1. **축구 선수 검색 작업**
   1. 작업 목표: 사용자 질의 기반 선수 검색 챗봇
   2. 작업 내용: 정형 DB 제작 및 데이터베이스 기반 선수 검색 LLM 모델
   3. 작업 사항
      1. 원본 데이터 수집 (FM 2023.csv)
         1. FM은 축구 시뮬레이션 게임인 “Footall Manager”의 약자
         2. 해당 게임에서는 실제 선수들의 데이터를 수치화하여 저장
         3. <https://www.kaggle.com/datasets/platinum22/foot-ball-manager-2023-dataset>
      2. 데이터 전처리
      3. 정형 데이터베이스 제작
      4. Text2sql 모델 생성
      5. SQL 구문 실행 체인 생성
      6. SQL 실행 결과 기반 출력 체인 생성
      7. 선수 이미지 크롤링 함수 생성
   4. 데이터 전처리 및 정형 DB 제작
      1. 변수 명 한글 매핑 (영어 -> 한국어)
         1. ex) “Name” -> “이름”, “Dribbling” -> “드리블” 등
      2. 몸무게 결측치 보완
         1. 몸무게가 0으로 입력된 선수 존재
         2. KNN 알고리즘을 이용하여 선수의 키로 몸무게 예측
         3. 단순 평균으로 보간하는 방식보다 신뢰성 확보 가능
      3. 잠재 능력 변환
         1. 나이가 어린 선수들 중 잠재력이 음수를 가지는 선수 존재
         2. 해당 게임에서 제공하는 기준을 이용하여 0~200 사이의 수치로 변환
      4. 대표 포지션 선정
         1. 선수들을 포지션 별로 구분하기 위해 대표 포지션 선정
         2. 원본 데이터의 각 포지션 점수를 이용하여
         3. GK(골키퍼) / D(수비수) / M(미드필더) / AM(공격형 미드필더) / ST(공격수) 로 구분
      5. 정형 데이터베이스 구축
         1. 선수의 개인 정보와 포지션 별 주요 스탯을 이용하여 테이블 생성
         2. 포지션 별 5개의 테이블을 데이터베이스로 구축
   5. SQL 구문 생성 체인 생성
      1. langchain의 ChatOpenAI 라이브러리 이용
      2. “gpt-4o-mini” 모델을 이용하여 한국어 버전 text2sql 모델 사용
      3. 프롬프트 엔지니어링을 통해 생성 SQL 구문에 정확성 향상
   6. SQL 구문 실행 체인 생성
      1. 앞선 체인에서 생성된 SQL 구문을 실행하여 데이터를 추출하는 커스텀 체인
      2. 추출한 데이터는 LLM모델이 인식하기 쉽도록 자연어 형식으로 변환
      3. ex) “선수의 이름은 손흥민이고, 포지션은 공격수이고,,,, “
   7. SQL 실행 결과 기반 최종 응답 생성 체인
      1. langcahin의 ChatOpenAI 라이브러리 이용
      2. “gpt-4o-mini” 모델을 이용
      3. 앞선 체인에서 추출한 데이터를 참조하여 최종 문장 생성
      4. 프롬프트 엔지니어링을 통하여 답변의 신뢰성 및 형태 조정
   8. 선수 이미지 크롤링 함수 생성
   9. 프롬프트 엔지니어링
      1. SQL 생성 체인과 최종 응답 생성 체인에서 LLM 모델 이용
      2. 각 모델의 출력 결과의 신뢰성/정확성/출력 형태를 개선하기 위한 프롬프트 구성
      3. SQL 생성 체인 프롬프트 예시
         1. “실제 테이블의 컬럼 이름만 포함하여 SQL 구문을 생성해주세요.”
         2. “한글로 입력된 선수나 국가의 이름은 가장 가까운 영어 이름으로 변환해 주세요”
         3. “축구와 관련이 있는 자연어가 입력될 때만 SQL 구문을 생성해주세요.”
      4. 최종 응답 생성 체인 프롬프트 예시
         1. “{{"Name": "이름", "설명": "마크다운 형식으로 자연스럽고 자세한 설명"}} “의 형태로 출력해주세요
         2. “각 선수의 설명에는 다음 요소를 포함해 주세요”
         3. “어떤 역할을 수행할 수 있는지, 어떤 스타일인지 요약해줘”
   10. 시스템 구성
       1. 사용자 자연어 질의 입력
       2. 사용자 질의 기반 SQL 구문 생성 체인
          1. 자연어를 SQL로 변환
       3. SQL 구문 실행 체인
          1. 플레이어 DB에서 SQL 구문 실행
       4. 최종 응답 생성 체인
          1. 추출한 데이터 기반 최종 응답 생성
       5. 자연어 출력
2. **전체 시스템 아키텍쳐**

도표, 평면도, 기술 도면, 개략도이(가) 표시된 사진

AI 생성 콘텐츠는 정확하지 않을 수 있습니다.