

RAG 테스트 문서 (가짜 데이터) - Google (Fictional)

Generated: 2025-12-25 06:00:36

목적: RAG(검색+생성) 기능 테스트를 위한 '가짜' Google 관련 문서입니다.

주의: 아래 프로젝트/수치/일정/제품명은 모두 예시이며 실제 Google의 사실/전략/실적이 아닙니다.

1) 가짜 개요(Fictional Overview)

- 회사: Google (테스트용 가상 요약)
- 가짜 프로젝트 코드명: Project SKYSEARCH
- 가짜 목표: 검색 품질(quality)과 응답 지연(latency)의 균형 최적화(예시)
- 가짜 KPI:
 - * 검색 응답 지연 P95: 220ms -> 180ms (가상 목표)
 - * 스팸 결과 노출률: 1.8% -> 1.2% (가상 목표)

2) 가짜 제품/기술 메모(Fictional Notes)

- 가짜 모델명: G-Reasoner-2 (예시)
- 가짜 인덱싱 전략: DeltaIndex + FreshBoost(예시)
- 가짜 리스크: 캐시 무효화 폭증, 질의 드리프트(query drift), 랭커 편향(bias) 증가

3) 가짜 운영 정보(Fictional Operations)

- 테스트용 RAG API Base URL 예시: <http://localhost:11434>
- 테스트용 기본 포트 예시: 11434
- 임베딩 모델 예시: nomic-embed-text
- 생성 모델 예시: gemma3:1b 또는 llama3.2:3b

4) 가짜 결론(Fictional Conclusion)

- 가장 큰 병목은 '캐시 무효화 폭증'으로 가정하며, 캐시 키 설계 개선이 최우선이다(가상).
- 로그 샘플링을 5%에서 8%로 늘리면 디버깅 속도가 개선될 수 있다(가상).

RAG 테스트 문서 (가짜 데이터) - 부록/용어/샘플 질문

Generated: 2025-12-25 06:00:36

A) 미니 용어집(Mini Glossary)

- P95: 자연 시간 분포의 95번째 백분위(예시 설명)
- Ranker: 문서를 정렬(랭킹)하는 모델/로직(예시 설명)
- top_k: 벡터DB에서 검색해오는 청크 개수

B) 샘플 QA (정답이 문서에 '명시'되도록 구성)

Q1) 가짜 프로젝트 코드명은 무엇인가?

A1) Project SKYSEARCH

Q2) 문서에서 가정한 가장 큰 병목은 무엇인가?

A2) 캐시 무효화 폭증

C) 테스트 팁

- 질문은 문서에 있는 문장/키워드를 그대로 묻는 형태가 가장 잘 동작합니다.
- 이미지 기반 PDF는 텍스트 추출이 어려울 수 있으며 OCR이 필요할 수 있습니다.