

LLM 기반 보호대상자 데이터 관리 시스템: SQL Agent 와 자동 보고 기능의 통합

박주연, 박해름, 유재현*
성신여자대학교 AI 융합학부 학부생
성신여자대학교 AI 융합학부 교수

20221365@sungshin.ac.kr, 20231367@sungshin.ac.kr,
*jhyoo@sungshin.ac.kr

LLM-based Protectee Data Management System: Integration of SQL Agent and Automated Reporting

Park Ju Yeon, Park Hae Reum, Yoo Jae Hyun*
School of AI Convergence, Sungshin Women's University

요 약

본 논문은 보호대상자 관리와 같이 데이터베이스 기반 의사결정이 중요한 영역에서, 비전문가도 직관적으로 정보를 활용할 수 있도록 하는 LLM 기반 데이터 관리 시스템을 제안한다. 제안된 시스템은 LangGraph 를 활용한 SQL Agent 를 통해 자연어 질의를 안전한 SQL 로 변환·검증·실행함으로써, 데이터베이스 구조를 이해하지 못하는 사용자도 원하는 정보를 손쉽게 조회할 수 있도록 한다. 더 나아가 프롬프트 엔지니어링 기법을 적용하여 데이터베이스에서 추출된 주요 지표를 요약·해석하고, 이를 일일 리포트 형태로 자동 생성함으로써 보호대상자의 상태 변화를 종합적으로 파악할 수 있게 한다.

I. 서 론

데이터베이스는 의사결정의 핵심 도구이지만, SQL 과 같은 질의 언어는 비전문가에게 큰 장벽이 된다. 특히 보호대상자 관리에서는 심박변이도, 스트레스 지수, 움직임 위험도, PPG 기반 위험도 등 다양한 생체 지표가 포함되어 있어 해석과 활용이 더욱 어렵다.

본 논문은 이러한 한계를 극복하기 위해 LLM 기반 보호대상자 데이터 관리 시스템을 제안한다. 제안된 시스템은 두 가지 기능으로 구성된다. 첫째, LangChain 과 LangGraph 를 활용한 SQL Agent 는 자연어 질의를 안전한 SQL 로 변환·실행하여 비전문가도 데이터를 쉽게 조회할 수 있도록 한다. 둘째, 프롬프트 엔지니어링 기반 리포트 생성기는 주요 지표를 요약·해석해 일일 보고서를 자동 제공함으로써 보호대상자의 상태 변화를 한눈에 파악할 수 있게 한다.

또한 LangGraph 의 상태 기계 구조를 통해 질의 실패 시 복구와 맥락 지속성을 보장하여, 기존 LLM 기반 SQL 생성의 한계를 극복하였다. 결과적으로 본 시스템은 실시간 질의 응답과 자동 보고를 통합하여 보호대상자 관리 업무의 데이터 활용성을 높인다.

II. 본론

2.1 이론적 배경

Text-to-SQL 연구는 자연어를 SQL 로 변환하는 데 초점을 맞추어 발전해왔다. 기존의 연구는 공개 벤치마크에서 우수한 성능을 보였으나, 실제 도메인 데이터베이스의 복잡성과 안전성을 충분히 반영하지는

못했다[1,2]. 이를 보완하기 위해 LangChain 과 LangGraph 가 제안되었으며, 본 연구에서는 이를 결합해 SQL Agent 를 구현하였다.

한편, LLM 을 활용한 자동 요약 및 리포트 연구도 활발히 진행되고 있다. 의료 기록이나 재난 상황 보고서 등에서 LLM 은 대량의 데이터를 빠르게 정리해 인간 친화적 형태로 제공하는 데 강점을 보여왔다. 이러한 흐름을 반영해 본 연구는 보호대상자의 데이터를 LLM 에 입력하고, 프롬프트 엔지니어링과 Few-shot 학습을 통해 일일 상태 보고서로 변환하는 방식을 채택하였다.

2.2 SQL Agent 설계

본 연구에 사용된 데이터베이스는 user 테이블(개인정보)과 스마트워치로부터 제공받는 데이터를 포함하는 event 테이블(심박변이도, 스트레스, 움직임 위험도, PPG 위험도, 위치 정보 등)로 구성된다. SQL Agent 는 LangGraph 기반으로 구현되었으며, 전체 과정은 네 단계로 요약된다.

첫 번째 단계에서는 사용자 발화를 입력으로 받아 인물명, 날짜, 관심 지표, 구역 조건, 위치 연결 여부와 같은 핵심 정보를 구조화된 슬롯 형태로 추출한다.

두 번째 단계에서는 데이터베이스 스키마를 반영하여 질의의 안정성을 확보한다. 구체적으로는 사용할 수 있는 테이블과 컬럼을 제한하고, 각 컬럼의 의미를 사전에 정의한다.

세 번째 단계에서는 언어모델이 실제 SQL 질의를 생성한다. 이 과정에서 출력은 단일 SELECT 문으로 한정되며, 생성된 질의는 후처리 모듈을 거쳐 검증된다. 검증 항목에는 스키마 적합성 여부, 질문에서 지정한

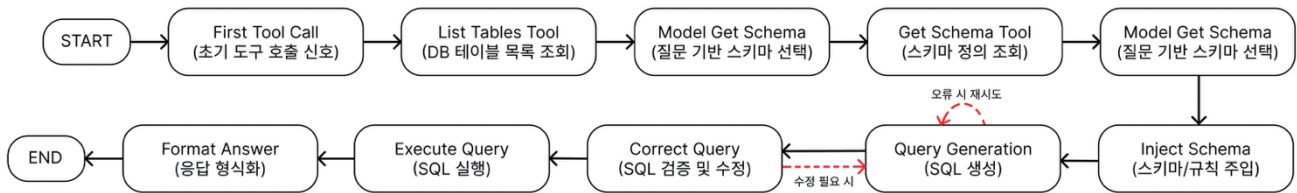


그림 1. SQL Agent 의 LangGraph 실행 흐름

지표와의 일치 여부, 그리고 DROP, UPDATE 와 같은 금지 명령어 포함 여부가 포함된다. 만약 오류가 발견되면 질의를 자동으로 수정하거나 재생성하도록 루프가 동작한다.

마지막 단계에서는 최종 확정된 질의를 데이터베이스에 실행하고, 그 결과를 사용자가 이해하기 쉬운 자연어 문장으로 변환한다.

2.3 LLM 기반 일일 리포트 생성

본 연구는 보호 대상자의 누적된 생체 데이터를 기반으로 LLM 모델을 사용해 일일 상태 리포트를 생성하는 것을 목적으로 한다. 사용자가 직접 데이터에 접근하고 위험 상황을 판단하는 데 소요되는 시간을 줄이고, 보다 직관적인 보고서 형태로 정보를 제공하기 위해 LLM 을 도입하였다.

LLM 이 제공하는 일일 리포트는 크게 네 가지 항목으로 구성된다. 첫째, 핵심 요약 항목을 통해 하루 동안 관찰된 전반적인 상태를 간단히 정리한다. 둘째, 위험으로 감지된 이벤트의 총 횟수와 각 이벤트의 발생 타임라인을 제시한다. 셋째, 보호 대상자가 불안정 상태의 조건을 두 가지 이상 동시에 만족한 경우를 별도로 표시하여, 불안정 데이터 간의 연관성을 파악할 수 있도록 한다. 넷째, 종합 의견 및 주의할 점과 제안을 제시한다.

최종 보고서 생성 과정은 데이터 입력→쿼리 생성→데이터 추출→LLM 입력→보고서 생성으로 이루어진다. 다량의 데이터를 정확하게 추출하고 보고서 생성에 필요한 특정 형식으로 제어해야 하기 때문에 직접 SQL 문을 작성할 수 있는 Raw SQL 방식의 SQLAlchemy 를 적용했다. 사용자의 프롬프트 입력을 단순화하기 위해 원하는 보호 대상자의 이름과 날짜만 입력하면, 해당 값들을 쿼리문에 바인딩하는 방식으로 실행한다. SQLAlchemy 는 데이터 베이스에 접근하여 위의 위험 이벤트 조건에 해당되는 데이터의 타임라인과 횟수를 추출하고 리스트 형태로 제공한다.

리스트 형태로 추출된 위험 상황 데이터는 자연어 생성 모델에 입력 데이터로 사용된다. LLM 이 보고서 형식을 빠르게 학습시키기 위해 FewShot 프롬프팅을 적용하였다. 이 과정에서 예시 보고서를 제공하여 모델이 항목별 작성 패턴을 학습하도록 하였다. 또한 SQL 실행 결과 일부를 함께 입력하여 데이터의 의미가 정확히 반영되도록 하고, 각 보고서 항목별 작성 과정을 단계적으로 안내함으로써 모델의 생성능력을 높였다. 이러한 프롬프트를 기반으로 LLM 은 최종 보고서를 생성한다.

2.4 연구 결과

본 연구에서는 보호대상자 관리 도메인에 특화된 SQL Agent 와 일일 리포트 기능을 대상으로 성능을 검증하였다. 먼저 SQL Agent 는 30 개의 자연어 질문-정답 쌍 실험에서 86.7%의 정확도를 달성하였고, SQL 유효성은 100%로 스키마 규칙을 위반한 사례가 없었다. 초기 질의 실패 시에도 83%의 비율로 복구에

성공했으며, 평균 응답 시간은 약 1.8 초로 실시간 활용 가능성을 확인하였다.

보고서 생성 기능을 평가하기 위해 [박주연, 2025-08-17]을 입력한 결과, 시스템은 핵심 요약, 주요 지표, 발생 타임라인, 종합 의견의 네 가지 항목으로 구성된 일일 보고서를 자동 생성하였다. 이는 데이터베이스에서 위험 이벤트를 추출하고, 프롬프트 엔지니어링을 거쳐 직관적인 텍스트 리포트로 변환할 수 있음을 보여주었다.

종합적으로 본 결과는 제안된 시스템이 세부 질의 응답과 전반적 상태 요약을 모두 지원함으로써, 보호대상자 관리 업무에서 정확성, 안정성, 실시간성, 직관성을 고루 갖춘 통합적 도구로 활용될 수 있음을 입증한다.

III. 결론

본 논문에서는 보호대상자 관리에 특화된 LLM 기반 통합 시스템을 구현하였다. LangChain 과 LangGraph 를 결합한 SQL Agent 는 질의 생성·검증·실행 과정에서 안정성을 확보하였으며, 프롬프트 엔지니어링을 적용한 일일 리포트 생성은 복잡한 데이터를 하루 단위로 종합해 직관적으로 제공할 수 있음을 확인하였다. 이러한 결과는 기존 Text-to-SQL 연구가 벤치마크 성능에 치중했던 한계를 넘어 실제 도메인에서의 적용 가능성을 보여주며, 데이터 기반 의사결정의 효율성을 높일 수 있음을 시사한다.

향후에는 정확도와 처리 속도의 정량 평가, 다중 데이터베이스 지원, 실시간 스트리밍 분석, 보안 강화를 통해 시스템을 확장할 계획이며, 이를 통해 보호대상자 관리뿐 아니라 의료·재난 관리 등 다양한 안전 분야에도 응용될 수 있을 것이다.

ACKNOWLEDGMENT

이 논문은 경찰청이 지원한 ‘사회적 약자 보호 강화 기술 개발(www.kipot.or.kr)’의 지원을 받아 수행된 연구결과입니다.

[과제명: 저전력 복합측위, 근접탐색 기술 기반 범죄피해

안전조치 대상자 위치추적 통합관제 플랫폼 개발/과제번호:

RS-2023-00236101]

참 고 문 헌

- [1] W. Hwang, J. Yim, S. Park, M. Seo, and M. Lee, "SQLova: Text-to-SQL Generation with Sketch-based Encoder-Decoder," *arXiv preprint arXiv:1902.01069*, 2019.
- [2] T. Yu et al., "RAT-SQL: Relation-Aware Schema Encoding and Linking for Text-to-SQL Parsers," *ACL*, 2020.