

# 20

## 웹 크롤링 기반 블루레이 추천 서비스

소속 정보컴퓨터공학부

분과 B

팀명 라피스 블루

참여학생 박덕형, 박태준, 안형찬

지도교수 조준수

### 과제 개요

#### 과제 배경

블루레이는 전 세계적으로 DVD를 대체한 대표적 고화질 영상 매체

수집 가치가 높지만 국내 인지도와 정보 접근성은 여전히 부족

맞춤형 추천 서비스 부재로 정보탐색에 많은 시간 소요



-블루레이 예시 이미지-

#### 연구 목표

블루레이와 영화 매핑 알고리즘 구현

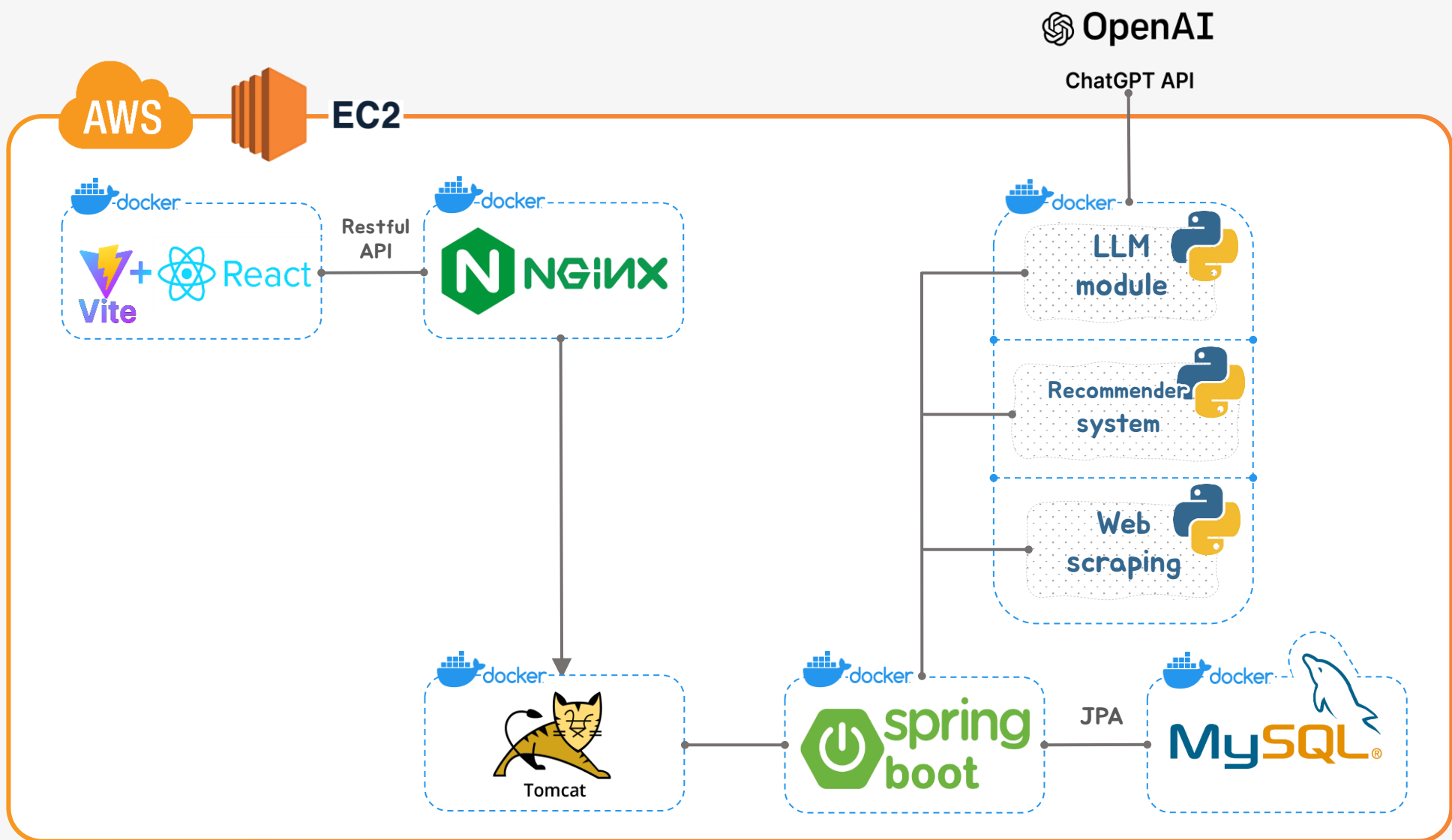
데이터 수집·정제 자동화 (웹 크롤링 + LLM)

추천 알고리즘 구현 (행렬분해 + 콘텐츠 기반 필터링)

상품 추천 및 블루레이 정보 제공 서비스 구현

### 시스템 구조 및 설계

#### 설계 구조도



#### 데이터 수집(웹 크롤링 + LLM)

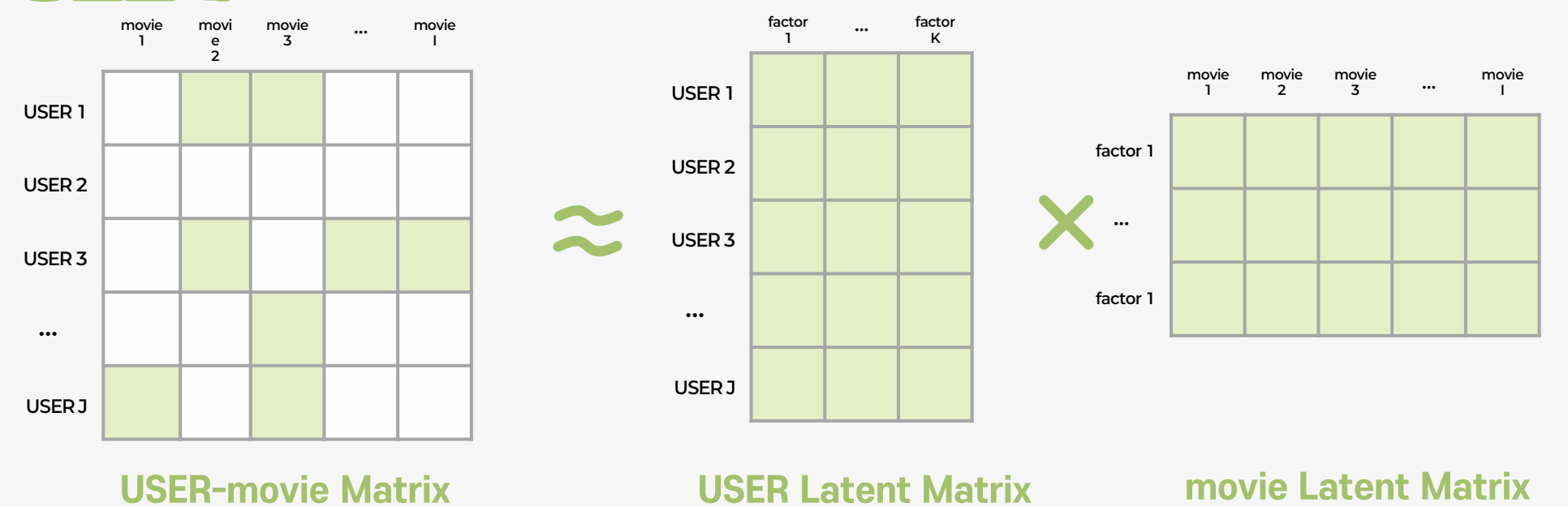


여러 온라인 소스로부터 웹 크롤링을 통해 블루레이 정보를 자동 수집 수집한 데이터를 LLM API를 이용하여 정제

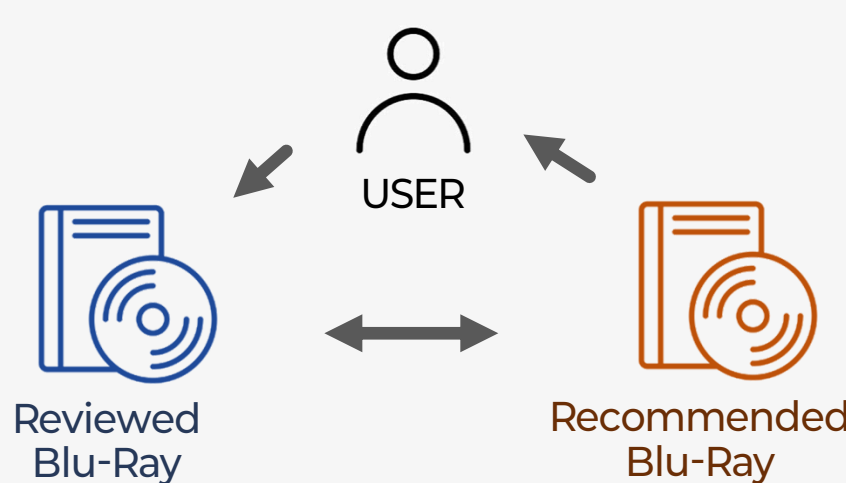
웹 크롤링을 통한 데이터 수집은 온전히 학습 목적으로 진행하였습니다.

#### 추천 알고리즘

##### 행렬분해



##### 콘텐츠 기반 필터링



무비렌즈 데이터셋을 활용, 행렬분해 알고리즘으로 사용자가 선호할 가능성이 높은 영화 후보군을 선별

영화 후보군에서 블루레이 데이터를 추출하고 콘텐츠 기반 알고리즘을 적용해 최종 블루레이를 추천

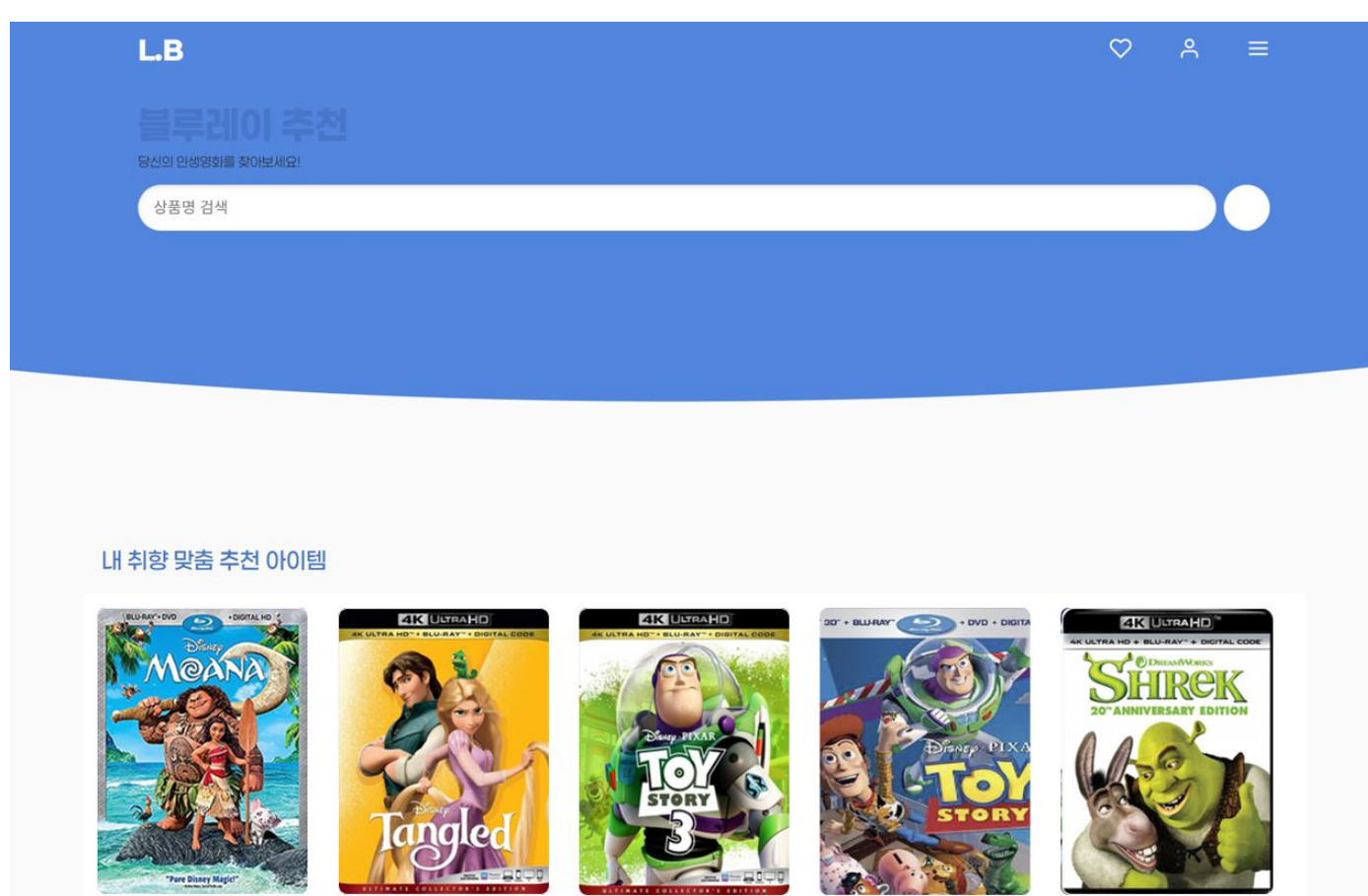
#### 데이터 매핑

$$\hat{y} = \alpha + \beta x + \sum_{i=1}^k \gamma_i z_i$$

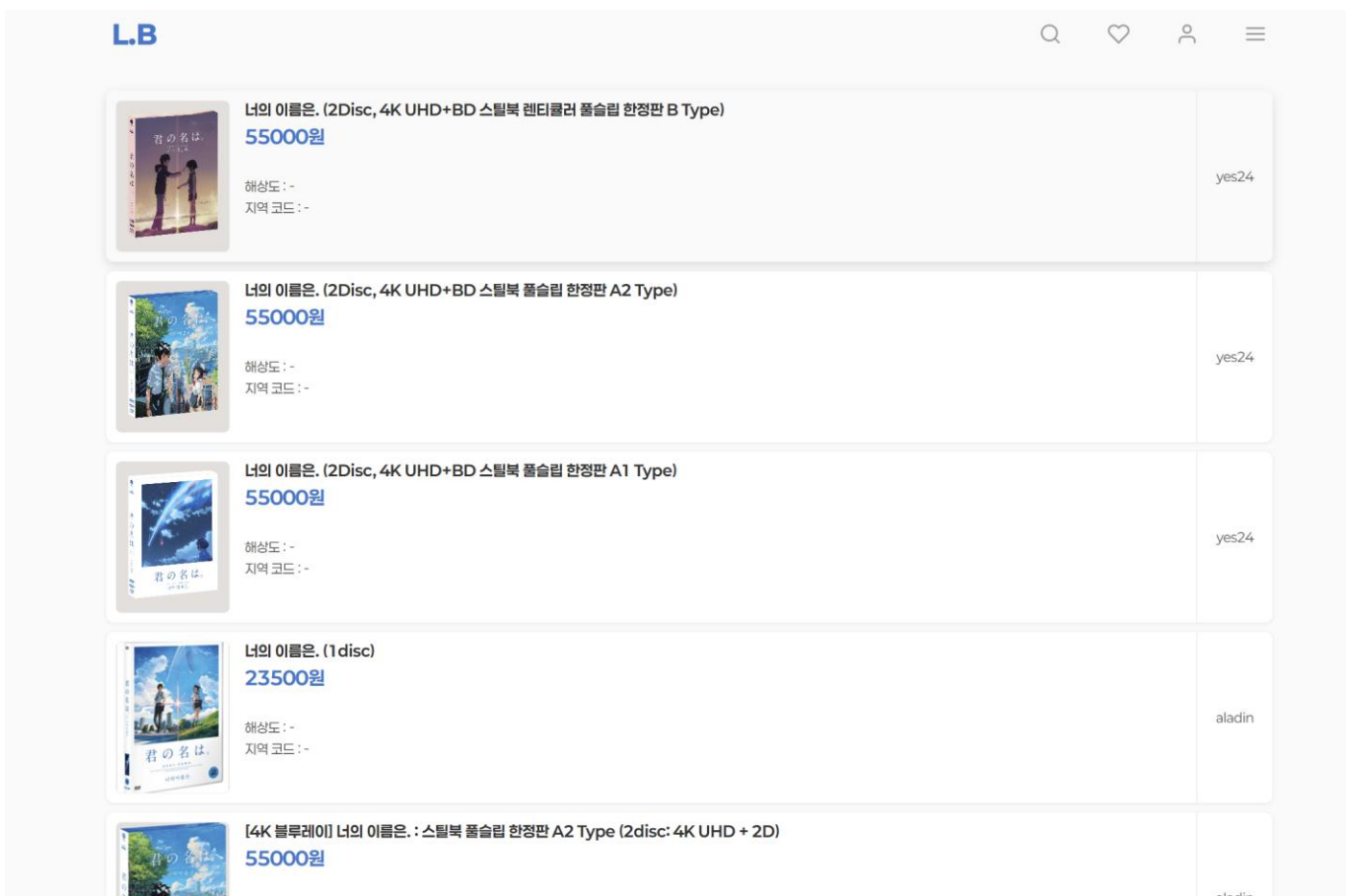
$\alpha$ : 절편(intercept)  
 $x$ : 블루레이 평점 ( $\beta$ : 가중치)  
 $z_i$ : 장르, 출시연도, 4K, 리마스터 여부 등과 같은 다른 특성들 ( $\gamma_i$ : 가중치)

블루레이 평점 -> 영화 평점  
블루레이의 정보를 가중치를 부여해서 영화 평점으로 계산

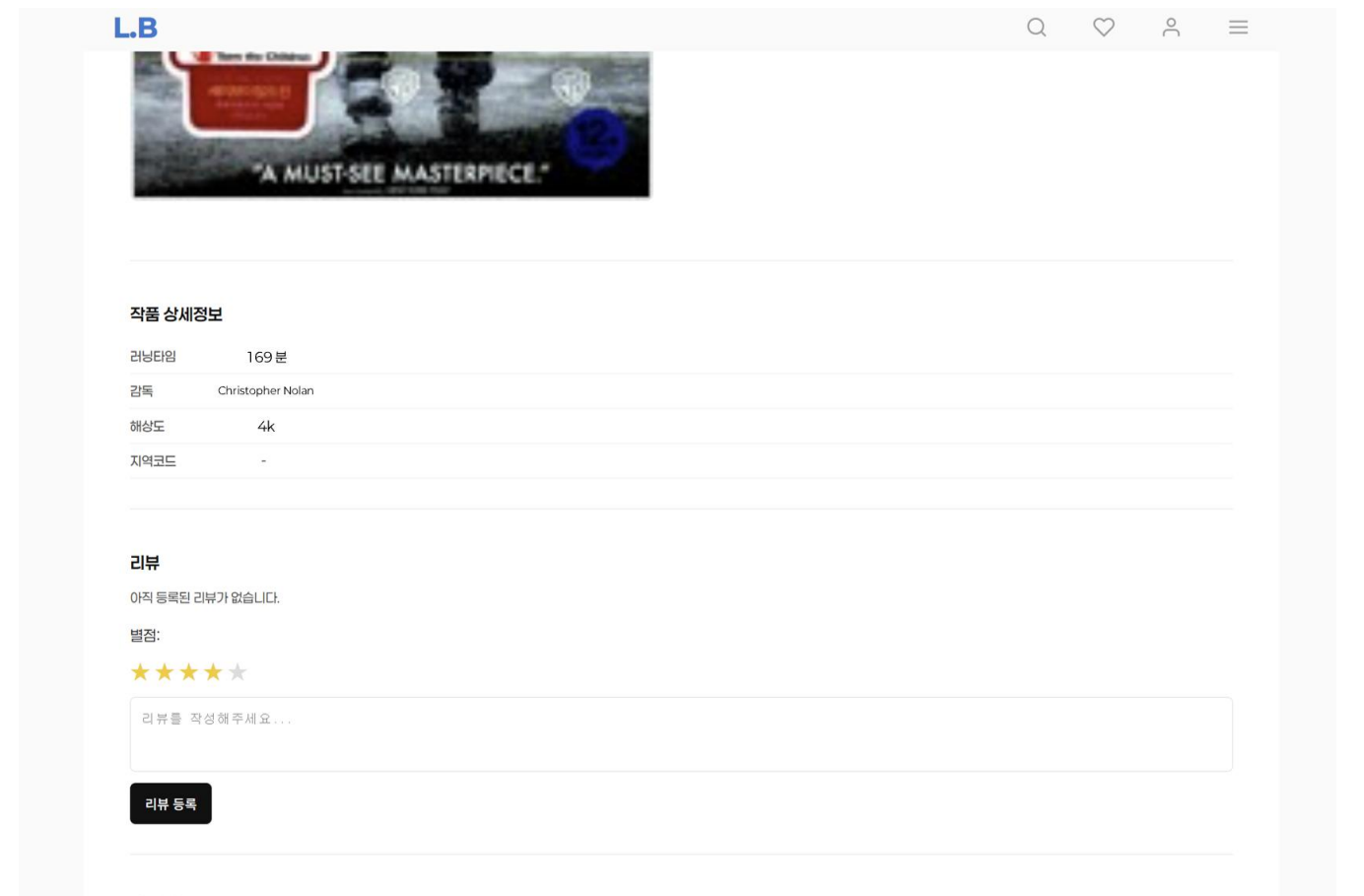
### 시스템 구현



-메인 페이지-



-검색 페이지-



-상세 페이지-