

Daily 협업일지(02/09)

[1] 오늘 날짜 / 이름 / 팀명

- 날짜: 2026.02.09
- 이름: 김슬기
- 팀명: 6팀

[2] 오늘 맡은 역할 및 구체적인 작업 내용

| 오늘 당신이 맡았던 역할은 무엇이었고, 어떤 작업을 수행했나요?

(예: 모델 학습 파라미터 조정, 결측치 처리, 발표자료 구성 등)

답변:

- 베이스라인 측정: Chunk Size 1000, Vector Search Only로 테스트 시 정확도 61% 기록. 구체적인 금액 검색에 실패함
- Hybrid Search 도입: rank-bm25를 적용하여 키워드 검색(BM25)과 벡터 검색을 5:5로 결합
- Chunk Size 최적화: 1000자에서 500자로 줄여서 실험
- Top-K 확장: 검색 문서 개수(K)를 5개에서 10개로 늘려봄

[3] 오늘 작업 완료도 체크 (하나만 체크)

| 진척 상황을 정량적으로 표시하고, 간단한 근거도 작성하세요.

- 0% (시작 못함)
- 25% (시작은 했지만 진척 없음)
- 50% (진행 중, 절반 이하)
- 75% (거의 완료됨)
- 100% (완료 및 점검까지 완료)

간단한 근거:

(40%) 초기 60%대였던 정확도가 70%대로 진입하며 시스템의 신뢰성이 확보됨

[4] 오늘 협업 중 제안하거나 피드백한 내용이 있다면?

| 오늘 회의나 메시지에서 당신이 제안하거나 팀에 피드백한 내용은 무엇인가요?

답변:

-

[5] 오늘 분석/실험 중 얻은 인사이트나 발견한 문제점은?

| EDA, 모델 실험 중 유의미한 점이나 오류가 있었다면 자유롭게 작성하세요.

답변:

- 벡터의 한계: "493,763,000원" 같은 숫자는 벡터 유사도로 찾기 힘듦. BM25를 섞으니 바로 찾아냄 (정확도 64%로 상승)
- 청크 크기: 청크가 너무 크면(1000자) 끝데없는 서식이 섞여 검색 품질이 떨어짐. 500자로 줄이니 예산

정보만 딱 잡혀서 정확도가 70%까지 오름

- Top-K: 정답이 7~8번째에 있는 경우가 많아 K=10이 유리함

[6] 일정 지연이나 협업 중 어려웠던 점이 있다면?

| 자기 업무 외에도 전체 일정이나 팀 내 협업에서 생긴 문제를 공유해 주세요.

 답변:

- BM25 토크나이저 설정에 따라 검색 결과가 달라져 튜닝이 필요했음

[7] 오늘 발표 준비나 커뮤니케이션에서 기여한 부분은?

| 슬라이드 제작, 발표 연습, 질문 정리 등 발표와 관련된 활동을 썼다면 기록하세요.

 답변:

-

[8] 내일 목표 / 할 일

| 구체적인 개인 업무나 팀 목표 기반 계획을 간단히 적어주세요.

 답변:

- 검색된 문서 중 불필요한 문서(서식, 별지 등)를 걸러내는 필터링 로직 추가
- 최종 파라미터 확정 및 중간 발표 리포트 작성