B) or a tie.
- Provide a brief justification (1-2 sentences) for your decision.
- Provide a final decision by considering the overall performance across all criteria.

3. Output Format:
* `<Winner>` should be either `"A"`, `"B"`, or `"C"`:

* `"A"` if Option A is better in that criterion.
* `"B"` if Option B is better in that criterion.
* `"C"` if it's a tie.
```

그림 3.5 LLM as Judge에 활용된 프롬프트 예시

2.1.2.2. 답변의 효용성 평가

- 사용자 질의와 검색 도우미가 생성한 답변을 사람이 명확한 가이드라인과 관련 도서 정보를 기반으로 평가하도록 한다.

- 총 6명의 평가자가 참여하며, 평가자에게 제공되는 가이드라인은 다음과 같다.

- 1점 (불만족): 질의의도와 일치하지 않은 답변 / 추천 근거가 질의의도와 일치하지 않음
- 2점 (보통): 질의의도와 맞으나 추천도서의 주요 내용을 파악할 수 있는 정보 부족으로 판단하기 어려움
- 3점 (만족): 질의의도와 일치하는 답변과 근거이며 추천도서 주요 내용을 파악할 수 있는 정보도 함께 제공

- 본문 데이터를 포함하여 답변을 생성한 본문 옵션을 포함해, 하나의 질의당 총 5개의 응답에 대해 평가한다.

(5): 제목 + 키워드 + 목차 + 본문

- 질의의 개수가 30개 이므로, 총 150개의 평가 샘플이 생성된다. 각 평가자는 총 25개의 샘플에 대한 평가를 하며, 해당 샘플들은 모두 다른 질의로 구성된다. 또한, 답변 옵션 평가에 대한 치우침 현상을 줄이기 위해, 각 평가자는 5개의 옵션 각각에 대해 5개의 샘플을 할당 받아 평가한다.

- 모든 평가 데이터에 대해 옵션 간의 평가 점수를 도출한 후 분석을 수행한다.

2.1.3. 평가 데이터 구축

- 평가 데이터는 평가 목적에 따라 총 두 가지로 구분된다.

- 첫 번째는 비교 평가를 위한 데이터로, 3가지 시나리오에 대해 각각 30개의 질의로 구성된 90개의 질의 데이터이다.

- 두 번째는 사용성 평가를 위한 데이터로, 위와 동일한 3가지 시나리오에 대해 각각 10개의 질의로 구성된 30개의 질의 데이터이다.

- 두 데이터의 구축 방법은 동일하다. 하지만, 비교 평가 데이터는 6,873건의 도서 데이터를 기반으로 생성되며, 사용성 평가 데이터는 활용할 본문 데이터 711건 기반으로 생성된다.

- 질의는 도서 분류별, 도서 제목, 키워드 데이터를 참고하여 비슷한 주제 및 키워드를 기반으로 범주화 하여 생성된다. 동일 도서 분류 내에서 3개의 시나리오별로 생성되며, 키워드나 답변이 최대한 겹치지 않도록 생성하였다.

- 질의의 시나리오는 '사서에게 물어보세요'의 질의 내용을 분석하여 구성하였으며, 기존 검색 시스템에서 검색이 어려운 시나리오를 중심으로 다룬다.

| 구분 | 내용 |
|---|---|
| 시나리오1<br>- 정확한 정보 없이, 내용의 일부나 키워드로 도서를 찾는 경우 | 사용자가 도서 제목, 저자 등의 메타 데이터를 기억하지 못하여, 내용 일부나 키워드만으로 도서를 찾으려는 경우가 많다. 하지만 기존 검색 시스템은 정확한 키워드 매칭이 필요하므로, 검색 실패율이 높다.<br><br>*예시) 일본어 번역본이며, 인체가 쓰면 쓸수록 퇴화한다는 내용의 도서* |
| 시나리오 2<br>- 2개 이상 주제에 대한 도서를 추천 받고자 하는 경우 | 사용자가 주제 중심으로 도서를 찾고자 하는 경우가 존재한다. 기존 시스템은 키워드 검색 기반으로 동작하므로, 문맥을 이해하여 관련 자료를 제공하는 것에는 한계가 있다. 특히, 다수의 키워드가 등장하는 경우 매칭률이 현저히 감소한다.<br><br>*예시) 결혼이민자 가족에 대한 현황과 지원정책 개선 방안에 대하여 전국 자료와 전북지역 관련 자료* |
| 시나리오 3<br>- 연령, 감성 등 개인적인 상황에 맞는 도서 추천을 받고자 하는 경우 | 사용자의 나이, 상황, 정서적 니즈에 맞는 개인화 추천 요청이 증가하고 있다. 하지만 기존 검색 시스템은 연령 대나 감성에 기반한 추천 기능이 존재하지 않는다.<br><br>*예시) 초등학교 2학년 자녀가 있는 학부모가 읽을 만한 도서* |

- 시나리오 1에서는 본문과 키워드 내용을 중심으로, 시나리오 2에서는 2개 이상의 주제와 키워드를 종합하여, 시나리오 3에서는 연령, 기분, 정서, 감성 중 최소 한 개 이상 연관된 질의로 생성한다.

### 2.1.4. 분석 결과

#### 2.1.4.1. 다양한 도서 데이터의 효용성 평가

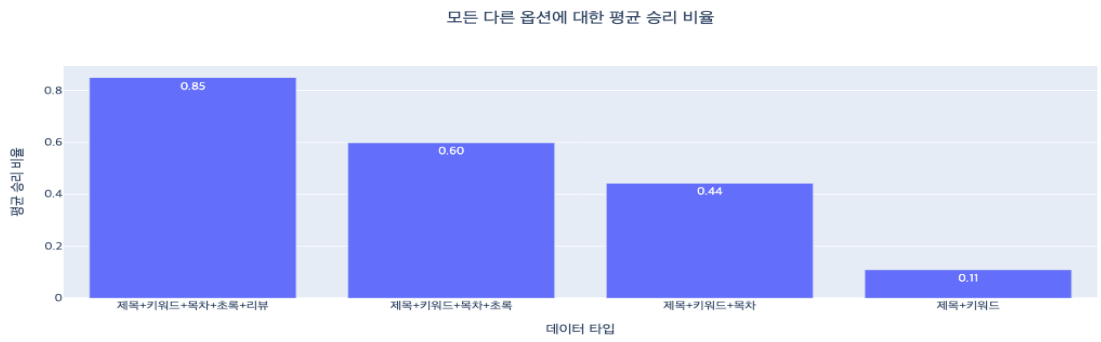



그림 3.6 데이터 활용 옵션에 따른 비교 평가 평균 승률

- 그림 3.6에서 확인할 수 있듯, 답변 옵션 (4) > (3) > (2) > (1) 순서로 우세하다는 결론이 도출되었다. 각 옵션에 대한 ELO Rating은 다음과 같다.

(1) 제목 + 키워드 : 747

(2) 제목 + 키워드 + 목차: 968

(3) 제목 + 키워드 + 목차 + 초록: 1063

(4) 제목 + 키워드 + 목차 + 초록 + 리뷰 :1224

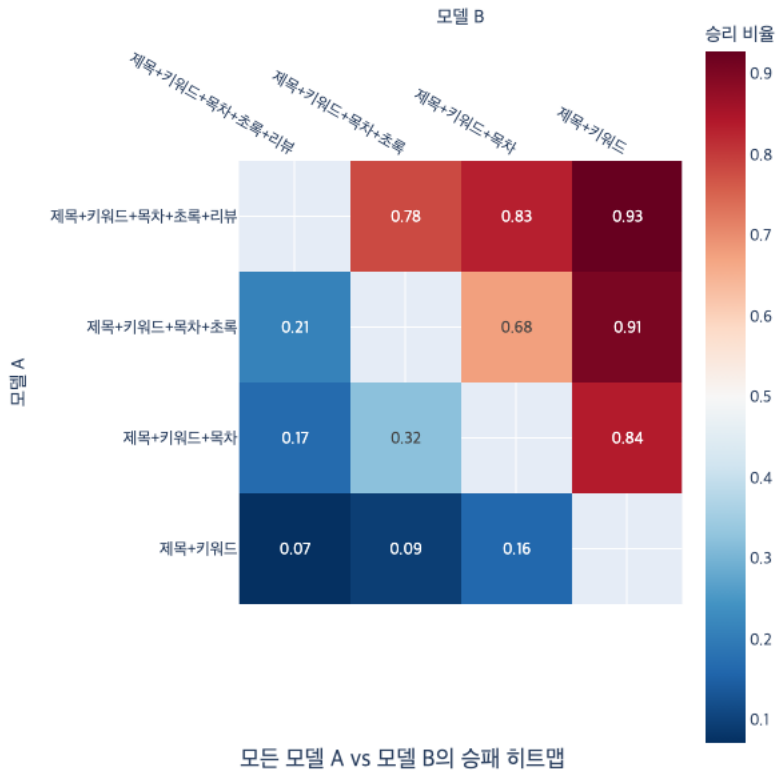



그림 3.7 다른 데이터 활용 옵션끼리 측정한 비교 평가 평균 승패 비율 히트맵

- 그림 3.7은 활용되는 데이터의 차이가 클 수록 유의미한 차이를 보인다. 도서의 제목, 키워드, 목차, 초록 및 리뷰를 모두 함께 활용하여 답변한 옵션 (4)은 제목과 키워드만을 활용한 옵션 (1)와 비교했을 때 93%의 승률을 보이며, 이는 목차와 초록을 함께 활용한 옵션 (3)와의 승률인 78%와 비교했을 때 상당히 높은 수치임을 알 수 있다. 이러한 경향성은 모든 답변 옵션에 대해 유지된다. 옵션 (3)의 옵션 (1)에 대한 승률은 91%인데 반해, 목차까지 활용한 옵션 (2)와의 승률은 68%을 보인다.

그림 3.8 시나리오 별 비교 평가 평균 승률




- 활용하는 데이터의 차이가 클 수록 답변에 대한 비교 평가 우위는 시나리오 1, 2, 3 각각에 대해서도 동일한 현상을 보이며, 이는 그림 3.8에서 확인할 수 있다. 특히 시나리오 2에서 각 옵션 간의 차이가 소폭 증가하는 경향성을 보인다.

### 2.1.4.2. 답변의 효용성 평가

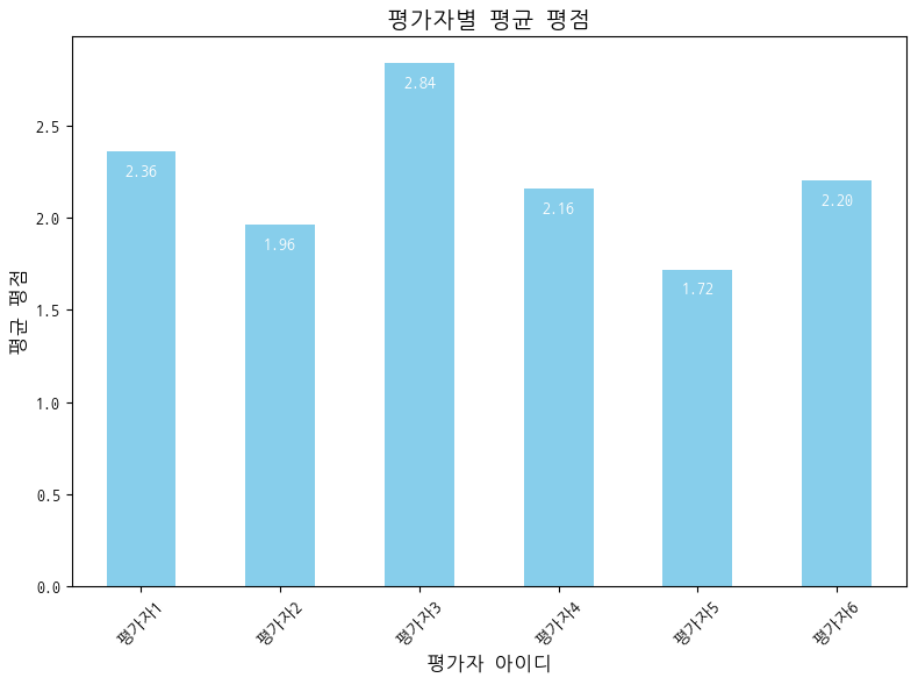



그림 3.9 평가자 별 평균 효용성 점수

- 평가자 별 평균 점수 측정 분포는 그림 3.9와 같다. 평균 점수가 가장 높은 평가자 3은 2.84점 > 평가자1(2.36) > 평가자6(2.20) > 평가자4(2.16) > 평가자2(1.96) > 평가자 5(1.72)이며, 모든 참가자의 점수 평균은 2.21점으로 나타났다.

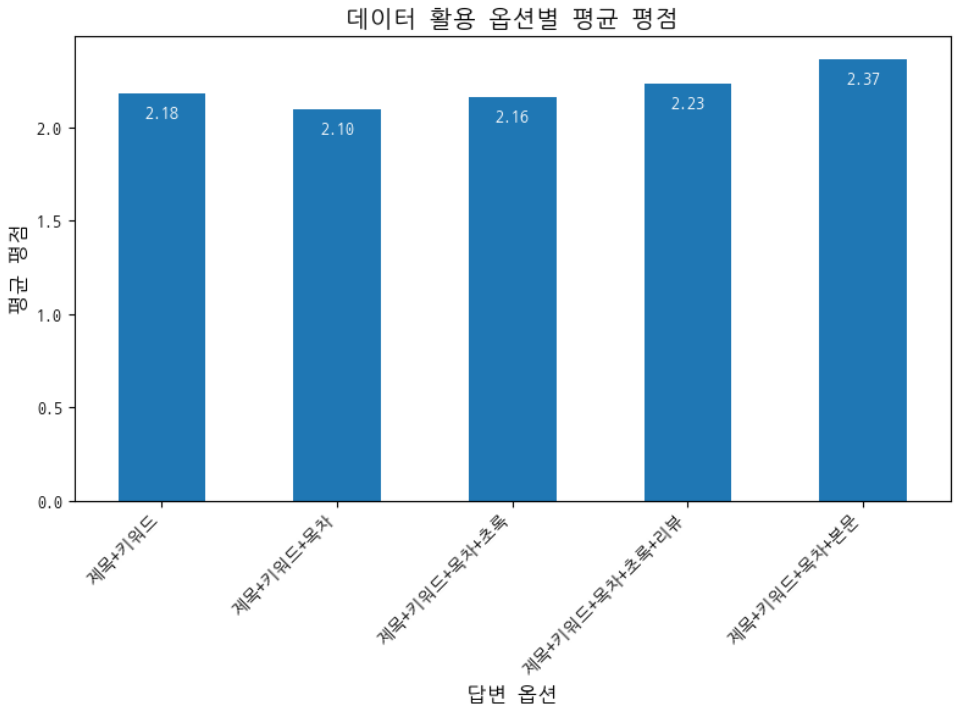



그림 3.10 데이터 활용 옵션에 따른 평균 효용성 점수

- 활용 데이터 별 평균 점수 분포는 그림 3.10과 같다. 옵션 5(2.37)가 가장 높은 점수, 그 다음으로 옵션 4(2.23) 순으로 정보가 많을 수록 높은 점수를 보인다.

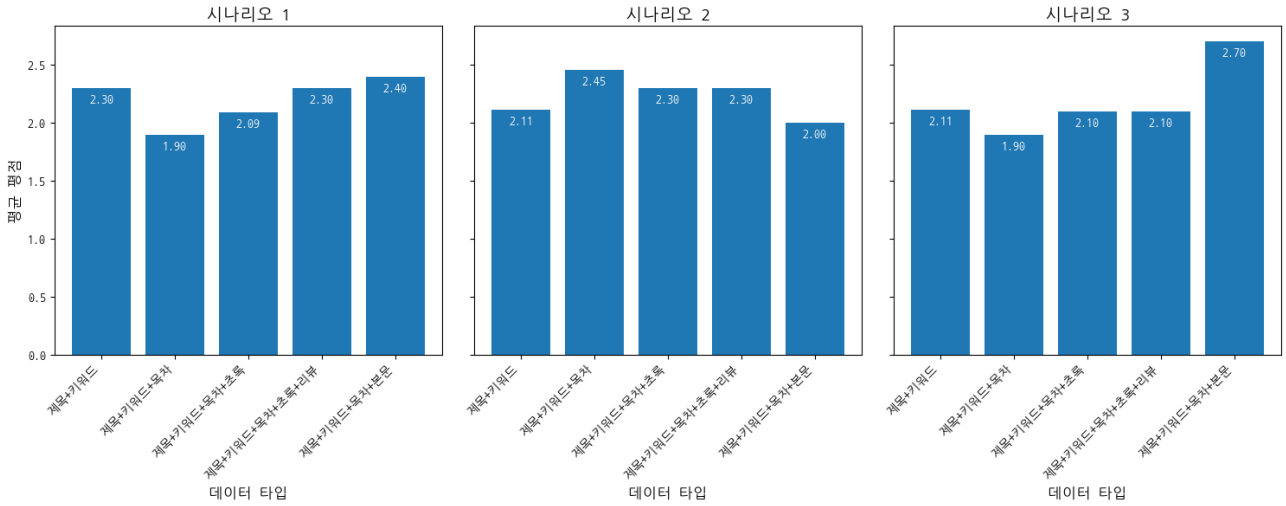



그림 3.11 시나리오 별로 구분한 데이터 활용 옵션에 따른 평균 효용성 점수

- 각 시나리오별 데이터 활용 옵션에 따른 평균 효용성 점수는 그림 3.11에서 확인할 수 있다. 시나리오 2에서는 옵션 2(2.45)가 가장 높은 평균 점수이고, 시나리오 3에서는 옵션 1-4의 평균점수(2.05)가 옵션5(2.70)에 비해 큰 차이로 낮은 것을 확인할 수 있는 걸로 보아, 개인적인 상황과 관련된 사용자 질의에 대해서는 본문 내용을 기반으로 상세히 답변하는 것을 선호하는 것을 확인할 수 있다.
- 전반적으로 활용 데이터가 많을수록 높은 점수를 보이는 경향을 보였다.

## 2.2 Human Alignment 측정

### 2.2.1. 목표

- LLM as Judge를 기반으로 한 비교 평가의 신뢰성을 확보하기 위해 사람의 평가와 비교하여 Alignment를 측정한다.

### 2.2.2. 측정 방법

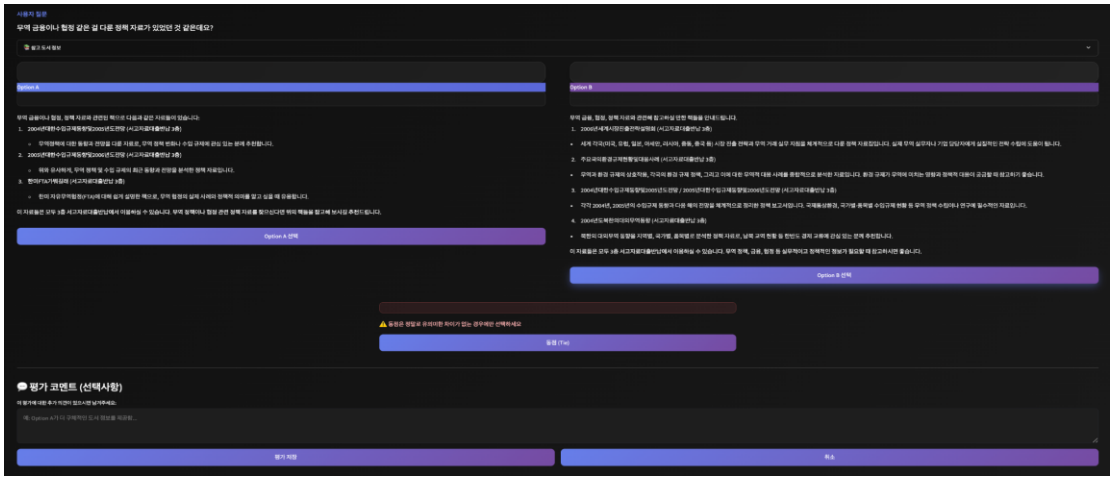



그림 3.12 Human - Pair-wise Evaluation (비교평가) 예시

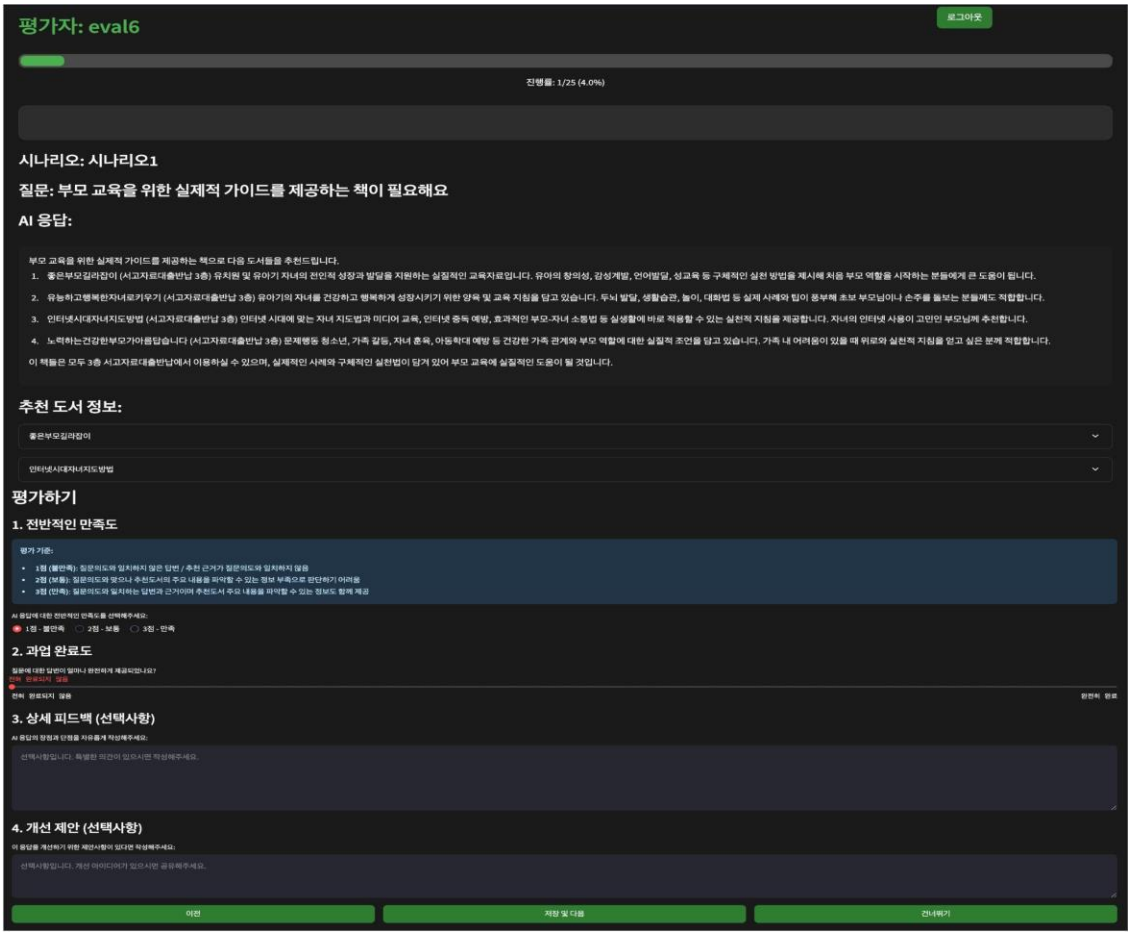



그림 3.13 Human - Point-wise  Evaluation (정성평가) 예시

- 평가자 간 일치도(IAA, Inter Annotator Agreement) 측정을 위해 총 9명의 평가자가 참여하며, 90개의 질의 전체에 대한 평가를 수행한다. 각 평가 샘플은 질의 한 개와 두 개의 답변 옵션으로 구성되며, 평가자는 LLM과 동일한 가이드라인과 도서 정보를 통해 비교 평가를 수행한다.

- 평가 샘플 구성 방법은 다음과 같다.

  - 모든 질의가 사용자에 의해 평가되는 것을 목표로 한다.
  - 각 질의는 4개의 답변 옵션이 존재하므로 순서를 고려하여 총 12개의 답변 옵션이 나오지만, 1080개의 평가 샘플을 9명의 평가자가 모두 평가하는 것은 많은 시간이 들어가는 작업이다.
  - 따라서, 각 질의에 대해 한 개의 답변 비교 쌍을 구축한다. 예를 들어, 어떤 질의는 (4)와 (3), 다른 질의에 대해서는 (2)와 (1)와 같이 평가 샘플을 구축하는 식이다. 이 때, 같은 비교 쌍에 대해서만 평가 샘플이 구축되지 않도록 하기 위해 (예를 들어, (1)은 항상 (2)와 비교 평가를 수행하는 상황이 나오지 않도록 하기 위해) 옵션 쌍에 따라 골고루 분배하여 구축하였다.
  - 각 질의에 대해 총 1개의 평가 샘플이 생성되며, 총 90개의 평가 샘플이 구축된다.

- 하나의 평가 샘플을 한 명의 평가자가 평가하게 되면 개인의 주관에 따라 사람 평가 결과가 설정될 수 있다. 따라서, 모든 평가 샘플에 대해 3명의 평가자가 평가하도록 설계한다.
- 총 90개 평가 샘플에 대해 총 3번의 평가가 이루어지며, 각 평가자는 30개의 평가샘플에 대한 평가를 수행한다. 총 270개의 평가 샘플은 순서를 고려하지 않은 모든 고유한 옵션 쌍 6개가 각각 45회씩 평가되도록 구성된다.

- 평가자 간 일치도(IAA) 측정은 다음과 같이 이루어진다.

  - 비교 평가를 수행하는 LLM과 사람 사이의 일치도 측정으로, LLM이 사람과 얼마나 유사한 평가를 하는지 측정한다. 각 평가 샘플은 3명의 사람이 평가에 참여하므로, 다수결 투표를 통해 각 평가 샘플에 대한 사람의 평가를 한 개의 라벨로 종합한다. 만약 모든 평가자가 다른 평가를 수행할 경우, 해당 샘플은 일치도 측정에서 제외한다.
  - 사람과 LLM의 평가결과를 비교하여 일치여부를 확인하며, Cohen's Kappa 계수 및 일치율을 측정한다 [20].

2.2.3. 분석 결과

- 사람의 평가를 종합한 결과, 모든 평가자가 다른 평가 결과를 도출한 2개의 샘플을 제외하여 총 88개의 평가 샘플에 대한 사람의 평가 결과를 얻었다.

- 88개의 평가 샘플에 대해 Cohen's Kappa 계수는 0.419로 중간 정도의 일치도를 보였으며, 평가의 정확한 여부를 측정했을 때 72%의 일치율을 보였다.

- 이러한 결과는 LLM as Judge가 사람의 평가를 어느 정도 대변할 수 있는 효율적인 방법론임을 보여준다.

# 3. 본문 활용 가능성 탐색

## 3.1 배경

지능형 검색서비스 구현에 있어, 기존 도서관 검색 시스템은 주로 책제목, 저자, 출판사, 키워드, 목차 등 정형화된 메타데이터에 기반하고 있다. 그러나 이러한 메타정보만으로는 사용자의 복합적인 질의나 문맥 중심 탐색 요구를 충분히 수용하기 어렵다. 특히 앞서 지적한 바와 같이, 이용자가 도서의 일부 내용이나 특정 장면을 기억하고 이를 단서로 정보를 탐색하려는 경우, 키워드 기반 검색은 응답률과 정확도 측면에서 한계를 보인다.

이러한 한계를 극복하고 검색 정확도와 만족도를 높이기 위해서는 도서의 본문 데이터를 적극 활용할 필요가 있다. 본문은 책의 내용을 직접적으로 담고 있는 핵심 데이터로서, 문맥 이해 및 의미 기반 검색에서 가장 풍부한 정보를 제공할 수 있는 자원이다. 국립중앙도서관이 보유한 방대한 디지털 도서 중 다수는 본문 데이터까지 포함하고 있으며, 이는 지능형 검색서비스와 대규모 언어모델(LLM)의 성능을 극대화할 수 있는 잠재적 자원이라 할 수 있다.

그러나 현재 도서 본문은 다양한 이유로 검색 서비스에 실질적으로 활용되지 못하고 있다. 가장 대표적인 이유는 많은 본문 데이터가 PDF 이미지 형태로 저장되어 있어 기계가 해석할 수 있는 텍스트 형태로 변환되지 않았다는 점이다. 이에 따라 기존 시스템은 본문 내용 전체를 검색 대상으로 삼지 못하고 있으며, 검색 시스템은 여전히 메타정보에 제한된 결과만을 제공한다.

본 연구는 이러한 문제의식 하에, 본문 데이터를 디지털 텍스트화하여 검색 서비스에 활용할 경우 메타정보 대비 어느 정도의 효용성과 성능 향상이 있는지를 실증적으로 검토하고자 한다. 이를 통해 본문 정보의 구조화 및 벡터화 가능성을 실험하고, 향후 본문을 중심으로 한 고도화된 의미 기반 검색서비스의 구현 가능성을 제시하고자 한다.

## 3.2 분석 결과

본 연구에서는 도서 본문 데이터를 지능형 검색서비스에 실질적으로 활용하기 위해 전처리 방식을 실험하였으며, 그 활용 가능성과 효과성을 정성적으로 평가하였다. 특히, 답변의 효용성 평가 섹션에서 확인할 수 있었 듯, 본문 데이터를 검색 도우미에 활용했을 때 다른 옵션 대비 가장 높은 평가 점수를 보이며 이는 연령, 감성 등 개인적인 상황에 맞는 도서 추천 시나리오에서 더욱 두드러진다.

또한, 답변의 효용성 평가에서 수합한 피드백 및 개선 사항을 분석한 결과, 사용자들은 자세한 추천 근거 및 관련 도서에 대한 내용 언급을 선호한다. 따라서, 본문 데이터를 검색 도우미의 답변에 활용하도록 함으로써 사용성을 증가시킬 수 있을 것이라 기대된다.

### 3.2.1 본문 전처리 및 정제 과정

도서 본문은 대부분 PDF 이미지 또는 스캔본 형태로 제공되기 때문에, 기계 판독을 위한 전처리 과정이 필수적이다. 이를 위해 본 연구는 다음과 같은 단계적 접근을 통해 본문을 디지털 텍스트화하였다:

- 본문범위 : 압축된 본문 데이터 파일이 매우 크기 때문에, 한정된 자원에서 전체 파일을 압축 해제하여 다루기 어렵다. 따라서, 파일 크기가 가장 작은 1,000건의 본문을 식별한 후 해당 파일들만 압축을 해제한다.
- 텍스트 추출 방식 검토 : 1,000건의 본문 PDF 문서들에 대해, 파이썬 라이브러리를 활용하여 텍스트를 추출한다. 오픈소스 라이브러리인 PyMuPDF 라이브러리를 활용한다.
- 에러 텍스트 검출 : 일부 추출된 텍스트는 인코딩 오류, 문단누락, 잡음 등으로 활용이 불가능한 수준이었다. 이에 따라 LLM을 활용하여 텍스트 오류를 자동 탐지하고 정상 여부를 판별하는 검수 과정을 적용하였다

그 결과, 전체 6,873건 중 약 711건의 도서 본문이 검색가능한 문맥단위로 정제되었으며, 이는 이후 벡터 임베딩 및 의미 기반 검색에 활용되었다.

### 3.2.2 본문 기반 검색의 활용 효과

답변의 효용성 평가 결과, 본문 데이터를 활용한 응답이 다른 데이터 조합 대비 가장 높은 평균 점수를 기록하였다. 특히 시나리오 3에서 그 효과가 두드러졌으며, 사용자의 감성적 상태, 상황적 배경에 맞는 도서 추천 요청에 대해 보다 유의미한 답변이 생성되었다.

*예시: “회사 퇴직을 준비하는 사람들을 위한 책 추천해줘”라는 질의에 대해, 목차나 키워드만으로는 적절한 응답을 생성하기 어려우나, 본문 내용을 활용하면 ‘퇴직 후 소득과 자산, 노후 빈곤 가능성에 대해 구체적으로 분석한 내용’을 근거로 삼아 보다 구체적이고 설득력 있는 추천이 가능하였다.*

### 3.2.3 본문 데이터의 필요성과 전략적 활용 영역

키워드가 제공하지 못하는 정보의 심층성과 다양성을 고려할 때, 본문 데이터는 다음과 같은 영역에서 검색 품질을 향상시킬 수 있는 자원임이 입증되었다:

- 문맥 기반 탐색: 사용자가 단편적 키워드나 기억의 조각으로 도서를 탐색할 때, 본문 내용을 통한 의미 기반 검색이 훨씬 높은 적합도를 보인다.
- 개인화 추천: 감성, 정서, 연령대 등을 반영한 질의에 대해, 본문은 사용자의 상황과 공감대를 형성할 수 있는 언어적 근거를 제공한다.

- 주제별 깊이 있는 분석: 동일한 주제(예: ‘기후변화’, ‘교육 불평등’)에 대해 서로 다른 맥락을 제공하는 도서를 식별하는 데 본문이 핵심적으로 기여할 수 있다.

특히, 향후에는 본문 기반 요약(Summarization), 중요 문단 자동 추출, 챕터 기반 검색 등의 고도화된 기술을 접목함으로써, 본문 활용의 정밀도와 확장성을 더욱 높일 수 있을 것이다.

# 제 4장. 결 론

## 1. 연구요약

본 연구는 국립중앙도서관의 도서 데이터를 기반으로, 문맥 이해 및 의미 기반 검색을 지원하는 지능형 검색 도우미를 구축하고, 그 성능과 효용성을 RAG Framework와 LLM 기반 평가 시스템을 통해 실증적으로 검증하는 데 목적이 있다.

우선, 기존 키워드 중심 검색의 한계를 극복하기 위해, 사용자의 자연어 질의를 이해하고 적합한 정보를 추출 및 생성할 수 있는 도서 검색 도우미를 설계하였다. 본 도우미는 키워드 검색과 벡터 검색을 결합한 하이브리드 검색 모듈과, 검색된 정보를 바탕으로 답변을 생성하는 생성 모듈로 구성된다. 하이브리드 검색을 구성하는 키워드 기반 검색에는 BM25 알고리즘을, 의미 기반 검색에는 OpenAI의 Dense Embedding 모델(text-embedding-3-large)을 사용하였으며, Dense 벡터 인덱싱 엔진은 오픈소스 라이브러리인 FAISS를 활용하였다.

데이터 분석 단계에서는 국립중앙도서관이 제공한 6,873건의 도서 메타데이터와 711건의 본문 데이터를 대상으로 전처리를 수행하였다. 도서 제목과 목차와 같은 기존 데이터 외에도, 가상의 초록과 리뷰를 OpenAI의 GPT-4.1 기반으로 합성하며 검색에 활용할 수 있도록 데이터베이스화하였다. 본문 데이터는 파일의 크기가 작은 711건의 PDF 데이터에서 텍스트를 추출한 후 LLM을 통해 텍스트 추출 오류 여부를 검증하는 절차를 통해 선별되었다.

검색 모듈은 아래 데이터 조합별로 유사도 계산을 수행하며, 사용자의 질의 유형에 따라 적절한 데이터베이스를 선택하여 검색할 수 있는 에이전트 방식으로 설계되었다. 이는 일반적인 Static RAG (특정한 고정된 데이터베이스 하나만을 사용하여 검색을 수행하는 방식) 파이프라인의 한계를 극복하고, LLM이 검색 도구 호출 여부, 질의 가공, 검색 대상 데이터베이스를 동적으로 선택할 수 있는 유연한 구조이다.

(1): 제목 + 키워드
(2): 제목 + 키워드 + 목차
(3): 제목 + 키워드 + 목차 + 가상의 초록
(4): 제목 + 키워드 + 목차 + 가상의 초록 + 가상의 리뷰

이후, 검색 도우미의 성능 검증을 위해 LLM as Judge 평가 시스템을 구축하였다. 이 시스템은 GPT 기반 평가 시스템이 두 가지 답변을 비교 평가하여 검색 데이터의 효용성과 생성 응답의 품질을 평가하도록 설계되었다. 평가 기준은 정확성, 관련성, 완전성, 명확성·간결성 등으로 구성되며, Pairwise 방식의 비교 평가를 통해 평가된다. 평가시스템의 신뢰성은 Human 평가 결과와의 평가 일치도를 측정함으로써 확보하였다. 총 90개의 비교 평가용 질의와 30개의 사용성 평가 질의를 시나리오별로 구성하였으며, 시나리오는 다음 세 가지 유형으로 구분된다:

| **구분** | **내용** |
|---|---|
| 시나리오1<br>- 정확한 정보를 몰라, 내용의 일부나 키워드로 도서를 찾는 경우 | 사용자가 도서 제목, 저자 등의 메타 데이터를 기억하지 못하여, 내용 일부나 키워드만으로 도서를 찾으려는 경우가 많다. 하지만 기존 검색 시스템은 정확한 키워드 매칭이 필요하므로, 검색 실패율이 높다.<br><br>*예시) 일본 번역책이며, 인체가 쓰면 쓸수록 퇴화한다는 내용의 도서* |
| 시나리오 2<br>- 2개 이상 주제에 대한 도서를 추천 받고자 하는 경우 | 사용자가 주제 중심으로 도서를 찾고자 하는 경우가 존재한다. 기존 시스템은 키워드 검색 기반으로 동작하므로, 문맥을 이해하여 관련 자료를 제공하는 것에는 한계가 있다. 특히, 다수의 키워드가 등장하는 경우 매칭률이 현저히 감소한다.<br><br>*예시) 결혼이민자 가족에 대한 현황과 지원정책 개선 방안에 대하여 전국 자료와 전북지역 관련 자료* |

| | |
|---|---|
| 시나리오 3<br>- 연령, 감성 등 개인적인 상황에 맞는 도서 추천을 받고자 하는 경우 | 사용자의 나이, 상황, 정서적 니즈에 맞는 개인화 추천 요청이 증가하고 있다. 하지만 기존 검색 시스템은 연령 대나 감성에 기반한 추천 기능이 존재하지 않는다.<br><br>*예시) 초등학교 2학년 자녀가 있는 학부모가 읽을 만한 도서* |

각 시나리오에 따라 검색 및 답변에 활용된 데이터 조합의 성능 차이를 비교하여 데이터 항목별 기여도를 정량적으로 분석하였다.

# 2. 연구결과

## 2.1 LLM Judge 기반 성능 평가

본 연구에서는 RAG 기반 검색 도우미의 성능을 정량적으로 검증하기 위해 LLM as Judge 평가 방식을 도입하였다. 그 결과, 도서 검색 시 활용되는 데이터 조합이 풍부할수록 응답 품질이 향상되는 경향성이 명확히 나타났다.

- ELO Rating 기준으로, 가장 간단한 데이터 조합(옵션 1: 제목+키워드)은 747점, 가장 많은 데이터를 활용한 조합(옵션 4: 제목+키워드+목차+초록+가상리뷰)은 1224점을 기록하였다.
- 특히 옵션 4는 옵션 1에 비해 93%의 우세 승률을 보였으며, 이는 옵션 간 정보량 차이가 클수록 검색 응답의 품질 차이도 뚜렷하게 나타남을 시사한다.
- 시나리오 유형별로도 동일한 경향이 유지되며, 시나리오 2(주제 기반 탐색)에서 데이터 조합 간의 성능 격차가 특히 크게 나타났다.

Human 평가 기반 효용성 평가에서는 사용자 질의에 대해 본문 데이터를 포함하여 생성된 응답(옵션 5)이 가장 높은 평균 점수를 기록하였으며, 특히 개인적 상황과 관련된 시나리오 3에서 우수한 평가를 받았다. 이는 본문이 사용자의 정서적·맥락적 요청에 대응하는 데 효과적이라는 것을 입증한다.

## 2.2 Human Alignment 측정 결과

LLM 평가의 신뢰성을 검증하기 위해 사람과의 일치도(Alignment)를 분석한 결과, 총 88개의 평가 샘플 중 72%가 사람과 동일한 판단을 내렸으며, Cohen's Kappa 계수는 0.419로 중간 수준의 일치도를 나타냈다. 이는 LLM as Judge 방식이 사람의 판단을 일정 수준 대변할 수 있는 평가 Framework임을 시사한다.

- 평가자는 총 9명이 참여하였으며, 각 평가 샘플에 대해 3인 이상의 다수결 평가 방식으로 사람 라벨을 생성하였다.
- 옵션 간 비교쌍의 균형 있는 분포와 순서 무작위화를 통해 평가 편향을 최소화하였다.

## 2.3 본문 데이터 활용 가능성 탐색

기존의 키워드 중심 검색 시스템은 복합적·주관적인 사용자 질의에 대응하는 데 한계를 보였다. 본 연구는 이에 대한 대안으로 도서 본문을 디지털 텍스트화하여 검색에 활용하고, 그 실효성을 실험적으로 검토하였다.

- 평가 결과, 본문 데이터를 포함한 응답이 전반적으로 가장 높은 품질 평가를 받았으며, 특히 감성·정서 기반 추천을 요구하는 시나리오 3에서 그 효과가 뚜렷하였다.
- 사용자 피드백에서도 "도서의 구체적인 내용 인용"이나 "추천 근거에 대한 서술"을 긍정적으로 평가한 점이 확인되어, 본문 정보가 검색 도우미의 실용성과 만족도 제고에 기여함을 보여준다.

## 2.4 종합 결론

본 연구는 국립중앙도서관 데이터를 활용한 RAG 기반 지능형 검색 도우미의 구현과 성능 검증을 통해 다음과 같은 주요 결론을 도출하였다.

① 복합 데이터 조합의 활용이 검색 응답 품질에 결정적 영향을 미치며, 초록, 리뷰 데이터를 추가로 활용하면 검색 및 응답 품질 향상에 도움이 된다.

② 벡터 데이터베이스와 LLM을 결합한 자연어 기반 검색 시스템은 실제 구현 가능성과 성능 측면 모두에서 유효성이 입증되었다. 키워드 기반 검색(BM25)과 의미 기반 검색(Dense Embedding)의

하이브리드 전략은 다양한 질의 유형에 대해 높은 검색 정확도와 유연한 응답 생성을 가능하게 하였으며, 사용자 질의에 대해 자연스러운 응답을 생성하는 데 효과적으로 작동하였다.

③ 본문 데이터는 개인화된 추천 및 문맥 기반 질의에 가장 중요한 자원이며, 향후 도서관 정보검색 시스템의 핵심 자산으로 활용 가능성이 높다.

따라서 본 연구는 기존 메타정보 기반 검색 방식의 한계를 보완하고, 사용자 중심의 고도화된 의미 기반 검색 환경으로의 전환 가능성을 기술적·실증적으로 제시한 점에서 중요한 시사점을 제공한다.

# 3. 기대 효과

본 연구는 RAG 기반 지능형 검색 도우미 구축 및 성능 평가를 통해 기존 키워드 기반 검색 시스템의 한계를 극복할 수 있는 실질적 대안을 제시한다. 특히, 다양한 데이터 조합과 LLM 기반 응답 생성을 결합함으로써, 다음과 같은 기대효과를 도출할 수 있다.

## 3.1 문맥 이해 기반의 검색 정확도 향상

기존 검색 시스템은 도서 제목이나 키워드와 같이 제한된 메타정보에 의존하며 키워드 매칭 기반으로 검색이 수행되기 때문에 사용자의 주관적·복합적 질의에 효과적으로 대응하지 못하였다. 그러나 본 연구에서 제안한 검색 도우미는 목차, 초록, 리뷰 등 다층적 데이터 활용하고 하이브리드 검색을 수행하여 의미 기반의 검색을 수행하며, 주어진 질의에 따라 동적으로 검색함으로써 높은 정확도의 검색 결과를 도출할 수 있었다.

- 기존의 검색 시스템이 다룰 수 있는 시나리오를 넘어, 문맥적 이해가 필요한 질의에 대해 높은 품질의 답변 결과를 생성한다.
- 전체 승률 기준, 데이터 조합이 증가할수록 응답 품질이 명확히 개선되었으며, 가장 많은 정보를 포함한 옵션4(타이틀+키워드+목차+초록+가상리뷰)의 승률은 0.85로 옵션1(타이틀+키워드)의 0.11 대비 7.7배 높은 성과를 보였다.

- 이는 지능형 검색서비스에서 단일 키워드의 한계를 넘어, 복합 데이터의 융합이 검색 품질을 향상하는데 주요한 요소임을 확인할 수 있었다.

## 3.2 시나리오별 최적화 전략 도출

시나리오 2(주제 중심 탐색)의 경우, 다중 키워드 조합이 요구되기 때문에 다른 시나리오보다 데이터 조합의 변화에 따른 성능 향상폭이 가장 뚜렷하였다. 실제로 옵션 4의 승률은 시나리오 2에서 0.91로 나타나, 모든 시나리오 중 가장 높은 수치를 기록하였다.

- 이는 정책 분석, 지역 비교, 사회 이슈 탐색과 같은 복합 주제 기반 질의에서 심화된 데이터 구조와 맥락 이해 능력을 갖춘 검색 시스템의 필요성을 보여준다.
- 향후 분야별 전문 도서관 및 아카이브 시스템에서, 주제 적합성에 따라 데이터 조합을 유동적으로 구성하는 검색 전략이 유효함을 시사한다.

## 3.3 데이터 항목별 활용도 분석을 통한 설계 가이드 확보

각 데이터 항목이 검색 품질에 미치는 영향을 분석한 결과, 목차(0.33), 가상리뷰(0.25), 초록(0.16) 순으로 승률 비율이 높게 나왔다.

- 목차는 도서의 전체 구조와 주제 흐름을 파악할 수 있어, 문맥 기반 검색에서 핵심적인 데이터로 활용 할 수 있음을 확인하였다.
- 가상리뷰는 사용자의 감성이나 상황에 따른 추천 품질을 보완하는 데 유용하며, 특히 시나리오 3(개인화 추천)에서 의미 있는 기여를 하였다.
- 초록은 정보 요약 기능은 있으나, 목차나 리뷰 대비 질의 적합성에서는 상대적으로 기여도가 낮은 것으로 나타났다.

이러한 분석 결과는 향후 지능형 검색시스템 설계 시, 우선적으로 수집·정제해야 할 데이터 항목의 우선순위를 판단할 수 있는 실증적 가이드라인으로 활용될 수 있다.

### 3.4 실용성과 확장성 측면의 기여

본 연구는 국립중앙도서관이 보유한 기존 데이터를 기반으로 가공 가능한 정보(예: 가상 초록·리뷰)를 생성하고 검색 효율성을 향상시키는 구조를 입증함으로써, 기존 데이터 환경에서도 RAG 기반 검색 기술이 효과적으로 적용될 수 있음을 실증하였다.

- 또한, 검색 도구 호출 여부, 질의 가공, 대상 데이터 선택을 LLM이 동적으로 판단할 수 있도록 설계함으로써, 유연성과 사용자 적응성이 높은 검색 도우미 구조를 제안하였다.
- 이는 향후 공공 데이터포털, 전문 분야 디지털 아카이브 등으로의 확장 가능성과 기술적 전이 가능성 측면에서도 중요한 시사점을 제공한다.

## 4. 한계점

- 본 연구에서 사용한 데이터 중 합성 데이터에 대한 엄밀한 검증이 부족하다. 본 연구에서 사용된 도서의 초록 및 리뷰는 목차 데이터를 기반으로 LLM을 통해 합성한 가상의 데이터이다. 이로 인해 도서 내용과 정확히 일치하지 않거나 부정확한 정보가 포함되었을 가능성이 있다. 보다 정확한 검색 및 추천 성능을 위해서는 실제 초록과 독자의 리뷰와 같은 신뢰성 높은 데이터를 활용하거나, 합성된 데이터를 검증할 수 있는 체계적인 방법이 필요하다.
- 질의 시나리오의 다양성이 부족하다. 본 연구에서는 실제 문의 사례를 기반으로 주요 시나리오를 도출하여 총 120개의 질의를 평가를 위해 제작하였으나, 도서 검색이라는 서비스의 특성상 사용자의 질의 형태는 매우 다양하여 본 연구에서 다루지 못한 유형의 질의가 존재할 가능성이 높다. 따라서, 실제 사용자의 검색 로그 분석이나 설문조사를 통해 보다 풍부하고 현실적인 시나리오를 추가로 발굴하고, 이에 따른 시스템 대응 능력을 강화하는 방식으로 검색 도우미를 고도화 할 수 있을 것이다.
- Human 평가 과정에서 평가 가이드라인과 무관한 주관적인 평가가 일부 이루어졌다. 본 연구에서는 LLM as Judge의 타당성과 사용성 평가를 위해 15명이 평가를 수행하며, 분명한 목적과 평가 기준에 의거하여 평가가 진행되어야 한다. 하지만, 비슷한 평가 샘플에 대해서도 큰 평가 점수 차이를 보이거나, 검색 도우미의 답변에 서지 정보가 없어 점수를 낮게 부여하는 등 제공된 가이드라인에 무관한 주관적인 평가가 이루어지기도 하였다. 이는 평가 배경 및 가이드라인에 대한 설명이 미흡했기 때문에 발생한 현상이며, 보다 객관적인 근거에 의해 평가가

이루어진다면 평가의 신뢰도가 상승할 것으로 예상된다. 평가 시스템 설계자가 평가자에게 자세한 내용을 설명하고 질의응답하는 과정을 거치며 평가 시스템에 대한 이해도를 높이는 과정이 필요할 것으로 생각된다.

- 본문 데이터 활용의 한계가 존재한다. 파일 용량 문제와 문서 파싱 기술의 이슈 등으로 인해 본 연구에서는 본문 데이터를 제한적인 수의 도서에만 적용하였다. 또한, 본문 데이터를 검색에 활용하는 방법론이 상대적으로 단순하며, 충분히 고도화되지 않았다. 더욱 효과적인 검색 성능을 달성하기 위해서는 도서 본문의 긴 텍스트를 효과적으로 요약하는 Book Summarization 기술과 같은 최신 방법론을 적용하거나, 본문 데이터를 검색 과정에 보다 정교하게 활용하는 방안에 대한 추가 연구가 필요하다.
- 이러한 한계점들은 향후 연구를 통해 충분히 보완될 여지가 있으며, 본 연구의 결과를 기반으로 더욱 발전된 형태의 도서 검색 도우미 개발이 기대된다.

# 5. 추후 연구 및 필요작업

본 연구를 통해 RAG 기반 지능형 검색 시스템의 구현 가능성과 데이터 조합의 실효성을 확인하였으나, 연구의 범위와 자원적 제약으로 인해 다음과 같은 한계점이 존재하였다. 이에 따라 향후 연구에서는 이들 한계점을 극복하고 시스템을 고도화하기 위한 작업이 지속적으로 요구된다.

## 5.1 합성데이터 검증 체계 도입 및 실데이터 확보

- 개선방안: 합성데이터(초록, 리뷰)의 정확성을 평가하기 위해 Human-in-the-Loop 기반 평가 Framework를 도입한다. 평가자는 실제 도서 내용을 읽고 가상 초록/리뷰와의 일치도, 정보 왜곡 여부, 표현의 적절성을 평가한다. 평가 내용을 기반으로 합성 방법론을 개선함으로써 합성데이터에 대한 결과 개선이 이루어질 것이다.
- 예시: “가상의 리뷰 A는 실제 독자 리뷰와 유사한 감성 어휘와 논점을 포함하고 있는가?”와 같은 평가 기준을 제시하고 정량화할 수 있다.
- 추가적 접근: 출판사 메타데이터, 서점 플랫폼(예: 알라딘, 예스24)의 사용자 리뷰를 활용하여 실제 데이터를 수집하고, 합성 데이터와 비교하여 정합도를 분석할 수 있다.

### 5.2 사용자 기반 질의 시나리오 확장

- 개선방안: 국립중앙도서관의 실제 검색 로그 데이터를 수집·분석하여, 실사용자가 자주 입력하는 질의 유형과 표현 방식을 정제한다. 또한, 설문조사나 사용자 인터뷰를 통해 인지적·정서적 검색 니즈를 반영한 추가 시나리오를 도출할 수 있다.
- 예시: 기존 시나리오 외에 “연휴에 가볍게 읽을 수 있는 여행 에세이 추천”, “노년층 대상의 건강 정보 책” 등 시간·장소·상황 기반 질의를 반영한 시나리오가 고려될 수 있다.

### 5.3 본문 활용을 위한 고도화된 텍스트 처리 기술 적용

- 개선방안: 본문 데이터를 보다 효과적으로 활용하기 위해 Book Summarization 기술이나 Long-context LLM을 활용하여, 긴 텍스트에서도 의미 단위 요약이나 문맥적 핵심 파악이 가능하도록 한다. 또한, PDF 텍스트 추출 시 오류를 자동 감지하고 보완할 수 있는 AI 기반 문서 정제 파이프라인을 구축하거나, OCR 기반 텍스트 추출 품질 개선 알고리즘을 병행적으로 적용한다.
- 예시: 300페이지 분량의 도서에서 주요 주제별 핵심 단락을 자동 추출한 후, 이를 질의 맥락에 맞추어 동적으로 응답에 반영하는 구조를 설계할 수 있다.

향후 연구에서는 데이터의 신뢰도 확보, 시나리오의 현실성 제고, 본문 데이터의 정교한 활용이라는 세 가지 축을 중심으로 검색 도우미의 지능화 수준을 단계적으로 향상시킬 수 있을 것으로 기대된다. 이러한 고도화 작업은 단순한 기술 개선을 넘어, 공공도서관 정보서비스의 질적 전환과 개인화·맥락화 기반 검색환경 구축의 기반을 마련하는 데 기여할 것이다.

## 6. 제언

### 6.1 서비스 적용 및 확장 방안

본 연구를 통해 검증된 RAG 기반 지능형 검색 도우미는 공공도서관, 전문 아카이브, 학술정보 플랫폼 등 다양한 정보서비스 환경에 적용될 수 있다. 특히 기존의 키워드 기반 검색 시스템이 다룰 수 없는 문맥 기반 탐색 문제를 해결할 수 있는 대안 기술로서 실용적 가치가 크다.

이를 바탕으로 다음과 같은 단계별 적용 및 확장 로드맵을 제안한다.

### 6.1.1 지능형 검색서비스 대상 선정

지능형 검색서비스를 실제 시스템에 적용하기 위해서는 초기부터 전체 소장자료를 대상으로 구현하는 것은 현실적으로 어렵다. 따라서 다음 기준에 따라 1차 적용 대상을 전략적으로 선정할 필요가 있다.

- 카테고리 기반 우선순위 선정
  - 적용 우선 대상은 단순히 데이터 보유여부는 물론, 사용자 열람횟수와 문의 빈도가 높은 카테고리를 중심으로 선정해야 한다. 예컨대, 사회과학, 역사, 문학 분야는 실질적 '사서에게 물어보세요'에서 수요가 높은 카테고리로 우선 검토될 수 있다.
- 시범 적용 대상: 목차 데이터를 포함한 도서
  - 본 연구 결과, '제목+키워드+목차' 조합은 단순 키워드 조합보다 응답 품질이 크게 향상되는 경향을 보였다.
  - 특히 시나리오 2(다중 주제 기반 탐색)에서 목차의 포함 여부가 검색 정합성에 결정적인 영향을 미치는 것으로 나타났다.
  - 목차 데이터는 본문에 비해 구조화되어 있고, 정제 비용이 낮으며, 초기 시스템 설계와 품질 검증에 적합한 형태를 띠고 있다.
  - 따라서 시범 적용 대상은 목차 데이터를 포함한 도서군으로 선정하고, 이를 기반으로 검색 시스템을 우선 구축할 것을 제안한다.

### 6.1.2 데이터 전처리 및 벡터DB 구축 전략

지능형 검색서비스의 품질은 데이터의 정제 및 표현 수준에 따라 결정된다. 이에 따라 다음과 같은 고도화 전략을 적용한다.

- 목차 벡터DB 구축
  - 제목, 키워드, 목차를 기반으로 벡터화를 수행한다.
  - 목차 데이터는 특수기호 제거, 구문 정규화 등의 정제 과정을 거쳐 문단 또는 소단락 단위로 분절되며, 의미 단위 텍스트로 재구성된다. 이후 임베딩 과정을 통해 벡터DB에 저장되어 문맥기반 검색이 가능하도록 처리된다.
  - 목차는 사용자가 원하는 정보 구조(주제 흐름, 장별 구분 등)와 가장 밀접하게 연결되므로, 의미 기반 유사도 검색에 중요한 역할을 한다.

- 문서 단위 또는 섹션 단위 벡터 생성이 가능하며, 구조화된 목차는 사용자 질의와 높은 문맥 일치를 제공한다.

- 초록 및 리뷰의 고도화된 합성 프로세스 적용

  - 구축한 목차 기반 벡터DB를 바탕으로, 해당 정보를 활용하여 도서 초록 및 리뷰 데이터를 생성하고, 이를 정제·검증하여 검색 성능을 보강하는 실험적 확장 단계를 진행한다.
  - 초록은 문서의 개요 및 전개 흐름을 파악하는 데 효과적이며, 사용자 리뷰는 사용자의 감성·상황 맥락을 반영한 개인화 추천 응답 생성에 중요한 역할을 한다.
  - 목차를 기반으로 LLM을 활용한 자동 초록, 가상 리뷰 생성 기법을 적용한다.
  - 가상으로 생성된 초록 및 리뷰의 정확성, 사실성, 표현 적절성을 평가하기 위해 다음과 같은 2단계 검증 체계를 도입한다.
    - Human-in-the-Loop 평가: 실제 도서를 읽은 평가자가 가상 초록과 리뷰의 실제 흐름의 정합성을 평가
    - 실제 리뷰·출판사 요약 데이터 비교 검증: 서점 플랫폼 리뷰, 출판사 초록 등과의 비교를 통해 신뢰도 점수화

- 본문 기반 벡터화 및 의미 검색 고도화

  - 본문 전체 또는 요약문을 대상으로 벡터화를 수행하여 긴 문서에서 핵심 정보만을 추출·검색할 수 있도록 한다.
  - 특히 Book Summarization 기술을 활용하여 장·절 수준의 핵심 문단을 자동 추출하고, LLM 기반 임베딩을 적용한다.
  - 이는 감성 기반 추천(시나리오 3)이나 사용자 맞춤형 응답을 생성하는 데 필수적인 데이터 자원으로 활용된다.

### 6.1.3 단계별 서비스 적용 방안

| 1단계 | 2단계 | 3단계 |
|---|---|---|
| ❶ 목차 기반 검색 시범 서비스 개발 | ❷ 본문기반 목차 생성 및 대상 범위 확대 | ❸ 통합검색 시스템 구축 및 확산 |
| ·목차 기반 벡터 DB 구축<br>·RAG 소장자료 검색 시범 서비스 | ·본문자료 기반 목차 데이터 생성 구축<br>·소장자료 검색 대상 전체 확대 | · 범용 검색 도우미 플랫폼화<br>· 합성데이터 생성 및 검증 모듈 개발 |

### ▣ 1단계: On-premise 기반 본 서비스 적용

1단계는 '목차 데이터를 포함한 도서'를 중심으로 선정된 카테고리에 대해 On-premise 기반의 지능형 검색 시스템을 구축하고, 이를 통해 실제 적용 가능성과 운영성을 검토하는 단계이다.

- 대상: 목차 데이터를 보유하고 있고, 사용자 검색 수요가 높은 분야의 도서 (예: 사회, 문학, 역사)
- 구축 항목 및 기술 구성
  - 벡터DB 구축
    - 제목+키워드+목차 조합의 의미 기반 임베딩 생성 및 인덱싱
    - FAISS 기반 유사도 인덱스 구성
- 구축 환경
  - 온프레미스 GPU 서버 및 벡터 검색 엔진(FAISS 등) 기반 구축
  - 벡터 검색 엔진(예: FAISS, Weaviate) 및 LLM 연동 검색 파이프라인 구축

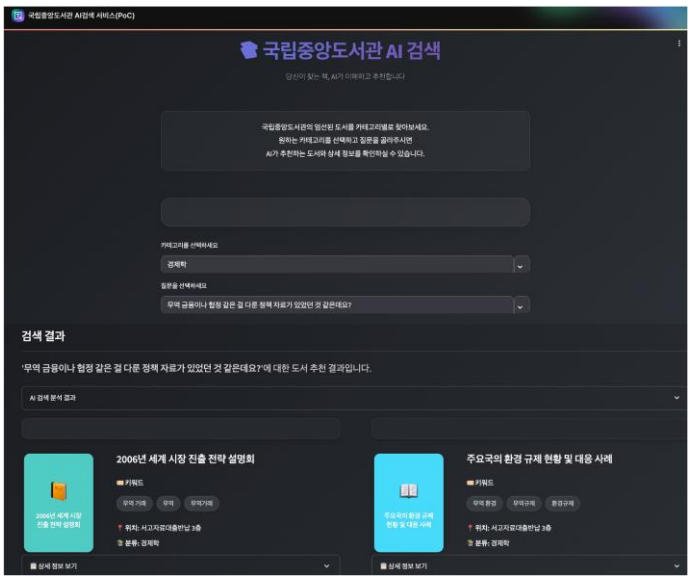



그림 4.1 지능형검색 반영된 사용자화면  예시

### ▣ 2단계: 본문기반 목차 생성 및 전체 소장자료 확대 적용

2단계는 1단계에서 축적된 시스템 구조와 사용자 피드백을 바탕으로, **전체 소장자료에 대한 목차 데이터 확대 적용**을 추진하고 벡터 검색 범위를 확장하는 단계이다.

- 목적 : 목차 데이터가 없는 도서에 대해 본문 기반 자동 목차 생성 → 전체 도서 벡터 검색 적용
- 적용방법
  - 본문 → 문단 분할 → 소제목 생성 → 종합 목차 구성
  - 자동 생성 목차에 대해 Human-in-the-loop 및 외부 검증 루프 적용
  - 생성된 목차를 기존 목차와 동일 방식으로 벡터화하여 통합 검색에 반영
- 통합연동
  - 기존 통합검색 포털과 연동(탭 형식 or 독립 메뉴)
  - '사서에게 물어보세요'와의 시너지 고려

### ▣ 3단계: 국립중앙도서관에서 제공하는 타 시스템 통합 및  확산

3단계는 국립중앙도서관의 RAG 기반 검색시스템을 관련 타 시스템과 통합하고, API 기반의 모듈화 및 플랫폼화를 통해 범용 도입이 가능한 구조로 전환하는 단계이다.

- 대상: 지자체 도서관, 분야별 전문 아카이브 (예: 대한민국 신문 아카이브, 도서관 정보나루 등)
- 플랫폼화 전략
  - 범용 검색 도우미 플랫폼 API화
    - 질의 처리, 검색 호출, 응답 생성 단계를 API로 분리하여 외부 서비스에서도 호출 가능하도록 구성
    - 사용자 기관에 맞는 맞춤형 파라미터 설정 가능(검색 대상 데이터, 응답 스타일 등)
  - 합성데이터 생성 및 검증 Framework 오픈소스화
    - 초록/리뷰/목차 합성 기술을 오픈 모듈로 제공하여 각 기관이 자체 콘텐츠에 적용 가능
    - 검증 루프(Human-in-the-loop 평가) 도구도 함께 배포하여 데이터 품질 유지

## 6.2 시스템 구축 예산 시나리오

- 데이터 건수 : 약 100만 건, 130TB 규모
  - 대상 : 목차와 키워드가 포함된 오프라인 도서 100만건
  - 용량 : 130TB 규모 (참고 : 6,300건에 대한 소장자료 용량 813GB)
- 내용
  - 벡터 임베딩 DB 구축 및 LLM 기반 응답 생성기 도입
  - 검색 UI/UX 고도화 및 인프라 자원 최적화
  - 운영 체계(보안, 백업, 장애 대응) 확립
- 기술 구성 및 요구 사항
  - GPU 서버 (HGX H200급) 2대, CPU / 색인 서버 6대
    - 소장자료에서 추출한 텍스트, 메타데이터, 임베딩을 영구 보존
    - 임베딩 AI
      - 소장 자료 내 텍스트를 숫자로 바꾸는 핵심 AI 엔진
      - 증분 문서에 대한 상시 재임베딩을 위한 자원
    - LLM
      - 검색 결과를 이용한 답변 생성 핵심 AI 엔진
      - 검색 결과에 대한 빠른 활용을 위한 고성능 GPU 서버 필요
    - GPU당 평균 700W 소모되며, 서버 1대당 일반적으로 8개의 GPU가 삽입되어 있음. 따라서 서버 1대당 최대 소비전력은 10kW 정도 요구됨
  - NVMe 스토리지(240TB), NAS 및 백업 장비(LTO 포함)
    - 저장 및 색인 서버의 데이터 유실에 대비한 장기 스토리지
    - 사고, 오염, 랜섬웨어에 대비하기 위한 예비 저장소
  - 고성능 네트워크 장비, 25kW 전력 및 냉각 설비
    - GPU와 데이터베이서스 간 빠른 통신으로 검색과 답변 생성의 데이터 통신 지연 시간 단축
    - 성능 저하 방지를 위한 냉각 시스템 요구됨
    - 전력 구성 요소
      - GPU 서버: 25kW
      - 시설 전체 에너지: 12.5kW
        - 전력효율지수(PUE) 1.5 기준

○ 냉각, 전원 변환, 보조 설비, 조명, 예비 포함

- 운영 인력 2인(인프라 + AI 서비스 관리)
- 연간 SW 라이선스, SLA, 보안 연동 등 서비스화 지원
  - RBAC, 암호화, 24x7 지원이 필요한 경우
    - 벡터 데이터 베이스, GPU 모니터링 등을 위한 관리 비용 지출

- 예상 예산

  - 초기 자본 지출: 약 9.2억 원
  - 구축 및 개발 인건비 관련 지출: 약 5.1억
  - 연간 운영비용: 약 3.45억 원
  - 데이터 정제/합성/생성 비용 (100만건 기준)

[자본 지출]

| 구분 | 사양 | 수량 | 예산 |
|---|---|---|---|
| GPU 서버 | HGX H200 | 2대 | 800,000,000원 |
| CPU/색인 서버 | 6대 | - | 126,000,000원 |
| NVMe 스토리지 | 240TB | 1식 | 44,000,000원 |
| 백업/NAS | 400TB NAS + LTO | - | 40,000,000원 |
| 네트워크·랙 | 스위치 포함 | 1식 | 35,000,000원 |
| 자본 지출 합계 | | | 919,000,000원 |

[구축 및 개발 인건비 관련 예상 지출]

| 구분 | 역할 | 인원 | 노임단가(6개월) |
|---|---|---|---|
| 프로젝트 매니저 | AI 기획 | 1 | 54,872,838원 |
| MLOps 엔지니어 | AI 모델 운영 | 2 | 121,772,940원 |
| AI 모델 엔지니어 | AI 모델 개발 | 2 | 83,321,484원 |
| 데이터 엔지니어 | 데이터 정제, 운영 | 2 | 93,014,196원 |
| 백엔드 엔지니어 | API 서버, 인증/권한, 서비스 연동 | 1 | 36,594,242원 |

| | | | |
|---|---|---|---|
| 프론트 엔지니어 | UI/UX 설계, 인터페이스 구현 | 1 | 40,363,560원 |
| QA | 기능 테스트, 품질 검증 | 1 | 21,423,342원 |
| 보안/인프라 | 계정 권한, 로그 관리, 보안 설정 등 | 1 | 59,142,600원 |
| 인건비 지출 합계 | | | 510,505,202원 |

- 2025년 적용 SW 기술자 평균임금 기준

[연간 예상 운영 비용]

| 구분 | 항목 | 예산(연간) |
|---|---|---|
| 전력비 | 25kW 장비 및 냉각<br>(예비·변동률 30%) | 81,000,000원/년 |
| 인건비 | 2명 (운영·인프라) | 200,000,000원/년 |
| 소프트웨어 지원 | 라이선스, SLA 등 | 60,000,000원/년 |
| 운영 지출 합계 | | 345,000,000원/년 |

[예상 데이터 정제 비용]

| 구분 | 항목 | 예산(연간) |
|---|---|---|
| 전력비 | 25kW 장비 및 냉각 (예비·변동률 30%) | 81,000,000원/년 |
| 인건비 | 2명 (운영·인프라) | 200,000,000원/년 |
| 소프트웨어 지원 | 라이선스, SLA 등 | 60,000,000원/년 |
| 운영 지출 합계 | | 345,000,000원/년 |

[데이터 정제/합성/생성 비용]

| 구분 | 항목 | 수량 | 예산 |
|---|---|---|---|
| 임베딩 | 제목 | 1,000,000 | 420원 |

| | | | |
|---|---|---|---|
| (OpenAI text-embedding-3-large) | 키워드 | 1,000,000 | 966원 |
| | 목차 | 1,000,000 | 1,106원 |
| | 초록 | 1,000,000 | 82,600원 |
| 합성 (OpenAI GPT-4.1) | 초록 | 1,000,000 | 9,436,000원 |
| | 리뷰 | 1,000,000 | 15,967,000원 |
| 생성 (OpenAI GPT-4.1) | 추천 문구 | 1,000,000 | 8,698,200원 |

- 데이터 정제 비용 중 각 문서 내 정보를 컴퓨터에서 가공 가능한 형태로 변환하는 추출비용은 도서/소장 자료 종류에 따라 높은 불확실성으로 예산에서 제외함.

[예상 WBS]

| 단계 | 기간(주) | 주요 내용 |
|---|---|---|
| 1. 요구사항 정의 / 인프라 셋업 | 2~4주 | 내부 문서 유형 분석, 문서 흐름 설계, 서버·스토리지 초기 설정 |
| 2. 파이프라인 구축 | 4~6주 | 문서 수집/파싱/색인 구축, 임베딩 및 검색 구성 |
| 3. 모델 통합 및 튜닝 | 4~6주 | LLM 선택 (오픈소스/상용), 프롬프트 엔지니어링, 성능 검증 |
| 4. API/UX/보안 개발 | 3~5주 | 사용자 입력창, 인증 및 접근 제어, 로그 모니터링 포함 |
| 5. 테스트 및 고도화 | 3~4주 | 실사용 테스트, 개선, 성능 보완, 이슈 대응 |
| 6. 교육 및 운영 이관 | 1~2주 | 운영 매뉴얼 작성, 담당자 교육, 모니터링 체계 정비 |

# [ 용어정의 ]

| 용어 | 풀네임 | 정의 |
|---|---|---|
| LLM-based Agent | Large Language Model-based Agent | 대규모 언어모델(LLM)을 기반으로 동작하는 지능형 소프트웨어 에이전트로, 사용자 질의 이해, 외부 데이터 검색, 정보 추론 및 답변 생성을 자율적으로 수행함 |
| LLM as Judge | Large Language Model as Judge | LLM을 평가자로 활용하는 방식으로, 검색 결과나 생성된 응답의 품질을 자동으로 평가하며 Human 평가를 보완함 |
| In-context Learning (ICL) | In-context Learning | 모델 재학습 없이 프롬프트에 예시나 지침을 제공하여 원하는 응답을 유도하는 학습 방식으로, Few-shot Prompting의 핵심 원리 |
| Dense Vector | Dense Vector Representation | 임베딩 모델을 통해 단어·문장·문서를 의미 기반으로 수치화한 고차원 벡터 표현으로, 의미적 유사도 계산에 활용됨 |
| BM25 | Best Match 25 | 정보검색에서 문서와 질의 간 관련성을 평가하는 랭킹 함수로, TF·IDF 및 문서 길이 정규화를 반영하여 키워드 검색 정확도를 높임 |
| Convex Combination | Convex Combination | 여러 점수나 벡터를 가중합하되, 가중치 합이 1이고 각 가중치가 0 이상일 때의 결합 방식으로, 서로 다른 검색 점수를 최적 비율로 통합할 때 사용됨 |
| Elo Rating | Elo Rating System | 경쟁 게임에서 플레이어 실력을 수치화하는 시스템으로, 검색·생성모델 평가 시 응답 간 승률을 기반으로 성능 비교와 순위화를 수행함 |

# [ 참고문헌 ]


[1] Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in neural information processing systems* 33 (2020): 9459-9474.

[2] Borgeaud, Sebastian, et al. "Improving language models by retrieving from trillions of tokens." *International conference on machine learning*. PMLR, 2022.

[3] Izacard, Gautier, et al. "Atlas: Few-shot learning with retrieval augmented language models." Journal of Machine Learning Research 24.251 (2023): 1-43.

[4] Stephen Robertson and Hugo Zaragoza. “The probabilistic relevance framework: BM25 and beyond.” *Foundations and Trends in Information Retrieval*. 3(4):333-389, 2009.

[5] Karpukhin, Vladimir, et al. "Dense Passage Retrieval for Open-Domain Question Answering." *EMNLP (1).* 2020.

[6] Borgeaud, Sebastian, et al. "Improving language models by retrieving from trillions of tokens." *International conference on machine learning*. PMLR, 2022.

[7] Bruch, Sebastian, Siyu Gai, and Amir Ingber. "An analysis of fusion functions for hybrid retrieval." *ACM Transactions on Information Systems* 42.1 (2023): 1-35.

[8] Wang, Zihao, et al. "Describe, explain, plan and select: interactive planning with large language models enables open-world multi-task agents." *Proceedings of the 37th International Conference on Neural Information Processing Systems*. 2023.

[9] Hao, Shibo, et al. "Toolkengpt: Augmenting frozen language models with massive tools via tool embeddings." *Advances in neural information processing systems* 36 (2023): 45870-45894.

[10] Shen, Yongliang, et al. "Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face." *Advances in Neural Information Processing Systems* 36 (2023): 38154-38180.

[11] Zheng, Lianmin, et al. "Judging llm-as-a-judge with mt-bench and chatbot arena." *Advances in neural information processing systems* 36 (2023): 46595-46623.

[12] Chan, Chi-Min, et al. "ChatEval: Towards Better LLM-based Evaluators through Multi-Agent Debate." *The Twelfth International Conference on Learning Representations*.

[13] Zhang Q, Wang Y, Jiang Y, Li L, Wu C, Wang Y, Jiang X, Shang L, Tang R, Lyu F, Ma C. et al. “Crowd Comparative Reasoning: Unlocking Comprehensive Evaluations For LLM-as-a-Judge”. July 2025.

[14] Chiang, Cheng-Han, and Hung-Yi Lee. "Can Large Language Models Be an Alternative to Human Evaluations?." *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2023.

[15] OpenAI, “Introducing GPT-4.1 in the API.” *OpenAI.* 14 April 2025, https://openai.com/index/gpt-4-1/

[16] Park et al. “An Empirical Study of Tokenization Strategies for Various Korean NLP Tasks.” *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. 2020.

[17] OpenAI, “New embedding models and API updates.” *OpenAI*. 25 January 2024, https://openai.com/index/new-embedding-models-and-api-updates/

[18] Douze, Matthijs, et al. "The faiss library." *arXiv preprint arXiv:2401.08281* (2024).

[19] Valentin Crettaz, “Elasticsearch hybrid search.” *Elastic Search*. 17 February 2025, https://www.elastic.co/search-labs/blog/hybrid-search-elasticsearch

[20] McHugh, Mary L. “Interrater reliability: the kappa statistic.” *Biochemia medica* vol. 22,3 (2012): 276-82.

# [ 별첨자료 - 시나리오 질의셋 ]

## 1. LLM Judge 평가

### 시나리오1

1. 무역 금융이나 협정 같은 걸 다룬 정책 자료가 있었던 것 같은데요?
2. 창업과 관련된 소상공인이나 제도 안내 책 찾습니다
3. 소득 분배의 불균형과 산업 구조적 원인에 대해 설명한 책이 있을까요?
4. 해외시장 진출 관련해서 수요 예측 자료 본 적 있어요
5. 소득계층별 조세 부담에 관한 구체적 연구 자료가 있을까요?
6. 중학교 교육과정에서 학습 설계나 평가 관련 자료가 있을까요?
7. 실업계 고등학교나 기술 관련 교육자료 찾고 있어요
8. 대학 전공 심화과정 관련 시험이나 평가 안내 책 있나요?
9. 학습 지도안이 잘 정리된 실무 중심 교재 없을까요?
10. 풍자나 사회 비판을 담은 독특한 문체의 책을 찾고 있습니다
11. 영문편지의 실제 예문이 담긴 책 없을까요?
12. 종교적 상징이 담긴 낭만적 서사시를 찾고 있습니다
13. 시대별 문화 흐름을 정리한 역사서가 필요해요
14. 유명한 목회자나 정치인의 생애를 다룬 전기성 기록 있나요?
15. 해외 운하 탐방기나 견문록 형태의 기록을 찾고 있어요
16. 한국의 현대미술 흐름을 정리한 작품집이 있던 것 같은데요
17. 명필의 글씨체나 서예 형식이 수록된 고문서 찾고 있어요
18. 조선시대 관료 복식에 쓰인 장식 문양에 대한 책이 있던데요
19. 일상에서 실천 가능한 생활체육 방법을 소개한 책이었어요
20. 노인 대상 건강보험 재정관리와 관련된 정책 보고서를 찾습니다
21. 고요한 국악기 '아쟁'과 관련된 전통 악보집 있나요?
22. 다양한 부위에 적용 가능한 전문 마사지 기법을 소개한 책이 있나요?
23. 국내 광산 정보를 정리한 오래된 책이 있었던 것 같아요
24. 도시 지역의 강우 흐름 처리 방안을 다룬 기술서가 궁금해요
25. 도시계획이나 건강한 도시 조성을 다룬 보건정책 자료 없을까요?
26. 새로운 항암물질에 대한 화학적 접근을 다룬 기술 개발서가 필요합니다
27. 주역과 관련된 해석이나 철학적 해설이 담긴 고전이 있었던 것 같아요
28. 논어를 독학하거나 쉽게 읽을 수 있는 해설 중심의 책이 필요해요
29. 타로를 통해 인간 이해나 심리를 철학적으로 풀어낸 책이 있나요?

30. 노자의 사상을 현대적으로 풀어낸 책이 있던데요

## 시나리오2

1. 산업 발전 관련 사례나 중장기 계획 다룬 책이 있을까요?
2. 중소기업 현황이나 정책 방향 관련 자료가 필요해요
3. 경제 정책이나 수입 정책 등 정책 성격이 강한 도서가 있을까요?
4. 외국인 투자 관련 노무, 제도 등 실무 정보를 담은 책 추천해주세요
5. 건강·복지 정책이 고령층 삶에 미치는 영향을 설명한 책이 궁금해요
6. 대학 시설 정책이나 행정 연구 중심 자료가 필요해요
7. 초등 수학 지도에서 교사용 보조 자료 중심 책 찾고 있어요
8. 가정 교육과 연계 가능한 효 교육 교재가 궁금해요
9. 청소년 대상 민주시민 교육 관련 자료 추천해주세요
10. 사이버윤리나 도덕 관련 교과서 기반 도서가 있을까요?
11. 한국 문학 중 사랑을 주제로 한 대표적인 소설이 궁금합니다
12. 작가 간 주고받은 편지 형식의 책이 있을까요?
13. 현대 시집 중 감성을 자극하는 작품 추천해주세요
14. 유머나 위트가 있는 문학 작품이 있을까요?
15. 궁중 이야기나 역사적 배경을 담은 문학 작품을 읽고 싶어요
16. 조선 시대를 전체적으로 서술한 역사 통사가 있나요?
17. 지역 독립운동과 관련 장소나 인물 중심 자료 있나요?
18. 서울 행정기관과 기관별 조직을 정리한 책이 궁금해요
19. 북한의 경제뿐 아니라 문화·사회적 맥락까지 다룬 책이 있을까요?
20. 서울을 배경으로 한 관광과 사진이 함께 담긴 작품집 있을까요?
21. 현대 건축물과 수상작 중심으로 정리된 책 있나요?
22. 대한민국의 전통 미술 및 전람회 중심으로 구성된 책이 궁금해요
23. 농산물 시장 개방에 따른 지역별 정책 대응을 분석한 자료가 있을까요?
24. 농산물 유통과 서비스 산업 개방을 함께 다룬 정책서 찾고 있어요
25. 중국 농림수산물 무역 구조와 생산 동향을 다룬 보고서 있나요?
26. 에너지 신기술 워크숍이나 기술 동향 정리한 보고서 없을까요?
27. 항균 성분 또는 식품 관련 기준 개정을 다룬 연구 보고서가 필요해요
28. 환경서비스와 국제 시장 개방이 관련된 정책 연구 자료 추천해주세요
29. 도덕과 윤리를 중심으로 한 철학 입문서나 교육 자료가 있을까요?
30. 가정윤리와 상담에 대한 내용을 담은 책을 찾고 있어요

# 시나리오3

1. 실패를 경험한 사람들을 위한 책 추천해줘
2. 첫 사회생활을 시작한 사람들을 위한 책 추천해줘
3. 청년 창업가를 위한 기업 분석 관련 책 있나요?
4. 회사 퇴직을 준비하는 사람들을 위한 책 추천해줘
5. 고령층이 읽기에 적당한 경제 흐름 설명 자료가 있을까요?
6. 죽음이 관련된 책 추천해줘
7. 결혼을 앞둔 부부를 위한 책 추천해줘
8. 고등학생이 이해할 수 있는 실업계 교육과정 중심 책 추천해주세요
9. 10대가 읽으면 좋을 책 추천해줘
10. 긴 연휴에 읽기 좋은 책 추천해줘
11. 전원이나 사찰 같은 조용한 공간의 시적 분위기를 담은 책을 찾고 있어요
12. 수험 생활 중 읽으면 위로가 되는 자전적 이야기가 필요해요
13. 어머니에 대한 그리움이 담긴 시를 추천해주세요
14. 해외 체류나 이국적 감성을 담은 시집을 찾고 있어요
15. 새해에 읽기 좋은 책 추천해줘
16. 서사적이고 고전적인 분위기의 영시 번역 작품을 추천해주세요
17. 인간관계 관련된 책 추천해줘
18. 일상 속 소비 습관을 기록하며 절약하는 방법을 다룬 책 없을까요?
19. 예비 부모를 위한 책 추천해줘
20. 중장년층이 즐기기 좋은 여가활동 관련 책 추천해줘요
21. 아이들과 함께 문화 공간을 체험하며 읽기 좋은 책이 있을까요?
22. 우울할때 읽으면 좋은 책 추천해줘
23. 감성적이고 표현의 폭이 넓은 시집추천해줘
24. 추운 겨울 마음이 따뜻하고 훈훈해지는 책
25. 쉽게 읽히는 책 추천해줘
26. 따뜻한 차와 함께 생각에 잠길 수 있는 철학적 성찰 책을 찾고 있어요
27. 자존감 높여줄 수 있는 책 추천해주세요
28. 20대가 읽으면 좋을 책 추천해줘
29. 초등학교2학년 자녀가 있는데 권장도서 좀 부탁합니다.
30. 50대 초반 여성이 공감하며 읽을 수 있는 책

# 2. 본문 활용 가능성 탐색

## 시나리오1

1. 여가와 관련된 사회적 변화나 여행 문화에 관한 책 있나요?
2. 워드프로세싱 자격 시험과 관련된 자료집 찾고 있어요
3. 부모 교육을 위한 실제적 가이드를 제공하는 책이 필요해요
4. 한국 문학사의 흐름과 작가별 정리를 다룬 책 있어요?
5. 임금교섭 관련 실무 지침이 담긴 도서 없을까요?
6. 한국 노사관계에 대한 이해를 돕는 책 추천해 주세요
7. 노년의 삶과 복지 문제를 다룬 책 추천해 주세요.
8. 영어 면접 준비와 인터뷰 전략을 함께 설명한 책이 필요해요
9. 성희롱 예방 교육을 위한 실제 사례 중심 자료 있나요?
10. 컴퓨터 사용이 청소년 건강에 미치는 영향을 분석한 연구 있어요?

## 시나리오2

1. 언론 보도가 재판에 미치는 심리적 영향에 대해 분석한 책은요?
2. “삶의 질과 관련된 국가 정책 변화를 다룬 책이 있나요?”
3. “환경 정책과 제도 개선 방향에 관한 보고서형 자료가 필요해요”
4. 공연예술 정책 및 아시아 지원기관 현황 자료를 찾습니다
5. 지역 관광 개발과 환경 갈등이 함께 다뤄진 책 있나요?
6. “조세 개혁이 복지 기반에 어떤 영향을 주는지 다룬 자료가 필요해요”
7. 국제 자매결연과 지역 교류 사례를 소개한 보고서가 있나요?
8. 원조교제 문제와 청소년 심리 간의 관계를 다룬 책 있나요?
9. 여성 삶의 변화상을 다룬 정부 정책 관련 자료가 궁금해요
10. 청소년의 심리와 법률적 의사결정 능력을 다룬 책이 궁금해요

## 시나리오3

1. 회사 퇴직을 준비하는 사람들을 위한 책 추천해줘
2. 죽음이 관련된 책 추천해줘
3. 어머니에 대한 그리움이 담긴 시를 추천해주세요
4. 예비 부모를 위한 책 추천해줘
5. 우울할때 읽으면 좋은 책 추천해줘
6. 추운 겨울 마음이 따뜻하고 훈훈해지는 책

7. 따뜻한 차와 함께 생각에 잠길 수 있는 철학적 성찰 책을 찾고 있어요
8. 20대가 읽으면 좋을 책 추천해줘
9. 초등학교2학년 자녀가 있는데 권장도서 좀 부탁합니다.
10. 50대 초반 여성이 공감하며 읽을 수 있는 책

2025년 연구보고서
국립중앙도서관 지능형 검색 서비스 적용방안 연구

인 쇄 2025년 9월 30일
발 행 2025년 9월 30일
발 행 처 국립중앙도서관
서울특별시 서초구 반포대로 201(반포동)
Tel. 02-590-0500
https://www.nl.go.kr
연구기관 주식회사 렛서
I S B N 979-11-6513-500-3(비매품/무료)

