뉴스 데이터 기반 맞춤형 기업 추천 서비스

단순한 스펙 매칭을 넘어, 당신의 가치관에 맞는 기업을 찾아드립니다.

9조

신정윤, 유재승, 이승훈, 정철희

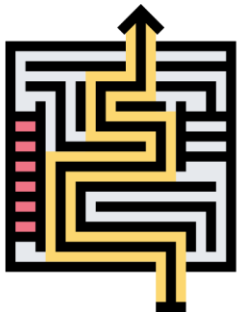
Problem Definition

기업 정보는 많지만, 기업 선택 과정에서 취준생은 방대한 정보 속 무엇이 중요한지 판단하기 어려워 개인의 가치관과 상황에 맞춘 기업을 선별하기 어려움



가치관 기반 추천의 부재

"채용 사이트에서 정보 탐색 시 내가 중요하게 생각하는 요소는 반영되지 않고, 단순히 조건만 매칭되면 기업 정보를 제공할 뿐이다."



기업 탐색의 어려움

"기업은 많고, 나는 무엇을 기준으로 선택해야 할지 모르겠다."



기업 평판 변화의 불확실성

"최신 기업 분위기/이슈를 일일이 파악하기 어렵다."

취준생의 가치관과 기업의 실시간 평판 정보를 통합하여 '나에게 맞는 기업'을 선별해주는 새로운 방식의 추천 시스템이 필요

Datasets

소스 : 네이버 뉴스 기사, 잡플래닛 기업 리뷰, 잡코리아 채용정보, 청년 패널 조사

기업 선정 : KMAC 선정 가장 존경받는 기업 Top 100 중 일부 선정

뉴스 데이터

외부 관점
언론, 전문가 시각을 통해 기업에 대한 객관적 평가 제공

객관성
보도·팩트 기반으로 작성되며 객관적 신뢰도 높음

변동성
단발성 사건·이슈·사고에 따라 관심도와 평판이 급격히 변화

기업의 최신 상태를 추천 시스템에 반영하기 위해 사용

리뷰 데이터

내부 관점
재직자·퇴사자의 실제 경험 기반으로 기업 내부 문화, 분위기를 반영

주관성
개인의 경험과 감정이 반영되어 주관성이 존재

지속성
기업의 장기적·근본적 특성을 보여줌

취준생 개인의 선호·상황에 맞춘 개인화 추천을 가능하게 하기 위해 사용

As-Is & To-Be

실시간 평판 기반 개인 맞춤형 기업 추천 서비스

	평판 분석	추천 기준	탐색 방식
As-Is	기업 평판 변화의 불확실성	가치관 기반 추천의 부재	기업 탐색의 어려움
	뉴스 데이터와 잡플래닛 리뷰 데이터를 주기적으로 수집하여 기업의 평판 요소별 감성 분석 및 점수 산출을 자동화	청년 패널 조사를 활용한 프로파일링 및 군집화를 통해 취준생의 가치관 파악	기업 정보와 취준생 정보를 하나의 그래프 구조로 통합
To-Be	실시간 데이터 기반 자동 감성 분석 및 평판 요소별 점수 업데이트 시스템 구축	사용자 가치관 – 기업 평판 요소 매핑 기반 추천 모델 구축	Link prediction을 통해 취준생에게 적합한 Top-3기업 추천

Expected Outcomes



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 분석 용이

자동화된 비교 분석

기업 간의 뉴스 데이터를 자동으로 비교 분석하여 취준생이 판단할 수 있도록 지원



기술적 효율성

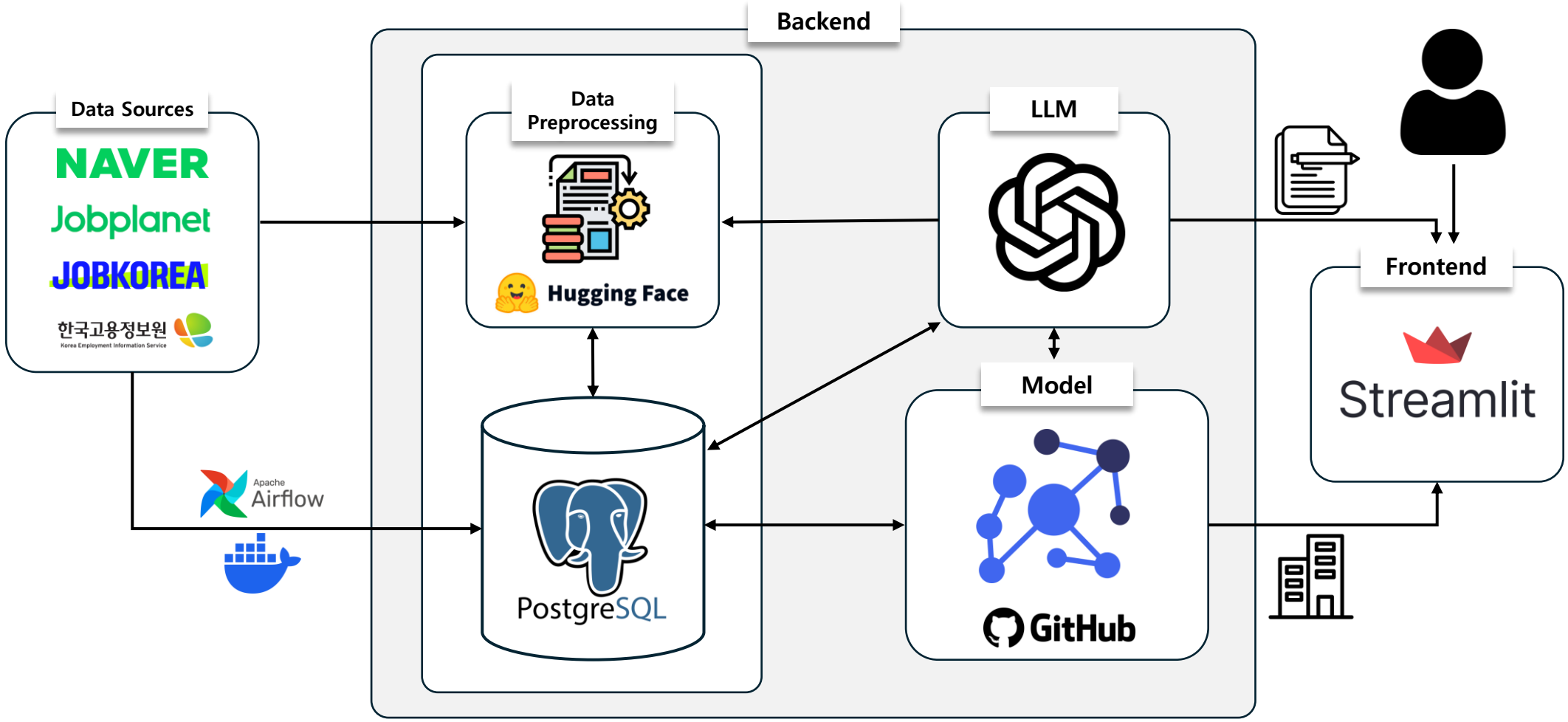
DB 인덱싱

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

Architecture



Data Collection & Preprocessing

데이터 수집 프로세스

수집 데이터

네이버 뉴스 본문
기업 외부 평판 및 최신 이슈 파악

잡플래닛 리뷰 데이터
기업 내부 평판 및 구성원 평가 반영

청년 패널 데이터 (2021-2023)
사용자 가치관 모델링 및 인구통계 정보 추출

채용 공고 데이터
추천 기업의 최신 채용 정보 제공

데이터 수집 방식

Automated

네이버 뉴스 / 잡플래닛 리뷰
Airflow + Docker를 활용하여
매일 자동으로 크롤링

Manual/File

청년 패널
연도별 데이터 파일 수집 후 통합

채용 데이터
수작업 수집

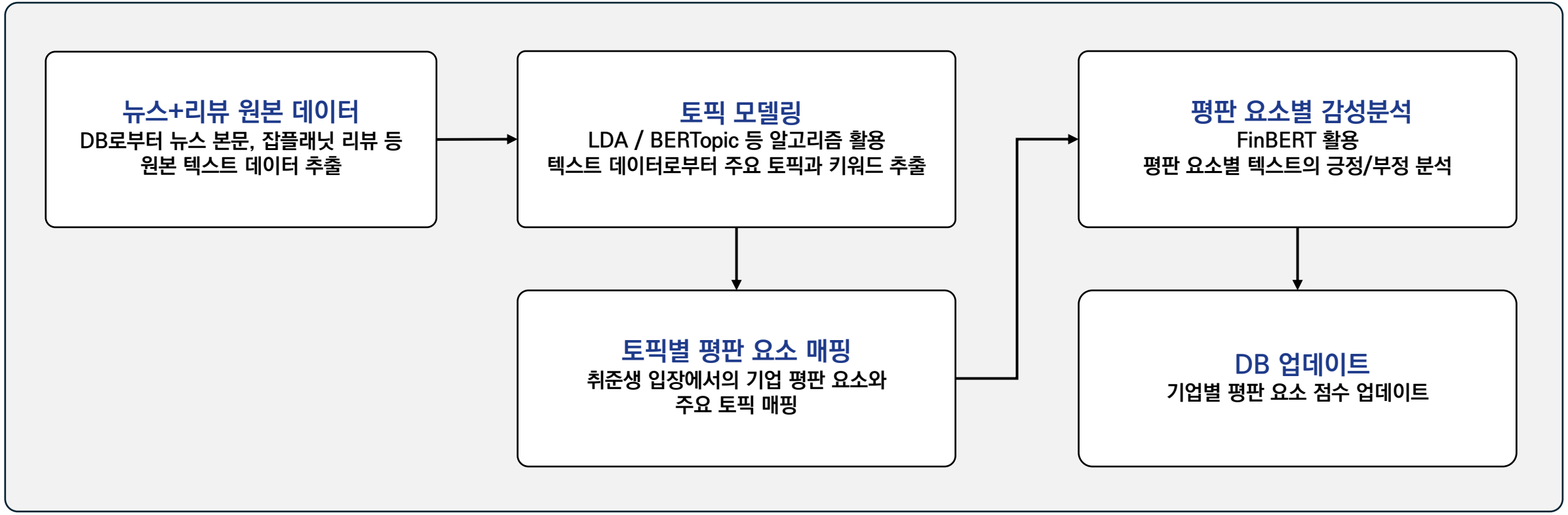
저장 구조



PostgreSQL

데이터 전처리
클러스터링 분석
지식 그래프 구축

Reputation Scoring Framework



Labeling Process

그래프에서 응답자의 선호 기업을 정답 Label로 사용하기 위해 평판 요소 기반 매핑



결측치 처리

청년 패널 데이터에서 응답의 일부가 결측

↓

나이, 성별, 학력 그룹 평균값으로 대체



사용자 클러스터

7개 평판 요소 벡터 기준 사용자 간 가치관 패턴 분석

↓

K-means로 사용자 군집 생성을 통해
유사 가치관을 가진 사용자 그룹 형성

↓

클러스터별로 평판 요소 벡터 평균값 계산



선호 기업 Labeling

지식그래프 학습용 선호 기업 정답 생성

↓

각 클러스터의 평균 평판 요소 벡터와
기업의 평판 요소 벡터를 코사인 유사도를 이용해 비교

↓

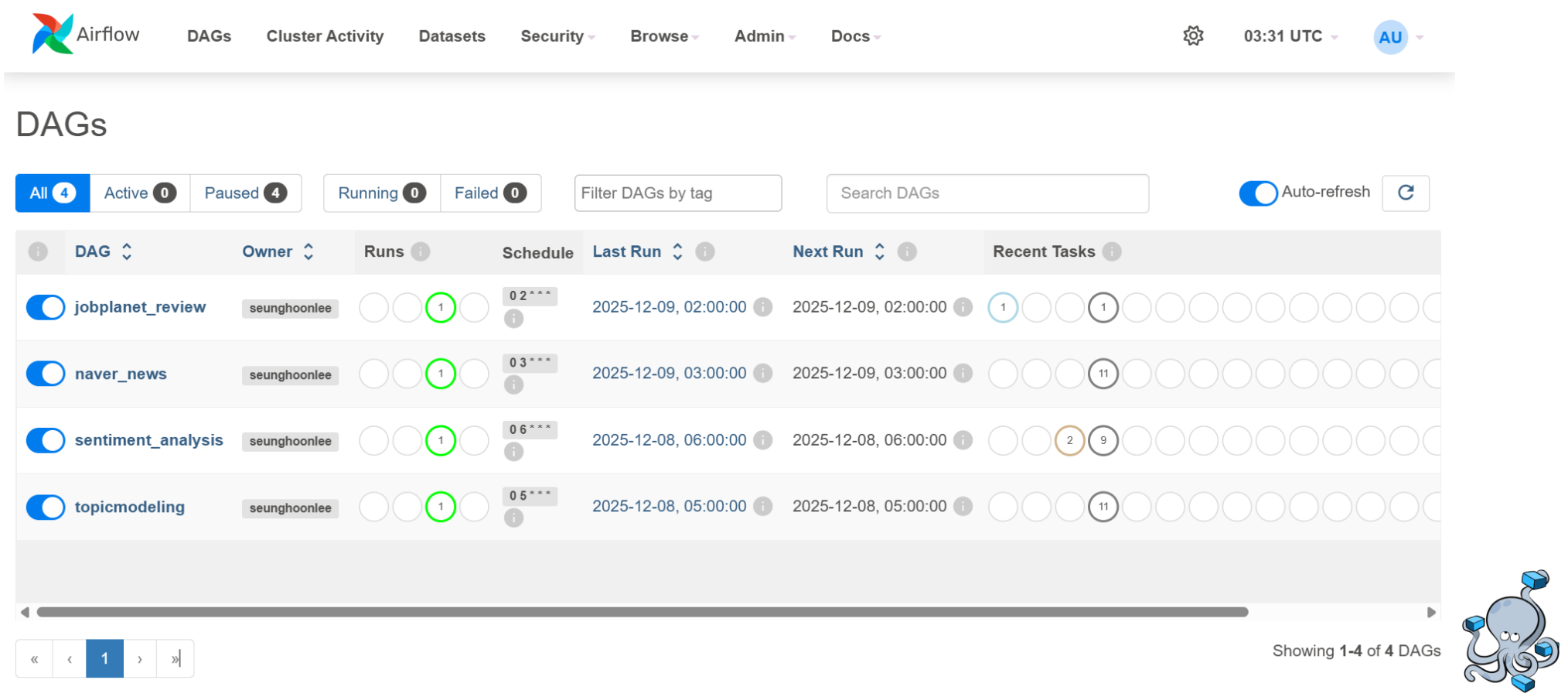
유사도 상위 top-3 기업을 해당 클러스터 사용자들의
잠재적 선호 기업으로 라벨링

↓

그래프 학습에 활용

Data Collection & Preprocessing

매일 자동 업데이트



DB management

- DB table

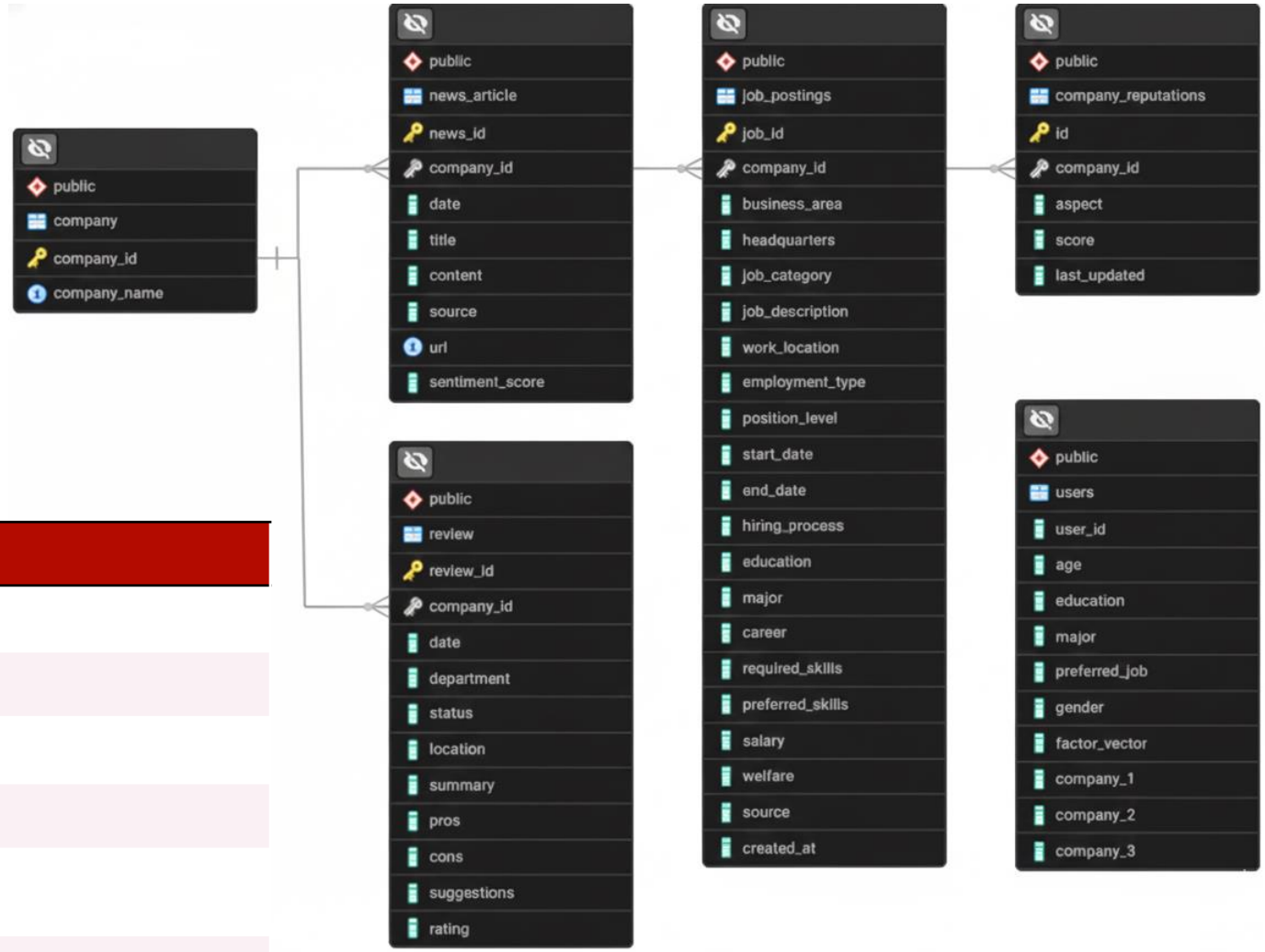
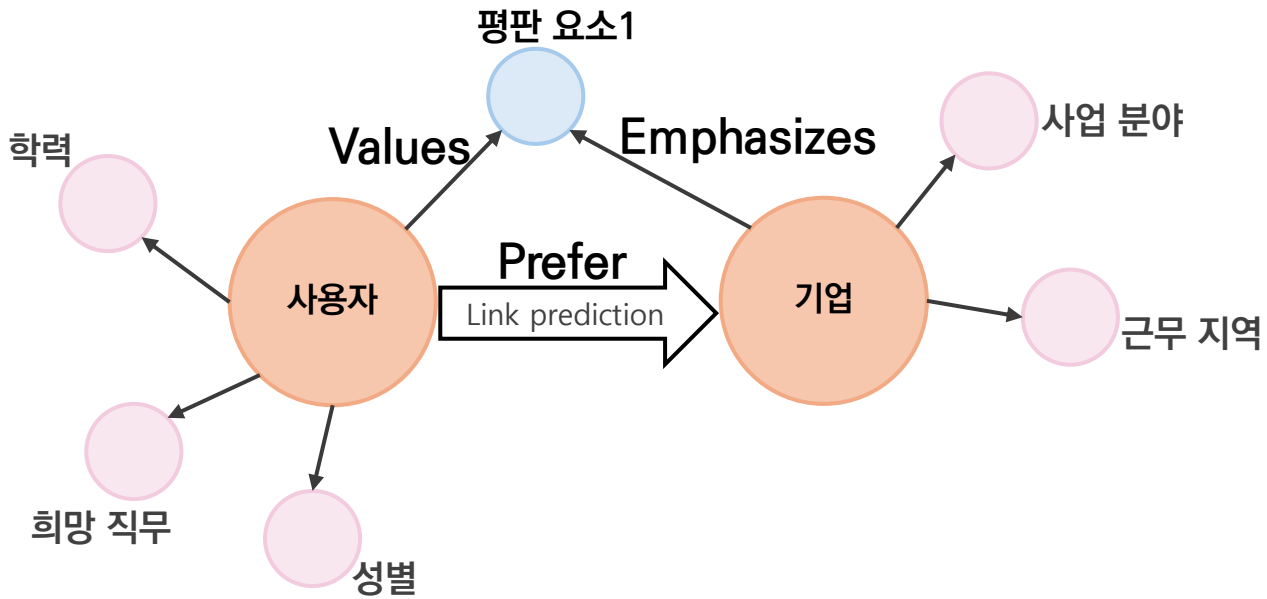


Table	설명
company	기업 관련 정보
news_article	뉴스 데이터 정보
review	리뷰 데이터 정보
job_postings	채용 공고 정보
company_reputations	기업의 실시간 평판 점수 정보
users	청년 패널 데이터 기반 사용자 프로필 정보

Method

- **지식 그래프 기반 추천**
  - Knowledge Graph Schema
    - 사용자, 기업, 평판요소 간의 관계를 지식그래프로 구성하고, entity와 relation을 임베딩 학습
    - 사용자 가치관과 기업 특성의 잠재적 연관성을 파악하여 학습된 임베딩을 이용해 (사용자 – prefer – 기업) 관계 점수를 예측하여 맞춤형 기업 추천
    - Entity : 지식그래프의 노드를 구성하는 객체
    - Relation : Entity 간의 의미 있는 연결 관계
    - Link prediction : 사용자와 특정 기업 사이의 prefer 관계가 얼마나 성립할 가능성이 높은지 점수로 계산하여 예측하는 과정



Method

- **지식 그래프 기반 추천**
  - Knowledge Graph Schema
    - User entity

엔티티 유형	설명
사용자	청년패널데이터 응답자 ID
나이대	응답자의 연령 구간(20대 초반/중반/후반)
성별	응답자의 성별(남/여)
학력	최종 학력 기준 응답 (고졸, 대졸, 석사학위 이상 등)
전공	전공 대분류 기준으로 분류(예: 컴퓨터·통신, 경영·경제, 사회과학 등)
희망 직무	취업 희망 직무(예: 애플리케이션 개발, 인공지능 등)
기업 선택 시 중요 요소	직업 안정성, 경제적 보상, 승진 기회, 직무 흥미, 자율성, 사회적 기여, 근무 유연성 등 주요 7개 요인의 중요도

    - 기업 선택 시 중요 요소는 개인별 평가 점수의 평균 이상인 값만 values relation으로 연결

Method

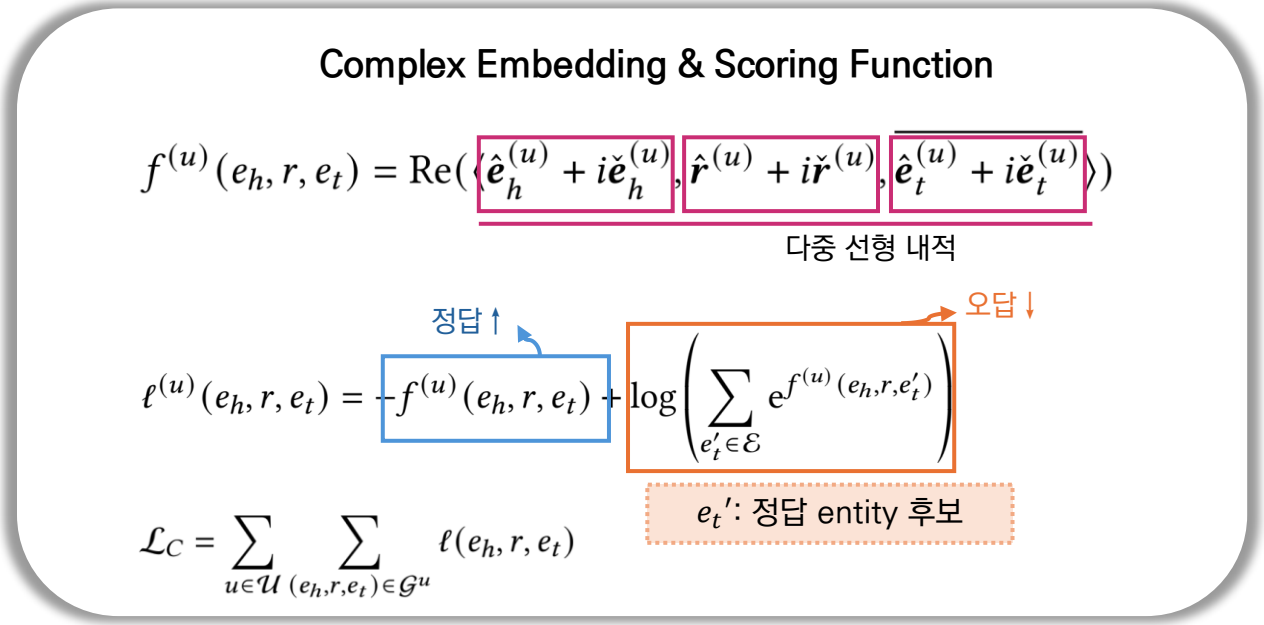
- **지식 그래프 기반 추천**
  - Knowledge Graph Schema
    - Company entity

엔티티 유형	설명
기업	기업명
사업분야	기업의 산업/업종 분류(자동차 제조 및 판매, 반도체 제조 등)
근무지역	기업의 주요 근무지
기업 평판 요소	직업 안정성, 경제적 보상, 승진 기회, 직무 흥미, 자율성, 사회적 기여, 근무 유연성 등 주요 7개 요인의 감성분석 결과 점수

    - 기업 평판 요소는 기업별 평판 요소 감성 점수 평균 이상의 요소만 emphasizes relation으로 연결

Method

- **지식 그래프 기반 추천**
  - Embedding & Link prediction
    - 그래프 내 엔티티와 관계를 저차원 벡터에 임베딩해서, 관계로 두 엔티티가 연결되는 정도를 계산하는 scoring function 기반 유추 진행

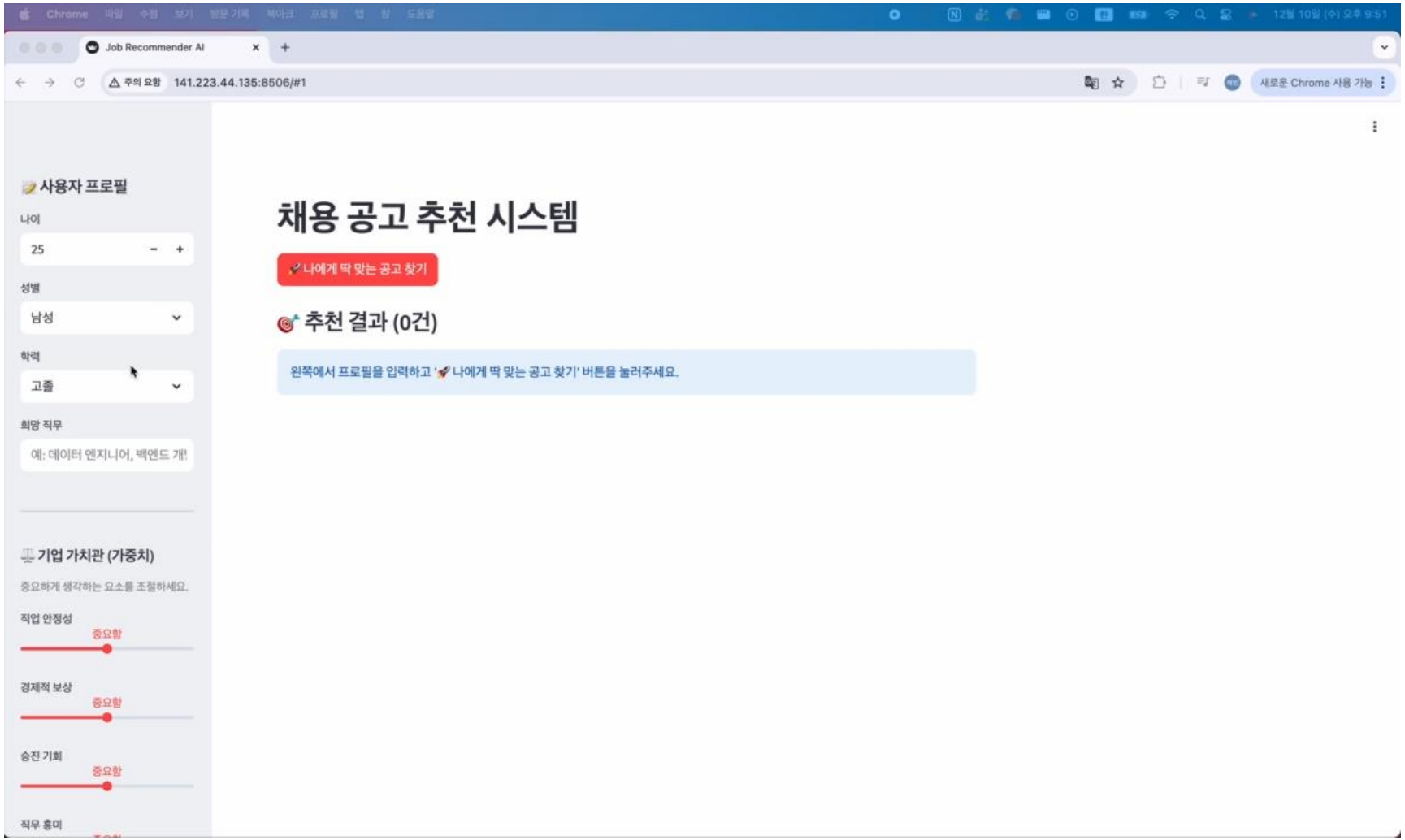


보조 손실

(user–educated–석사 학위)
같은 실제 fact triple을
재구성하도록 embedding 학습

- 선호 기업에 대한 정보가 없는 새로운 사용자의 input이 들어왔을 때, 나머지 user 속성을 이용해 관계 역추정을 통해 user embedding 후 Link prediction을 통해 선호 기업 top – 3 예측
- 예측된 상위 3개 기업에 대해 사용자 특성에 부합하는 채용 공고를 DB에서 조회하고, 이를 기반으로 LLM이 개인 맞춤형 기업 추천 보고서를 생성

Demo


Job Recommender AI
141.223.44.135:8506/#1
사용자 프로필
나이
25
성별
남성
학력
고졸
희망 직무
예: 데이터 엔지니어, 백엔드 개!
기업 가치관 (가중치)
중요하게 생각하는 요소를 조절하세요.
직업 안정성
중요함
경제적 보상
중요함
승진 기회
중요함
직무 흥미
채용 공고 추천 시스템
나에게 딱 맞는 공고 찾기
추천 결과 (0건)
왼쪽에서 프로필을 입력하고 '나에게 딱 맞는 공고 찾기' 버튼을 눌러주세요.

Result

1. 취준생에게 가장 알맞은 회사를 추천 가능하게 되어 취준생이 자신이 바라는 회사를 찾는데 매우 유리
2. 최신 데이터 업데이트 및 한눈에 들어오는 그래프와 보고서를 제공하여 취준생에게 신속한 정보 제공
3. 대규모 뉴스 및 리뷰 데이터를 확보하여 향후 다양한 데이터 분석 목적으로 활용 가능

- **향후 계획**
  1. 더욱 신속한 뉴스 및 리뷰 업데이트 주기 적용
  2. 청년패널 데이터 외 달리 참고할수 있는 더욱 대규모 데이터셋으로 GNN모델 재학습

Thank you

Thank you for listening to the presentation

Pohang University of Science and Technology