

Ch_10 Retriever 서기

Retriever의 정의

: 정보를 찾아서 가져오는 기술

문서를 저장하는 기능보다는 어떻게 하면 더 정확하고 관련성 높은 정보를 검색해 올 것인가에 초점을 둬!

핵심 기능 세 가지

- 문서 검색: 유사한 문서 추출
- 검색 결과 확장: 질문 의도분석
- 컨텍스트 제공: 모델이 학습하지 않은 최신 정보 제공

MultiQueryRetriever: 사용자의 불완전함 보완

- 질문을 확장하는 리트리버
- 놓치는 정보를 없애는 역할
- 질문을 여러 관점으로 확장

EnsembleRetriever: 검색의 균형과 신뢰도

→ 상호 보완적인 알고리즘을 결합해 검색 결과의 신뢰성 확보

- 특정 알고리즘의 편향성 때문에 발생하는 검색 실패를 방지
- 하이브리드 검색을 구현

ContextualCompressionRetriever: 고효율, 고품질의 답변

- 사용자의 질문 맥락의 맞춰 내용을 compress함
- llm은 불필요한 정보에 집중하지 않고 필요한 정보에만 집중할 수 있게 함
- 할루시네이션 현상을 해결

리트리버를 하나만 쓰는 게 아니라 여러 개를 엮어 체인처럼 씬!

→ 멀티쿼리가 질문을 늘리고, 인셈블이 문서를 대량으로 뽑고, 많은 문서들을 compress하는 것처럼 하나의 흐름으로 사용!