4차 5기 실무 Project

|  |
| --- |
| 스낵 리뷰 |

**2019년 11월 4일**

**Open API를 활용한 빅데이터 전처리 시각화**

리뷰 쓰고 가조

김동환

조규상

**목 차**

**1. 프로젝트 개요**  **1**

1.1 프로젝트 기획 배경 및 목표 1

1.2 구성원 및 역할 2

1.3 프로젝트 추진 일정 3

**2. 프로젝트 현황**  **4**

2.1 시장 분석 4

2.2 경쟁 제품 장단점 분석 6

2.3 차별화 핵심 전략 기술 9

**3. 프로젝트 결과**  **9**

3.1 주요 기능 및 상세 동작 9

3.2 사용 개발 도구 및 목표성 13

3.3 핵심 알고리즘 16

3.4 주요 동작 19

3.5 회차별 멘토링 결과 19

**4. 기대 효과**  **21**

4.1 향후 개선 사항 21

4.2 기대 효과 23

**5. 개발 후기**  **24**

**6. 강사 및 멘토 의견**  **25**

**1. 프로젝트 개요 [스낵 리뷰]**

Open API를 활용한 빅데이터 전처리·시각화 과정을 통해 배운 웹 프로그래밍 및 R 이용한 데이터 시각화를 실습해본 프로젝트입니다.

1.1 프로젝트 기획 배경 및 목표 [한국 문화 콘텐츠를 다루는 리뷰 플랫폼]

기획 배경 :

* 콘텐츠 판매 목적이 아닌 리뷰만을 다루는 리뷰 플랫폼의 부재
* 기존 리뷰들은 단순히 평점만으로 대상을 평가 => 세분화된 평가 지표를 제공
* 타 플랫폼들의 리뷰 정보들은 데이터들을 시각화하는 요소가 적음 => 시각화 제공

목표 :

* 영화, 도서 플랫폼 등의 사이트들을 모델로 하여 동일하면서도 추가적인 기능을 추가한 웹사이트 개발.
* 웹 개발과 R을 이용한 데이터 전처리 및 시각화 등 수업에서 배운 내용 등을 실습
* React, Spring Boot 등 수업에서 다루지 않았던 Tool을 함께 공부하면서 프로젝트에 적용

1.2 구성원 및 역할

|  |  |  |  |
| --- | --- | --- | --- |
| 이름 | 전공 | 역할 | 구현 부분 |
| 김동환 | 문헌정보학과 | 팀장 | 백엔드 구현 |
| 조규상 | 컴퓨터정보학과 | 팀원 | 데이터 전처리 및 시각화  프론트 기능 구현 및 디자인 |

1.3 프로젝트 추진 일정

|  |  |  |  |
| --- | --- | --- | --- |
| 구분 | 기간 | 활동 | 비고 |
| 사전  기획 | 9/20(금)  ~  9/30(월) | 프로젝트 기획 및 팀 구성 |  |
| 9/20(금) | PJT주제 선정, 팀(PM/팀원) 구성 | 2인/팀 |
| 9/27(금) | 프로젝트 멘토링  [프로젝트 방향 설정 및 현업프로젝트 소개] | 현업 멘토 참여 |
| PJT  수행  /  완료 | 9/30(월)  ~  11/5 (화) | 프로젝트 수행 |  |
| 9/30(화)  ~  10/18(금) | 프로젝트 설계 |  |
| 10/18(금) | 프로젝트 멘토링  [프로젝트 점검 및 기술자문 ] | 현업 멘토 참여 |
| 10/18(금)  ~  11/4(월) | 구현 및 테스트 |  |
| 11/5 (화) | 팀별 최종 발표 (구축 완료 보고) | 최우수 한 팀 선발  멘토 평가 |

**2. 프로젝트 현황**

2.1 시장분석

현재 ‘리뷰’ 라는 정보는 주로 블로그에서 사용자들이 개인적인 후기를 작성하거나 콘텐츠 제공 플랫폼에서 유저들의 선택을 도와주기 위한 참고 지표 정도로 다루고 있다.

그 예시로 ‘네이버 시리즈’, ‘리디 북스’, ‘왓챠’ 등이 있으며, 이들은 모두 콘텐츠 제공 플랫폼이기 때문에 리뷰의 시각화나 정렬에 신경을 쓰지 않은 곳들이 많으며, ‘오직 리뷰만을 위한 플랫폼’ 은

식당과 같이 특정 주제만에 한정된 플랫폼 밖에 존재하지 않는다.

그러므로 콘텐츠 제공에 신경쓰지 않고 문화 콘텐츠들(도서, 영화, 만화 등)의 리뷰만을 집중하여 제공, 관리해주는 플랫폼은 현재 콘텐츠 제공 플랫폼들이 과밀화된 시장상황 속에서 더욱 가치가 있는 신규 시장을 확보 할 수 있을 것이다,

2.2 경쟁 제품 장단점 분석

1. 네이버 시리즈

장점 :

* 리뷰를 단 유저들의 성별, 나이대를 시각화로 보여줌.
* 콘텐츠 제공 플랫폼들 사이에서 가장 규모가 크며, 리뷰 수가 많다.

장점 :

* 단순 리뷰들을 페이지 순으로 정리
* 해당 작품의 특징이나 평가를 알기 힘들다.
* 리뷰들의 정렬도 존재하지 않아서 많은 리뷰들의 가치가 사실상 무용이 된다.

1. 리디북스

장점 :

* 스포일러가 있는 리뷰를 유저가 스스로 가릴 수 있는 독특한 기능이 존재한다.
* 별점의 비율을 보여주면서 호불호가 갈리는 작품인지 알 수 있다.

단점 :

* 작품 별 특징을 알려주는 시각화가 없기 때문에 해당 작품의 특징이나 평가를 알기 힘들다.
* E-Book컨텐츠에만 제한이 되어있기 때문에 다른 서적들에 대한 리뷰를 남길 수 없다.

1. 왓챠

장점 :

* 평점을 시각화해서 보여줌으로써 호불호를 알 수 있다.
* 기존의 한 줄의 짧은 리뷰가 아닌 사각형 코멘트 창을 이용하여 특색 있는 디자인을 사용하였다.

단점 :

* 리뷰가 정렬되지 않아서 많은 리뷰들이 어떤 내용들을 담고 있는지 파악하기가 어렵다.
* 평가요소나 시각화가 없기 때문에 해당 작품만의 특징이나 평가를 알기 힘들다.

2.3 차별화 핵심 전략 기술

* ‘스낵 리뷰’는 해당 작품만이 가지는 특징을 레이더차트, 워드클라우드로 보여줄 수 있다.
* 리뷰 뿐만이 아니라 이벤트 크롤링 등을 이용하여 해당 콘텐츠에 대한 프로모션 정보들을 제공해줄 수 있다.

**3. 프로젝트 개발 결과**

3.1 주요 기능 및 상세 동작 [소제목 : 12pt / 맑은 고딕]

주요 기능 1): 영화 API를 통해 현재 한국 영화 박스오피스 순위 1~10위의 해당하는 영화들을 게시한다

스크린샷이(가) 표시된 사진

자동 생성된 설명

주요기능 2) 페이지 상세 보기

스크린샷, 텍스트이(가) 표시된 사진

자동 생성된 설명

- 페이지 상세보기 창을 이용해, 해당영화의 키워드와 독자들이 평가하는 스파이더 맵 차트를 보여준다(미구현)

3.2 사용 개발 도구 및 목표성

사용 개발도구 및 프로그래밍 언어 : R, Java, JavaScript, Spring boot

목표 : 수업 때 배운 R과 Java를 실습 및 React와 Spring boot와 같은 수업에서 다루지 않았지만 실무에서 널리 활용되는 언어들의 활용

* 1. 핵심 알고리즘

1. 유저들이 입력한 평점이 DB에 저장되어 레이더 차트에 반영
2. 유저들이 입력한 댓글들의 내용들을 전처리해서 워드클라우드에 적용
3. 영화, 책 API를 통해서 데이터를 얻어오고 시각화

3.4 주요 동작

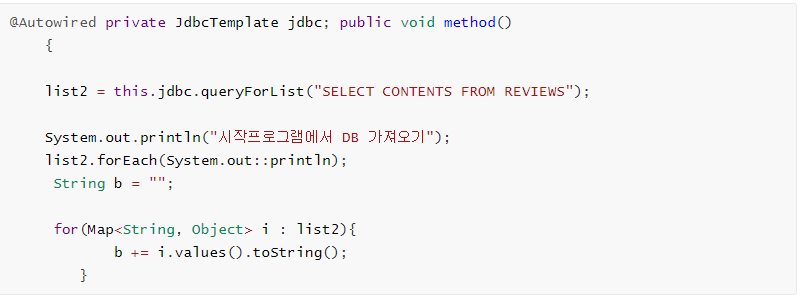
워드클라우드 :



DB의 리뷰데이터들을 하나의 String으로 합친 뒤, 전처리를 하여 워드클라우드를 실행,

But 중간의 개발환경을 바꾸는 과정에서 DB 설계를 재구성하고, 시간이 부족했기 때문에 React에서 데이터를 가져오는 과정을 구현하지 못 하였기 때문에, 따로 자바 프로그램에서 이미지를 만들고 프론트에서 사용하는 방식으로 처리하였다.

실행 코드 1 :



b 라는 String 변수에 DB에서 불러온 리뷰의 내용들을 모두 합쳐서 저장.

실행 코드 2 :

b 변수를 이용하여 R의 전처리 과정을 실행



영화 API 처리

- 백엔드에서는 직접 API를 통해 데이터를 받아온다. 영화진흥원에서 제공하는 API에서는 영화포스터를 제공하지 않는 만큼 IMDB API를 추가적으로 이용하여 포스터를 받아올수록 처리하였다.

1. MovieRestController



1. MovieService



- resTemplate을 통해 API와 간편하게 통신할 수 있었다.

1. Movie.js



- async, axios 를 이용하여 비동기 처리를 하였으며 추가적으로 데이터의 처리가 필요한 부분은 parseInfo를 통해서 수정하였다.

3.5 회차별 멘토링 결과

|  |  |
| --- | --- |
| 회차 | 내용 |
| 1회차 | 주제 관련 피드백 : 주제를 좁히고 수업에서 배운 내용을 토대로 기본을 구현하는 것을 중요시해라.  구현 방식 피드백 : 다른 API를 사용해서 기본 정보를 제공하는 것이 좋다.  결과 => API 사용 결정, 주제 축소 및 방향성 조정 |
| 2회차 | 구현 방향 피드백 : 새로운 것을 개발 하는지, 원래 있던 것을 연습하는 용도로 만들어보는 건지 구분  -> 연습하는 용도로 기존 사이트를 구현해보는 방향으로 전환  -> 기존 대상을 따라하되 기존 대상에 없는 기능들을 1~2가지를 추가하는식으로 개발  기능 관련 피드백 : 백엔드와 프론트를 나누어서 구현하는 것이 분산처리 아키텍처이며, 좋은 개발 방향이다.  발표 관련 피드백 : 기능에 대한 세부 설명이 필요하다 (ex방사선 차트)  결과 => 방향을 기존에 있던 것을 연습하는 방향으로 확정, 개발환경 조정 |

**4. 기대 효과**

4.1 향후 개선 사항

* 전체 리뷰를 볼 수 있고, 리뷰를 쓸 수 있는 리뷰 페이지 구현
* DB에서 데이터를 가져와서 실시간으로 시각화 요소들을 업데이트 하는 기능 구현
* 각 콘텐츠 사이트들에서 이벤트들을 크롤링해서 이벤트 정보 페이지 구현
* 소설, 만화 등의 콘텐츠들의 페이지 구현

4.2 기대 효과

구현이 완료가 된다면 시각화 요소들을 통해서 해당 콘텐츠의 특징을 한 눈에 파악할 수 있게 될 것이며, 유저들의 편의성이 증가할 것이다.

또한 영화 뿐만이 아니라 도서, 만화 등 국내 문화 콘텐츠들을 전부 다룰 예정이므로, 규모가 큰 전문 리뷰 플랫폼이 될 것이다.

이벤트 정보 등의 크롤링을 통해서 유저들의 편의성을 더욱 더 증가시켜줄 수 있을 것이다.

**5. 개발 후기**

|  |  |
| --- | --- |
| 성명 | 후기 |
| 김동환 | 배운 것을 다시한번 복습하려 진행했던 프로젝트였지만, 생각보다 쉽지 않았습니다.  처음 계획했던 초안을 포기하고, 다시 시작한 만큼 부족한 결과물이지만 남은 교육기간 짬짬히 채워나갈 예정입니다. |
| 조규상 | React나 Spring boot와 같이 수업에서 배우지 않았지만 실무에서 정말 많이 사용되는 도구들을 공부할 수 있어서 좋았고,    수업에서 배운 R 시각화를 Java와 연동시켜볼 수 있어서 좋은 경험이었습니다. |

**6. 강사 의견**

|  |  |  |
| --- | --- | --- |
| 평 가 요 소 | 배점 | 평 |
| 1. 아이디어 :  유사한 서비스 존재 유무 및 체계성 | /20 |  |
| 2. 개발 :  실제 구현 정도 및 배포 유무, 코드의 무결성 및 난이도, 현업적용도, 실무기술 반영정도 | /30 |  |
| 3. PJT 수행력 :  일정관리 및 역할분담, 목표 일정 달성도, 팀내 참여도 등 | /30 |  |
| 4. 준비도 :  프리젠테이션 및 프로젝트 준비 정도 | /20 |  |
| 계 | /100 | 강사 의견 필수 |