

RECOBOOK



teamreco

전혜진 김형원 안성근 이인영

CONTENTS

- i. 주제 선정 배경
- ii. 프로젝트 목표
- iii. Step 1 : 데이터 구축
- iv. Step 2 : 도서 한줄평 감정분석
- v. Step 3 : 콘텐츠 기반 필터링 추천 시스템
- vi. Step 4 : Flask 서버 기반 챗봇
- vii. Step 5 : App 배포
- viii. 개선점







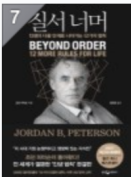



주제 선정 배경

N | 책 추천



네이버 베스트셀러 정보 ①

다른 사이트를 보시려면 클릭하세요. [다른 사이트 더보기](#)

교보문고	에스24	알라딘	인터파크도서	반디앤루니스	더보기
<div>1</div>  <p>울재 클래식스 3...</p> <p>도서구매</p>	<div>2</div>  <p>신곡(La Divina ...</p> <p>도서구매 e북구매</p>	<div>3</div>  <p>작은 별이지만 ...</p> <p>★★★★★ 10.00</p> <p>도서구매 e북구매</p>	<div>4</div>  <p>귀멸의 칼날 23/...</p> <p>도서구매</p>	<div>5</div>  <p>달러구트 꿈 백...</p> <p>★★★★☆ 8.53</p> <p>도서구매 e북구매</p>	
<div>6</div>  <p>아들이 돈 공부...</p> <p>도서구매 e북구매</p>	<div>7</div>  <p>질서 너머</p> <p>도서구매 e북구매</p>	<div>8</div>  <p>운의 알고리즘(...</p> <p>도서구매 e북구매</p>	<div>9</div>  <p>서른살, 비트코...</p> <p>★★★★★ 10.00</p> <p>도서구매</p>	<div>10</div>  <p>2021 선재국어 ...</p> <p>도서구매</p>	

주제 선정 배경

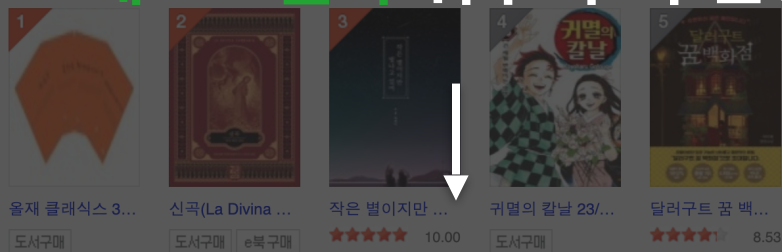
N | 책 추천



네이버 베스트셀러 정보

다른 사이트를 보시려면 클릭하세요. 다른 사이트 더보기

베스트셀러 위주의 책 선정

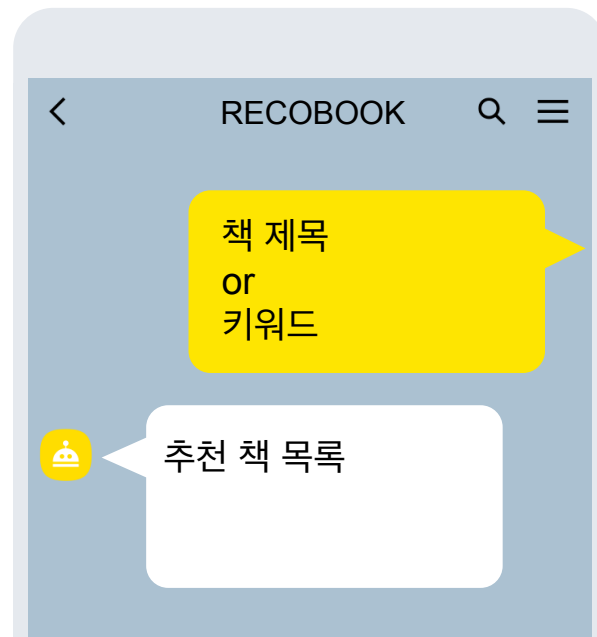
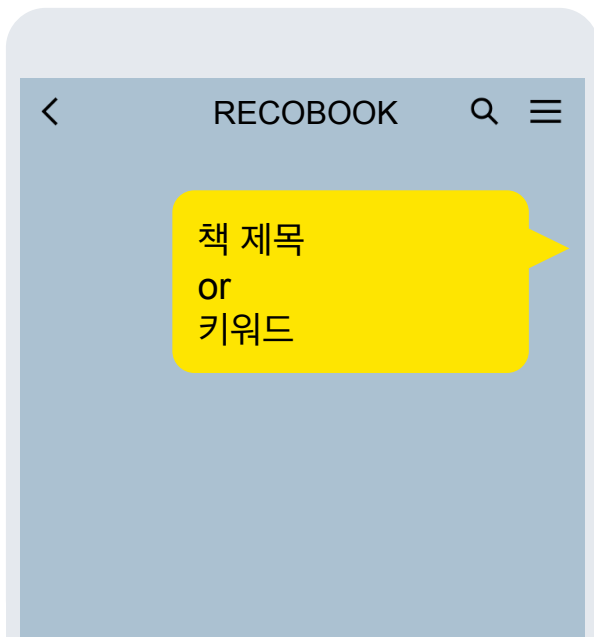


새로운 도서 추천 시스템 개발



프로젝트 목표

컨텐츠 기반 책 추천 모델을 구축하여 챗봇을 이용한 서비스를 개발한다



Step 1. 데이터 구축

Yes24 베스트셀러 800권 크롤링

: 책 id, 책 제목, 작가, 출판일, 카테고리, 리뷰등록자, 리뷰내용, 리뷰등록일

id	name	출판일	분류	작가	순위	평점	한줄평	댓글등록자	댓글등록일
99300613	귀멸의 칼날 23	2021년 04월 23일	['국내도서', '만화	고토게 코요하루	1	평점5점	재밌스	w*****4	2021-04-29
99300613	귀멸의 칼날 23	2021년 04월 23일	['국내도서', '만화	고토게 코요하루	1	평점5점	사요나라	0***a	2021-04-28
99300613	귀멸의 칼날 23	2021년 04월 23일	['국내도서', '만화	고토게 코요하루	1	평점5점	마지막권	g*****l	2021-04-28
⋮									
97255028	긴긴밤	2021년 02월 03일	['국내도서', '어린이	루리	79	평점5점	책 읽으면서 울어본 적 없는데 진짜 평	h*****4	2021-03-11
99015280	2021 이동기 영	2021년 04월 07일	['국내도서', '수험	이동기	80	평점5점	좋아용	w*****4	2021-04-29
⋮									
84807166	나는 무조건 합격	2019년 12월 16일	['국내도서', '자기	이윤규 저	800	평점4점	좋은데 유튜브에 있는 내용과 비슷한 존 유튜브 참고하셔도 좋을듯해요	안*꽃	2020-08-17
84807166	나는 무조건 합격	2019년 12월 16일	['국내도서', '자기	이윤규 저	800	평점4점	공부에 도움	w*****k	2020-08-16

Step 2. 한줄평 감정분석

1) 네이버 영화리뷰 긍/부정 모델

train data (훈련용 리뷰 개수) : 150000 -> 전처리 후 : 145393

test data (테스트용 리뷰 개수) : 50000 -> 전처리 후 : 48852

LSTM 모델로 학습 후 학습된 모델 적용하여 **sentiment_predict** (감정분석) 함수 생성

sentiment_predict : 토큰화, 불용어제거, 정수 인코딩, 패딩 후 예측값 추출

2) 영화리뷰 모델을 바탕으로 도서 한줄평의 긍/부정 평가

sentiment_predict 사용하여 긍/부정 확률 추출

책 ID로 데이터 그룹화하여 긍/부정 확률 평균 구하고 분류 기준으로 중복제거

Step 3. 콘텐츠 기반 필터링 추천 시스템

1) 제목 기반 책 추천 결과 도출


```
find_sim_book(merge_data, genre_sim_sorded_ind, user_book, 10)
```

2) 키워드 기반 책 추천 결과 도출

```
find_keyword_book(merge_data, genre_sim_sorded_ind, keyword, 10)
```


- merge_data : 전체 데이터
- genre_sim_sorded_ind : 코사인 유사도가 높은 순서대로 정렬된 행의 위치 인덱스
- user_book, keyword : input 값 (제목, 키워드)
- 10 : 상위 10개 책 추천

Step 4. Flask 서버 기반 챗봇

 레코북

재밌게 읽었던 책이나 원하는 키워드를 입력해주세요! 책을 추천해드립니다!

2021년 05월 24일

 레코북

안녕하세요 레코북입니다.

사용자님께서 재미있게 읽으신 책을 입력해주세요!

*제목으로 찾기: 책의 제목 입력

*키워드로 찾기: #형태로 태그 입력

오후3:23


ex) 나의라임오렌지나무
ex) #만화#국내도서

오후3:23

전송

Step 4. Flask 서버 기반 챗봇

1) 제목 기반 책 추천




레코봇

오후3:23 수상한학원

name
푸른사자와니니3
찰리와초콜릿공장
수상한기차역
유령고양이후쿠코1
십년가게4
요괴의아이를돌봐드립니다4
117층나무집
트리깍의생물
비밀의보석가게마석관1

오후3:23



레코봇


오후3:26 해리포터

name
재혼황후5
범인없는살인의밤
룬의아이들-블러디드4
빨강머리앤
프로젝트헤일메리
호밀밭의파수꾼
악녀는마리오네트세트
봉제인형살인사건
루팡의딸

오후3:26

Step 4. Flask 서버 기반 챗봇

2) 키워드 기반 책 추천

 레코북

오후3:28

#추리 #공포

name
이상한과자가게전천당10
이상한과자가게전천당10
파친코1
나미야잡화점의기적(봄날에디션)
이상한과자가게전천당9
이상한과자가게전천당9
구미호식당
범인없는살인의밤
시선으로부터,
봉제인형살인사건

오후3:28

 레코북

오후3:30

#교육 #수학

오후3:30

Step 5. App 배포



pythonanywhere 사용하여 웹 애플리케이션 배포

주소 : <http://staroot.pythonanywhere.com>

개선점

1. 데이터 추가

현재 베스트셀러 순으로 800권 크롤링. 추후 만권 이상의 데이터로 다시 학습시켜 범위 확대.

2. 추천 시스템 성능 향상

긍/부정 점수 이외의 다양한 성향분석을 통한 추천 성능 향상.

3. 구매 링크 추가

추천 책 리스트에 구매 링크를 추가하여 바로 구매할 수 있는 방식 추가.

THANK YOU

