뉴스 데이터 기반 맞춤형 기업 추천 서비스

9조
신정윤, 유재승, 이승훈, 정철희

문제 정의

취준생을 위한 기업 추천 서비스의 필요성

정보 접근성 제한

취준생은 기업에 대한 정보 접근성과 비교 기준이 제한적이어서 자신에게 적합한 기업을 체계적으로 탐색하기 어려움

정형화된 정보 제공

기존 채용 플랫폼은 연봉, 복지, 근무지 등 정형화된 정보 위주로 제공되어 기업의 전반적인 이미지나 최근 동향을 반영하기 어려움

맞춤형 서비스 부족

따라서 다양한 데이터를 종합해 취준생 관점에서 맞춤형으로 기업을 추천해주는 서비스가 필요함

기업 평판 정량화의 중요성

비정형 정보의 가치

취준생은 기업 선택 시 연봉이나 복지 외에도 기업 평판, 사회적 이슈 등 비정형 정보를 함께 고려해야 함

뉴스 데이터의 의미

기업 뉴스 데이터는 기업의 이미지와 신뢰도를 간접적으로 보여주는 핵심 자료가 될 수 있으나, 이러한 정보는 뉴스나 기사 속에 흩어져 있어 정량적 비교가 어려움

정량화의 필요성

따라서, 뉴스 데이터를 기반으로 기업 평판을 정량화하는 것이 중요함

주제 소개

- 뉴스 데이터 기반 맞춤형 기업 추천 서비스
  - 뉴스 데이터를 수집하여 기업 평판 요소별로 감성 분석을 수행
  - 기업 외부 평가 중심의 뉴스 데이터뿐 아니라, 기업 내부 평가를 반영하기 위해 잡플래닛 리뷰 데이터를 부가적으로 활용
  - 분석된 결과를 바탕으로 실시간 평판 점수를 산출
  - 개인의 관심 요인을 반영해 취준생에게 적합한 상위 기업(Top-k) 을 추천
  - 추천 결과는 채용 정보 + 평판 분석 리포트 형태로 제공

핵심 기술

뉴스/리뷰 데이터베이스 구축 및 관리

대규모 뉴스 데이터를 효율적으로 저장하고, 조건 조회가 가능하도록 설계

채용정보 데이터베이스 구축 및 관리

기업별 채용 정보 데이터를 효율적으로 저장하고, 조건 조회가 가능하도록 설계

감성 분석 및 임베딩 기반 벡터 인덱싱

문서 단위로 감성 점수 및 의미 벡터를 생성하여 유사 이슈 기반의 실시간 평판 계산 및 유사 기업 탐색 지원

추천 및 보고서 생성 (LLM 기반)

평판 점수가 높은 기업 Top-k의 뉴스·채용 정보를 통합하여 LLM이 자동으로 취준생 맞춤형 추천 보고서를 생성

주제 소개

- 데이터
  - 소스 : 네이버 뉴스 기사, 잡플래닛 기업 평판 리뷰, 잡코리아 채용정보
  - 기업 선정 : KMAC 선정 가장 존경받는 기업 Top 100 중 일부 선정

실시간 기업 평판 반영

뉴스는 기업의 경영, 인사, 사회적 이슈, 투자 동향 등을 즉시 반영하는 가장 민감한 정보원

- 감성 분석을 통해 긍정/부정 기사 비율을 계산하여 기업 이미지 변화를 수치화 가능
- 시간별, 주제별로 변화된 기업 이미지를 추적하여 실시간으로 평판 변화 감지 가능

기업 신뢰도 및 사회적 반응 측정

신제품, 규제, ESG, 인사 등 다양한 주제가 공개되어 기업 브랜드 평판 형성에 직접적영향을 미치는 공개 커뮤니케이션 채널

- 소비자, 직원, 시민 등 다양한 이해관계자의 의견을 반영하여 기업의 사회적 반응을 측정 가능
- 기업의 신뢰도를 객관적인 데이터로 백업하여, 취준생의 판단 근거로 제공

→ 뉴스 + 리뷰 데이터를 기반으로 기업의 실시간 평판 점수를 산출하고, 이를 채용 데이터와 연계하여
취준생 맞춤형 Top-k 기업 추천 및 리포트 자동 생성을 목표로 함

기대 효과

취준생 관점의 정보 접근성 향상

실시간 평판 파악
복잡한 뉴스 검색 없이 실시간 평판 점수를 통해 기업의 최신 이미지를 빠르게 파악 가능

정량적 이해
지원을 고려하는 기업의 최근 동향과 사회적 평가를 정량적으로 이해 가능

맞춤형 리포트 제공
LLM 기반 맞춤형 요약 및 추천 기능을 통해 개인의 직군·관심 분야에 따라 자동으로 기업 비교 리포트를 제공

기업 분석 및 비교 자동화

통일된 평판 점수 체계
여러 기업의 뉴스 데이터를 동일한 평판 점수 체계로 관리하여 기업 간 비교 및 산업 내 동향 분석이 용이

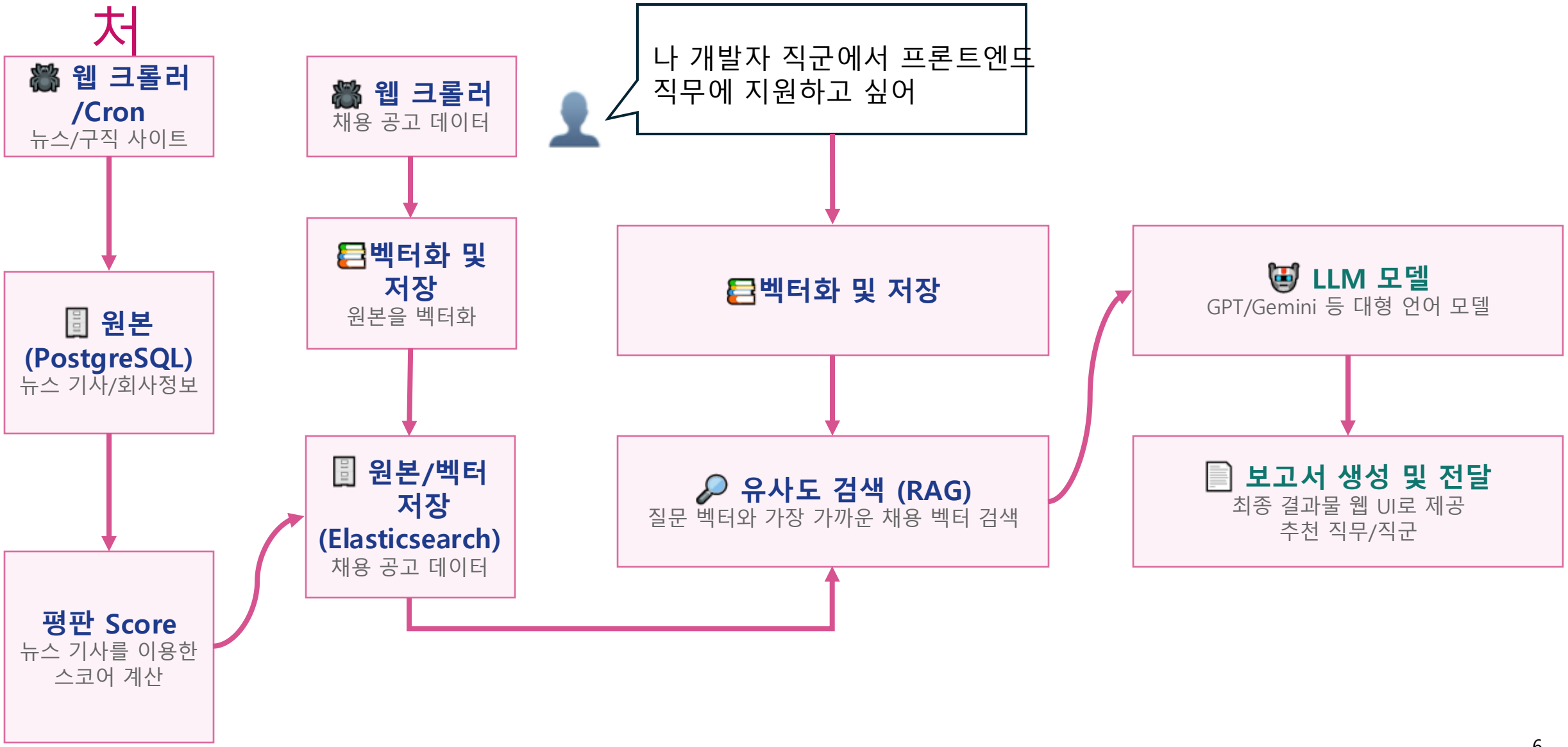
자동화된 비교 분석
기업 간의 뉴스 데이터를 자동으로 비교 분석하여
취준생의 판단할 수 있도록 지원

기술적 효율성

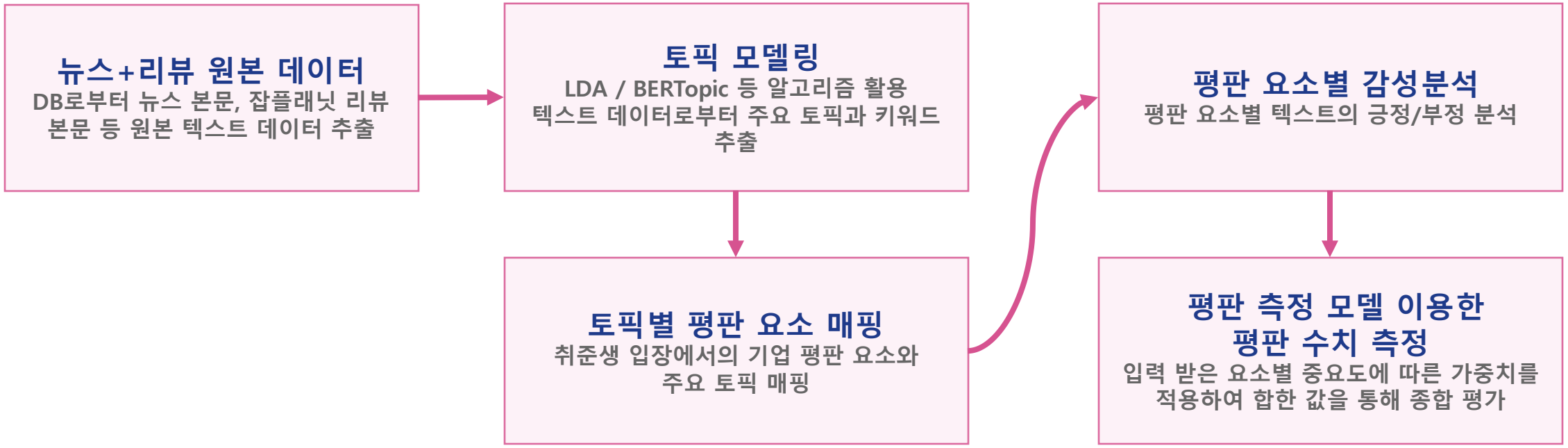
DB 인덱싱
하루 수백 건씩 쏟아지는 뉴스 데이터를 DB 인덱싱을 통해 효율적으로 관리 가능

빠르고 정확한 평판 계산
감성 분석 + 임베딩 기반 벡터 검색으로 대규모 비정형 뉴스 데이터에서도 빠르고 정확한 평판 계산 가능

통합 아키텍처


웹 크롤러 /Cron
뉴스/구직 사이트
원본 (PostgreSQL)
뉴스 기사/회사정보
평판 Score
뉴스 기사를 이용한 스코어 계산
웹 크롤러
채용 공고 데이터
벡터화 및 저장
원본을 벡터화
원본/벡터 저장 (Elasticsearch)
채용 공고 데이터
나 개발자 직군에서 프론트엔드 직무에 지원하고 싶어
벡터화 및 저장
유사도 검색 (RAG)
질문 벡터와 가장 가까운 채용 벡터 검색
LLM 모델
GPT/Gemini 등 대형 언어 모델
보고서 생성 및 전달
최종 결과물 웹 UI로 제공
추천 직무/직군

평판 점수 측정 모듈 프레임워크



$$\text{평판 점수} = \sum_{k=1}^{K} \omega_k S^{(k)}$$

$S^{(k)}$: k번째 평판 요소의 감성 점수
ω_k : k번째 평판 요소에 대한 가중치

가중치 학습

- Pairwise Logistic Ranking
  - 기업 i의 평판 점수
    - 점수 : $\hat{R}_i = \sum_{k=1}^{K} \omega_k S_i^{(k)}$ $S_i^{(k)}$: i 번째 기업의 k번째 평판 요소 감성 점수
    - 제약 : $\omega_k \geq 0,\ \sum_{k=1}^{K} \omega_k = 1$
      ($W \in \Delta$) 로 표현
  - 학습 목표
    - 실제 순위에서 $i \succ j$ (가 보다 상위) 인 모든 쌍 집합 P에 대해 $\hat{R}_i > \hat{R}_j$ 가 되도록 W를 학습
  - 손실 함수
    - $\mathcal{L}(W) = \sum_{(i,j)\in P} \log\left(1 + \exp\left(-\left(\left(S_i - S_j\right)^T W\right)\right)\right)$ s.t. $W \in \Delta$
    - $S_i = \left[S_i^{(1)}, \ldots, S_i^{(K)}\right]^T$
  - 프로젝션 경사 하강법을 이용하여 최적화
    - $W^{(t+1)} \leftarrow \Pi_\Delta\left(W^{(t)} - \eta \nabla_W \mathcal{L}(W^{(t)})\right)$
    - 초기값 $\omega_k^{(0)} = \frac{1}{K}$
    - Π_Δ : 제약 조건을 만족하는 공간으로의 프로젝션

타당성 검증: 설문조사 기반 기업 순위와의 비교

- 뉴스 기반 평판 측정 결과의 신뢰도 검증
  - 뉴스 데이터를 기반으로 산출된 기업 평판 점수를 통해 매긴 기업 평판 순위와 기존 설문조사 기반 기업 평판 순위가 유사한 경향을 보이는지 비교·검증
  - 통계적 검정을 통해 모델의 타당성 확보
- 데이터 구성
  - 2024년도 데이터를 train set으로 사용
  - 2025년도 데이터를 test set으로 활용하여 외적 타당성 검증 수행
- 통계적 검정 방법

스피어만 순위상관계수 (Spearman's ρ)

평가 내용: 두 순위 집단 간의 비선형적 일치도 평가

의미: ρ 값이 유의하게 높을 경우, 뉴스 기반 평판 순위가 설문조사 순위와
통계적으로 강한 양의 상관관계를 가져 일치도가 높음을 의미

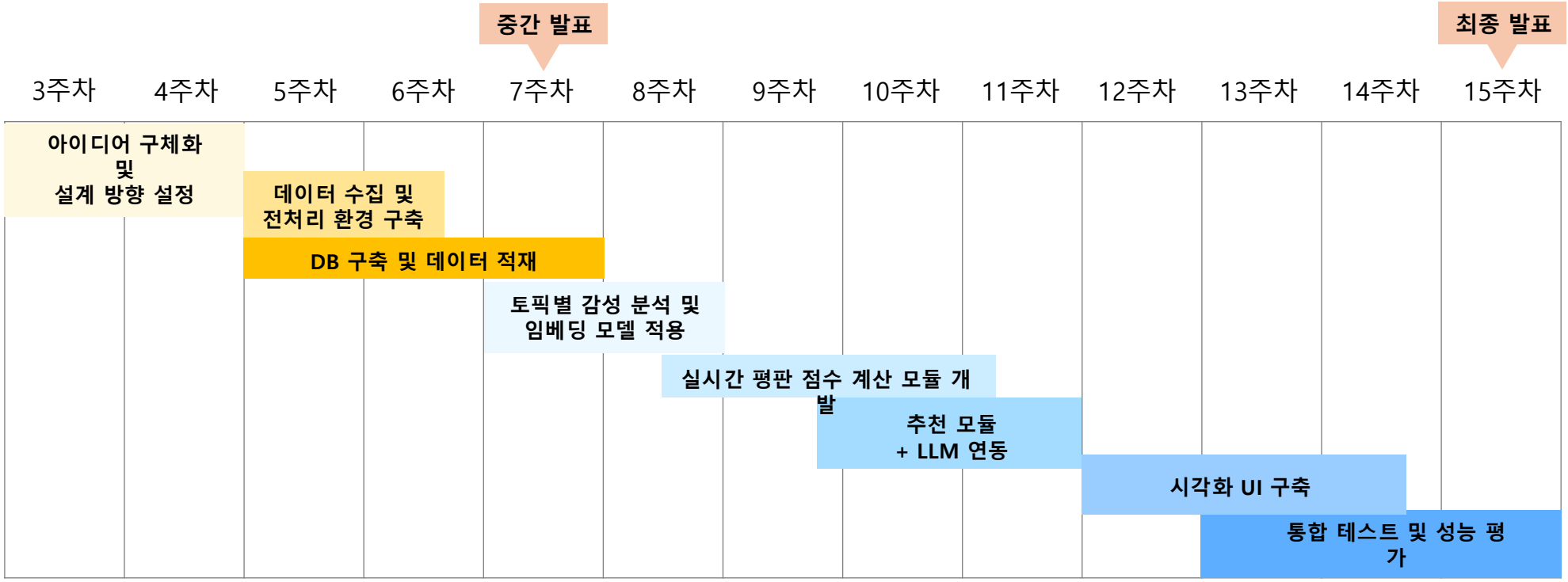
켄달의 타우 (Kendall's τ)

평가 내용: 순위 쌍 간 일관성과 불일치 정도를 기반으로 순위 상관성 측정

의미: τ 값이 유의미하게 크면 높은 순위 일치도를 가짐을 의미하며, 이는 설문조사 기반 평판 순위 결과와의 일관성을 가짐을 나타냄

결론: 두 지표 모두 유의한 양의 상관을 보일 경우, 뉴스 기반 평판 모델은 설문조사 결과와 일관된 순위 경향을 나타내며 모델의 타당성이 확보되었음을 의미

향후 일정



Proposal

- 주제 및 아키텍처 설계
- 데이터 수집 및 DB 구축
- 초기 평판 점수 계산 구조 시연

Final

- 실시간 평판 점수 산출
- Top-k 추천 및 LLM 기반 리포트 생성
- Streamlit 대시보드 시각화