

AI 데이터 분석가 '물어보새' 등장 1부

RAG와 Text-To-SQL 활용

#AI #Artificial Intelligence #RAG

2025.07.26
이경연

<https://techblog.woowahan.com/18144/>

1 문제 정의

구성원 약 95%가 데이터를 활용해 업무를 수행

구성원 절반 이상이 SQL문 사용에 대해 어려움을 가지고 있음

데이터에 대한 신뢰도가 있는 지에 대한 고민

2 프로젝트 목표 및 핵심 요소

목표

자연어를 통해 신뢰도 높은 데이터를 추출할 수 있게 하여 구성원의 데이터 리터러시 상향 평준화

****데이터 리터러시**

데이터를 통해 유의미한 정보를 추출하고 해석하며 신뢰성을 검증하고, 데이터 탐색과 분석을 통해 통찰을 도출하고 합리적인 의사결정을 내릴 수 있는 소통 능력



체계화

데이터 카탈로그, 데이터
거버넌스 비즈니스 용어 등을
활용한 데이터 용어 표준화



자동화

24/7 무중단 운영



효율화

사내 도메인 이해, 빠른 검색



접근성

Slack 기반 Q&A
인터페이스

3 물어보새 기반 기술

LLM

딥러닝 알고리즘 기반의
대규모 언어 모델

GPT-4o

RAG

모델이 데이터를 직접
학습하지 않고 필요한 사내
데이터 검색으로 LLM 답변
성능 개선

Langchain

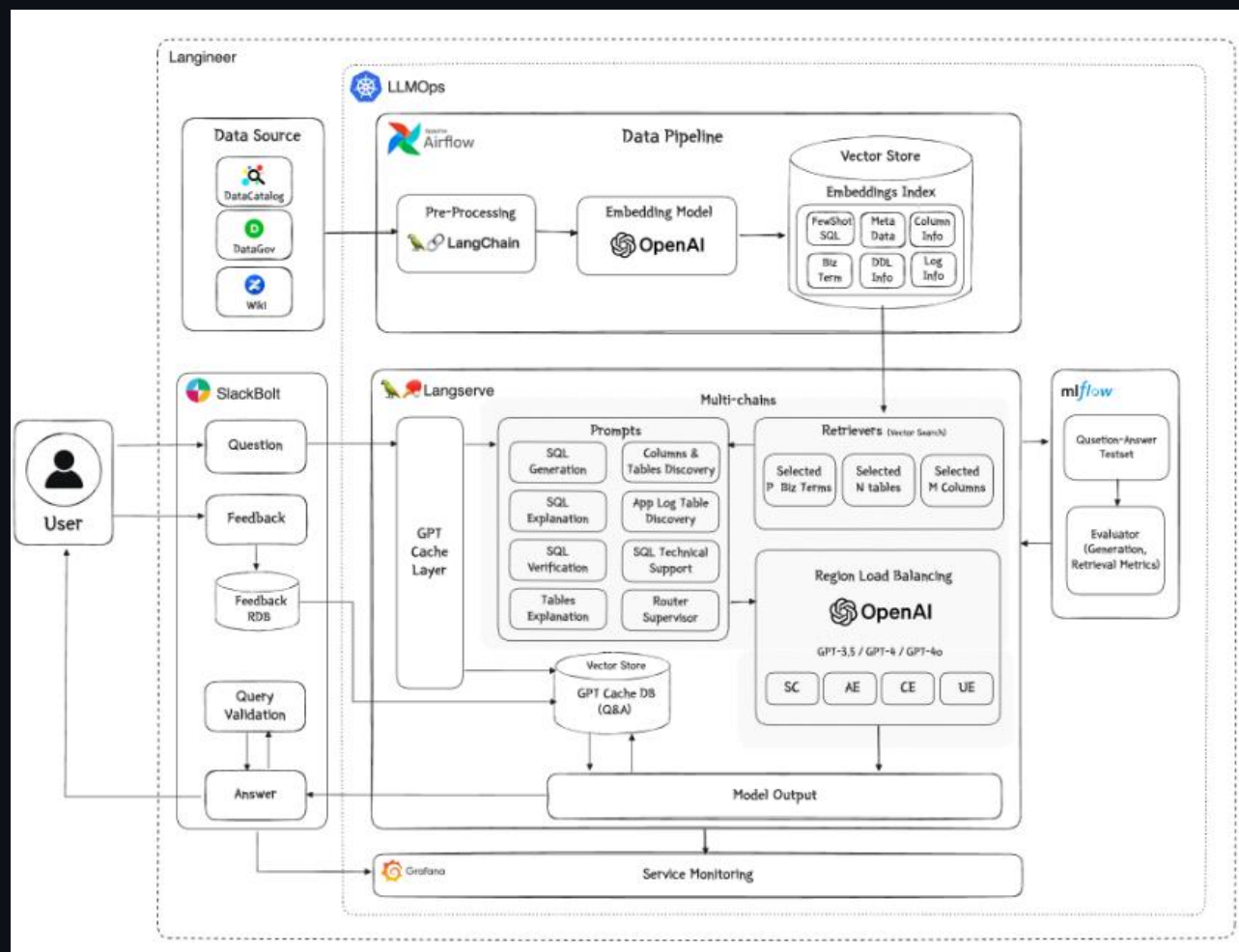
LLM과 애플리케이션 개발을
지원하는 모듈화된 체인의
오픈 소스 프레임워크

LLMOps

LLM 배포, 관리, 모니터링을
위한 운영 기법과 도구로
파운데이션 모델 없이도
엔지니어링 영역에 적용

4 물어보세 아키텍처

- 1 벡터 스토어 기반 비정형 데이터 파이프라인 구축
- 2 다양한 리터러시 기능 제공을 위한 RAG 기반의 멀티 체인 구조 개발
- 3 LLM 서비스 개발, 배포, 운영을 위한 LLMOps 구축
- 4 Slack 기반의 다양한 응답 기능 도입



5 물어보세 Text-to-SQL 구현 전략

데이터 보강

메타데이터 상세화 (테이블 목적, 컬럼 설명, 비즈니스 키워드 등)
비즈니스 용어집 및 few-shot SQL 예시 구축

검색 알고리즘 개발

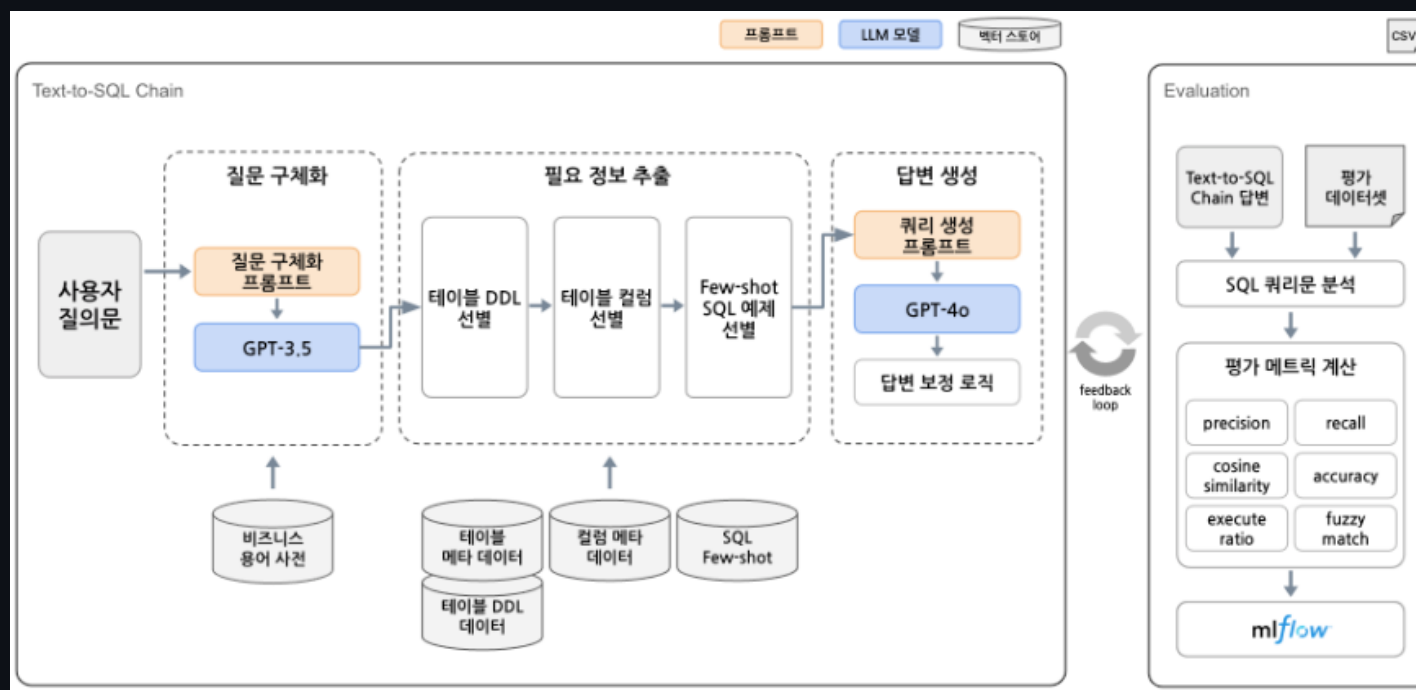
질문 의도 파악 및 메타데이터 기반 유사도 분석으로 정보 선별

프롬프트 엔지니어링

LLM 기반 ReAct 방식의 추론과 검색 과정 적용
“데이터 분석가 페르소나” 지정 → 프롬프트 품질 향상

실험 및 평가시스템 구축

사내 리더보드, 수집 개 지표, A/B 테스트 500회 이상 수행
테스트 → 승인 → 운영 배포로 반복 개선



6 물어보새 프로젝트 결과 및 확장 가능성



성한영(HANYOUNG)/ON/BUDS팀 12:54

2024년 7월 1일 기준 푸드 배달, 1인분, 주류, 면, 커피 메뉴 배달 주문수 그리고 각각 푸드 배달 주문수 대비 비율도 알려줘



물어보새 12:54

답변을 물어오고 있어요. 잠시만 기다려주세요

```
WITH FoodDeliveryOrders AS (
    SELECT
        ord_id,
        ord_date,
        ord_time,
        ord_status,
        menu_id,
        menu_name,
        menu_category,
        menu_is_alcohol,
        menu_is_solo,
        menu_is_noodle,
        menu_is_coffee
    FROM
        food_delivery_orders
    WHERE
        ord_date = '2024-07-01'
        AND ord_status = 'Y'
        AND menu_is_alcohol = 1
        AND menu_is_solo = 1
        AND menu_is_noodle = 1
        AND menu_is_coffee = 1
        AND menu_category IN ('BEER', 'BREAD')
), MenuInfo AS (
    SELECT
        menu_id,
        menu_name,
        menu_category,
        menu_is_alcohol,
        menu_is_solo,
        menu_is_noodle,
        menu_is_coffee
    FROM
        menu_info
    WHERE
        menu_category IN ('BEER', 'BREAD')
), MenuOrderCounts AS (
    SELECT
        ord_id,
        ord_date,
        ord_time,
        ord_status,
        menu_id,
        menu_name,
        menu_category,
        menu_is_alcohol,
        menu_is_solo,
        menu_is_noodle,
        menu_is_coffee
    FROM
        FoodDeliveryOrders
    WHERE
        ord_date = '2024-07-01'
        AND ord_status = 'Y'
        AND menu_is_alcohol = 1
        AND menu_is_solo = 1
        AND menu_is_noodle = 1
        AND menu_is_coffee = 1
        AND menu_category IN ('BEER', 'BREAD')
), OrderInfo AS (
    SELECT
        COUNT(ord_id) AS total_food_delivery_orders,
        SUM(menu_is_alcohol) AS alcohol_orders,
        SUM(menu_is_solo) AS solo_orders,
        SUM(menu_is_noodle) AS noodle_orders,
        SUM(menu_is_coffee) AS coffee_orders
    FROM
        MenuOrderCounts
)
SELECT
    total_food_delivery_orders,
    alcohol_orders,
    solo_orders,
    noodle_orders,
    coffee_orders,
    ROUND(CAST(alcohol_orders AS DOUBLE) / total_food_delivery_orders, 3) AS alcohol_order_ratio,
    ROUND(CAST(solo_orders AS DOUBLE) / total_food_delivery_orders, 3) AS solo_order_ratio,
    ROUND(CAST(noodle_orders AS DOUBLE) / total_food_delivery_orders, 3) AS noodle_order_ratio,
    ROUND(CAST(coffee_orders AS DOUBLE) / total_food_delivery_orders, 3) AS coffee_order_ratio
FROM
    OrderInfo
LIMIT 100
```

이 SQL 쿼리는 2024년 7월 1일 기준으로 각 푸드 배달, 1인분, 주류, 면, 커피 메뉴 아이템의 주문 수와 각 메뉴 아이템 주문 수 대비 '푸드 배달' 주문 수의 비율을 계산합니다. 데이터



쿼리문 생성 및 해설 기능

약 30초에서 1분 이내에 사용자에게 답변 제공

평가 및 한계

첫 입사자 혹은 다른 도메인 업무 담당자는 물어보새로 업무 이해도를 높이는 데 도움됨

일부 구성원들은 비즈니스 로직의 정확성과 질문 이해도의 개선이 필요하다고 의견 제시

확장

점진적으로 테스트 참가자와 대상 조직을 늘려가면서 쿼리문 생성 답변뿐만 아니라 데이터 디스커버리 영역으로까지 기능을 확장

감사합니다.