



RAG를 활용한 의회 회의록 기반 기사문 생성 시스템 개발



임회정 전남대학교 공공데이터분석센터(교수), 광주광역시 61186, 대한민국

황법량 전남대학교 데이터사이언스 대학원(석사과정), 광주광역시 61186, 대한민국

ID: KS031-A

서론

의회 회의록은 의원과 집행부 공무원 간 질의응답으로 구성되며, 공공서비스에 대한 다양한 의견과 문제제기가 담긴 중요한 자원이다. 그러나 대부분의 시민은 물론 언론인들조차 이를 제대로 읽고 해석할 시간이 부족하다.

본 연구는 이러한 한계를 해결하기 위해, 많은 사람들이 관심 가질 만한 관점에 입각해 회의록을 기사문 형태로 요약·재구성하는 RAG 시스템을 구현하려 한다.

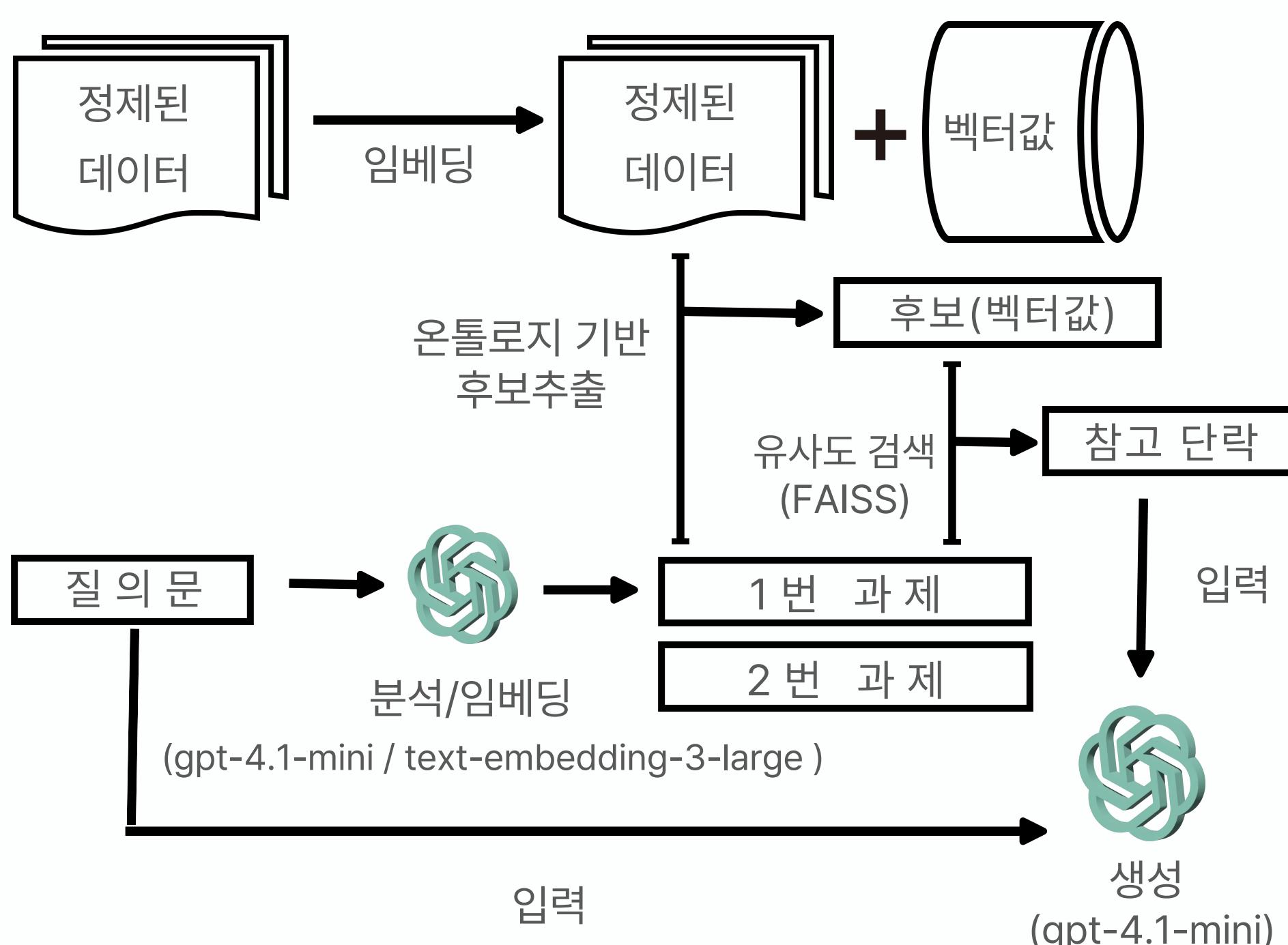
데이터 전처리



의회 회의록은 사회자의 의사진행 발언을 기준으로 발언순서별 분할이 가능하다. 본 연구에서는 이렇게 나누어진 단락을 검색의 최소 기준으로 삼는다. 이어서 분할된 단락에서 LLM을 이용해 해당 단락의 발언자, 주제 등의 정보를 추출한다.

이러한 추출 과정에서 API를 이용해 LLM을 호출할 경우 비용이 발생하는데, 본 연구에서는 이 비용을 절감하기 위해 오픈소스 LLM gpt-oss-20b를 사용했다.

RAG 구조



입력된 질의문은 1차적으로 LLM에서 유사도 검색에 적합한 형태의 과제로 분석 및 변환된다. 과제문별로 유사도 검색을 통해 최대 5개의 참고단락을 추출한다. 이렇게 추출된 단락과 질의문을 LLM에 입력해 기사문을 생성한다. LLM은 API로 gpt-4.1-mini를 호출해 사용했다. 텍스트 임베딩에는 text-embedding-3-large를 사용했다.

평가

평가 대상

본 연구에서는 2022년 7월부터 2025년 10월까지의 기간 동안의 광주광역시의회와 서울특별시의회 회의록을 대상으로, 100개의 예시 질의문에 대한 응답 과정에서 추출된 참고 단락의 품질을 평가하였다. 아무런 정제작업을 하지 않고 회의록을 임베딩하여 벡터값을 추출한 다음 임베딩한 질의문과 유사도 검색을 수행하는 Naive RAG와 본 연구에서 구현한 Ontology RAG의 성능을 비교했다.

평가 방법 및 기준

질의문 예시

A의원(OO의회 XX당 소속)의 □□ 주제에 대한 견해는 무엇인가?

평가
(gpt-4.1-mini)

- 발언자, 의회, 정당을 정확하게 찾아냈는가?
- 질의문의 주제와 참고단락의 주제가 유사한가?

참고단락 예시

B의원(OO의회 XX당 소속)이 □□ 주제에 대해 발언한 내용

평가용 질의문 (100개)

- 특정 의회에서 논의된 특정 주제에 대한 내용 (의회+주제): 40개
- 특정 발언자가 특정 주제를 언급한 내용 (인물+주제): 20개
- 특정 정당의 특정 주제에 대한 입장 (정당+주제): 10개
- 두 대상 사이의 입장 차이 (인물, 의회, 정당 + 주제): 30개

결론

평가지표	Naive	Ontology	변동
사실성 오류율	32.2%	7.05%	25.15% 감소
관련성 점수 평균	5.77/10	6.54/10	7.66% 개선

Ontology RAG는 총 15개의 질의문에 대해서 참고할 단락이 없다는 결과를 출력했다. 이는 질의문이 묻고 있는 내용이 존재하지 않는 것인데 Naive RAG는 유사도 검색에만 의존해 완전히 잘못된 단락을 참고 단락으로 추출했음을 의미한다. 해당 단락들을 추가적으로 오류로 판정하고 관련성 점수를 0점으로 부여했다. 평가결과 사실성 오류율이 크게 감소했으며 관련성 점수 평균 또한 일정한 개선이 있었다.



본 연구에서 사용한 코드와 데이터 그리고 평가 결과는 GitHub 공개되어 있다. 각 언어별로 번역한 상세 설명 또한 첨부되어있다.