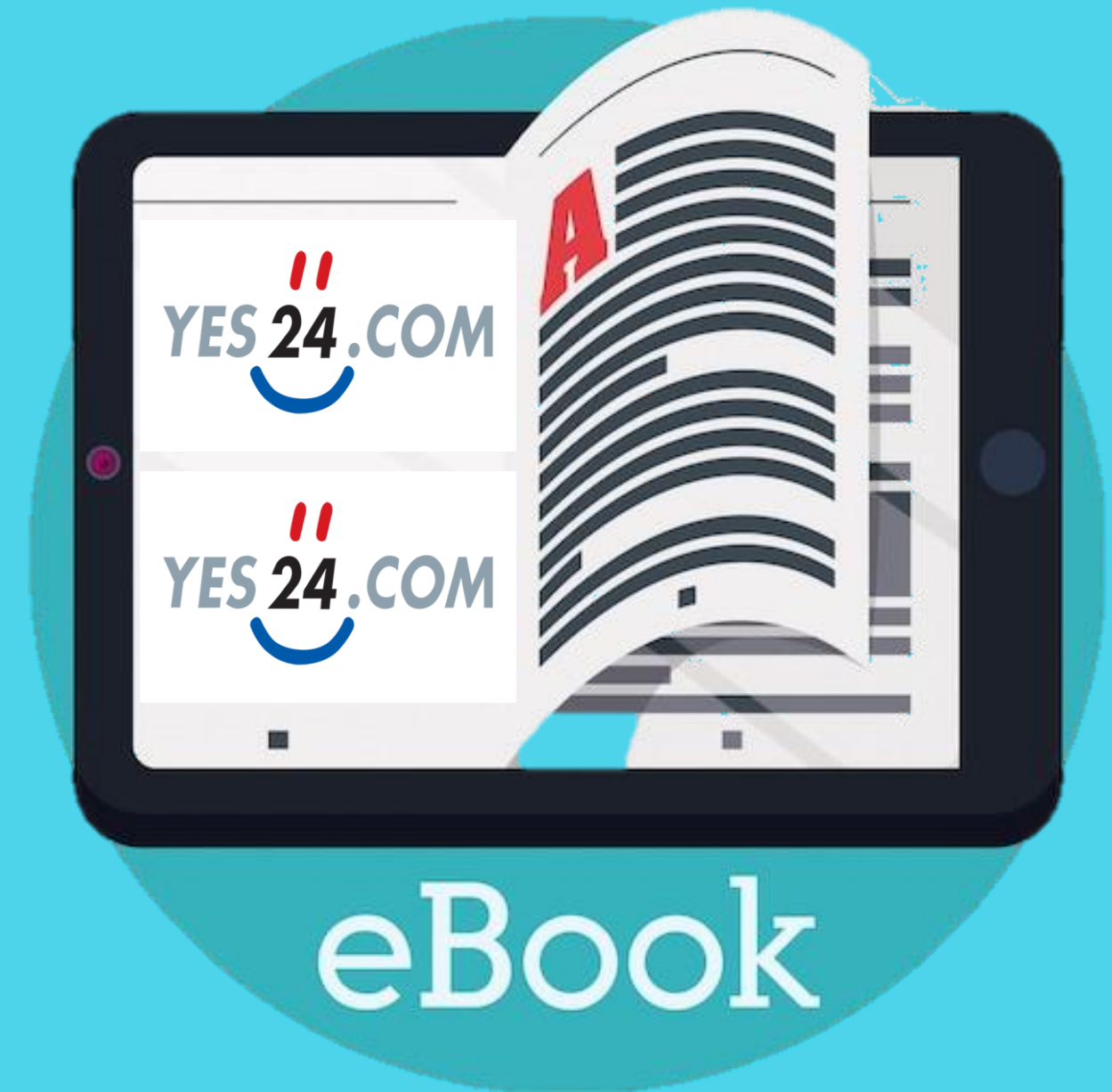


고마워, 부들북 서비스.



부들부들 조 #부경대 #학생들



제목과 미리보기를 활용한
맞춤형 소설 추천 시스템

팀원 소개

* A: 멘토, B: 멘티



전승훈(B)

- 데이터 전처리
- 모델 튜닝



강창훈(B)

- 데이터 크롤링
- PPT 제작



김규형(A)

- 데이터 크롤링
- 데이터 전처리
- 모델 튜닝
- 웹앱



김연지(B)

- 데이터 크롤링
- PPT 제작



김지원(B)

- 데이터 전처리
- 모델 튜닝



정성원(A)

- 데이터 전처리
- 모델 튜닝

TIME STAMP

6.13 (월)	6.14 (화)	6.15 (수)	6.16 (목)	6.17 (금)
<p>조 편성</p> <p>첫 인사 & 자기소개</p> <p>조 이름 결정</p> <p>조장 결정</p>	<p>프로젝트 주제 선정</p> <p>프로젝트 계획 수립</p> <p>멘토&멘티 기반</p> <p>역할 분류</p> <p>프로젝트 기획안 제출</p>	<p>웹 소설에서</p> <p>e-book으로 방향 수정</p> <p>YES24 e-book 페이지</p> <p>탐색</p> <p>주 알고리즘 결정</p> <p>크롤링 시작</p>	<p>크롤링 팀 :</p> <p>1차 크롤링 (도서 100개)</p> <p>2차 크롤링 (전체 도서)</p> <p>전처리 팀 :</p> <p>데이터 전처리 개별 학습</p>	<p>크롤링 팀 :</p> <p>데이터 크롤링 개별 복습</p> <p>PPT 제작 시작</p> <p>전처리 팀 :</p> <p>데이터 전처리 수행</p> <p>중간 발표</p>
6.20 (월)	6.21 (화)	6.22 (수)	6.23 (목)	6.24 (금)
<p>전처리 된 데이터</p> <p>형태소 분석</p> <p>(Okt 사용)</p> <p>알고리즘 코드 작성</p> <p>(TF-IDF)</p> <p>유사도 확인</p> <p>웹앱 초안 작성</p>	<p>전처리 된 데이터</p> <p>형태소 분석</p> <p>(Mecab 사용)</p> <p>알고리즘 코드 작성</p> <p>(Doc2Vec)</p> <p>모델링 및 성능 확인</p> <p>웹앱 수정</p>	<p>형태소 분석기 회의</p> <p>임베딩 기법 회의</p> <p>코드 재검토</p> <p>프로젝트 마무리</p> <p>발표 내용 회의</p>	<p>PPT 마무리</p> <p>본인의 맡은 역할 시연 및</p> <p>코드 중심의 질의 응답</p> <p>상호간의 노하우 공유와</p> <p>배움의 시간</p>	<p>최종 발표</p>

목차 소개



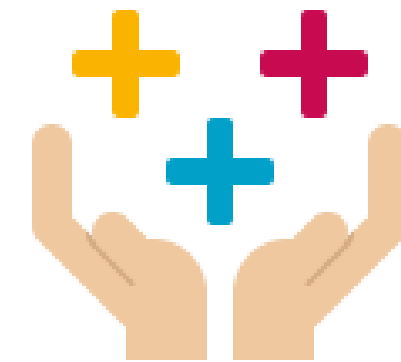
01. 프로젝트 배경

- 기획 배경 및 목적
- E-book 기존 서비스와의 비교
- 우리 서비스만의 차별화 핵심 전략



02. 프로젝트 과정 및 결과

- 서비스 도식화
- 데이터 수집 및 전처리
- 추천 시스템 모델 수행



03. 기대효과

- 기대효과
- 개선사항
- 활용



BOODLE
SERVICE

1. 프로젝트 배경

2. 프로젝트 과정 및 결과

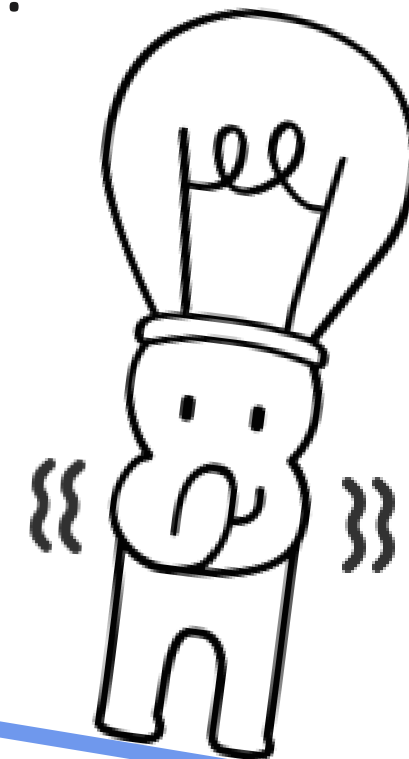
3. 기대효과

1



2

'태그'나 '분류'로는
좀처럼 원하는 게
안 나오네...



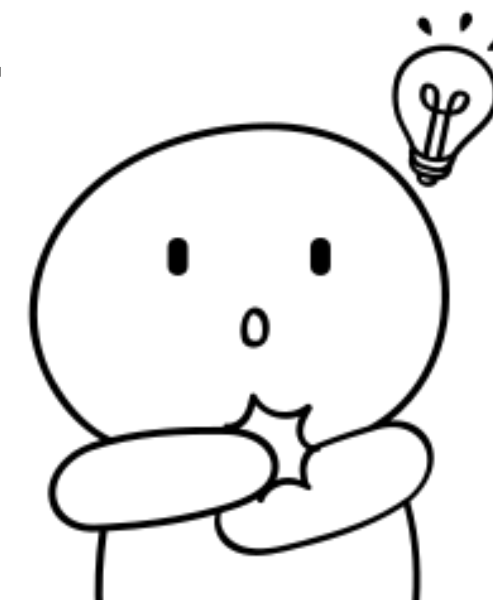
3

책 **내용**을 토대로
유사도를
책정해보면 어떨까?



4

책 **미리보기**를
토대로 한 번
찾아보자!





BOODLE
SERVICE

1. 프로젝트 배경

문제상황

불분명한 '태그'와
'도서추천'으로
원하는 도서를
찾기 어려움.



이 책을 구입하신 분들이 **함께 산 책**

이 책을 구입하신 분들이 **많이 산 책**

1/4



이방인 - 세계문학전집 266
알베르 카뮈 저 | 민음사
★★★★☆ 16건
6,300원



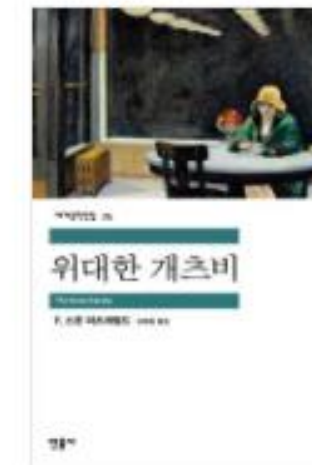
인간 실격 - 세계문학전집 103
다자이 오사무 저/김춘...
★★★★☆ 8건
5,600원



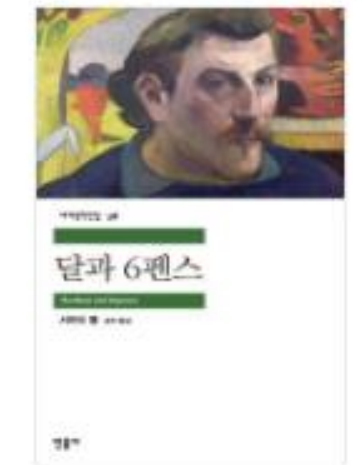
오만과 편견 - 세계문학전집 088
제인 오스틴 저/윤지관...
★★★★☆ 15건
7,700원



1984 - 세계문학전집 077
조지 오웰 저/정희성 역...
★★★★☆ 14건
6,650원



위대한 개츠비 - 세계문학전집 075
F.스콧 피츠제럴드 저/...
★★★★☆ 3건
5,600원



달과 6펜스 - 세계문학전집 038
서머셋 모 저/송무 역 |...
★★★★☆ 16건
6,300원

→ 구매내역에 기반한 추천



BOODLE
SERVICE

1. 프로젝트 배경

문제상황

불분명한 ‘태그’와
‘도서추천’으로
원하는 도서를
찾기 어려움.



이 상품의 태그

#2PM준호의책장 #알쓸범잡 #20세기고전 #방탄책방 #방탄RM이추천한책 #북클럽러버의선택
#안사면0원사면영원 #요즘책방:책읽어드립니다 #아이돌셀러 #방탄소년단이읽은책 #성장 #빨간책방



설민석의 조선왕조실록
설민석 저 | 세계사
0원 (0% 할인)



팩트풀니스
한스 로슬링, 올라 로슬링...
11,880원 (10% 할인)



유혹의 기술
로버트 그린 저/강미경 ...
14,500원 (0% 할인)



뇌, 욕망의 비밀을 풀다
한스-게오르크 호이젤 ...
12,600원 (0% 할인)

➔ 책의 **내용**에 기반한 태그 및 추천이 아닌,
수식어와 **평가**에 기반한 태그

- 국내도서
 - + 가정 살림
 - + 건강 취미
 - + 경제 경영
 - + 국어 외국어 사전
 - + 대학교재
 - + 만화/라이트노벨
 - + 사회 정치

- 소설/시/희곡
 - + 장르소설
 - + 테마소설
 - + 고전문학
 - + 시/희곡
 - + 비평/창작/이론
 - + 한국소설
 - + 영미소설
 - + 일본소설
 - 중국소설
 - 프랑스소설
 - 독일소설
 - 러시아소설
 - + 스페인/중남미소설
 - 북유럽소설

➔ 광범위한 책 분류



BOODLE
SERVICE

1. 프로젝트 배경

소비자분석

자신이 감명 깊게
읽은 책과 유사한
내용의 책을
원하는 경향이
있음.

N | 이상한 나라의 앨리스 같은 책



2007.11.20.

관련 + 교훈

Q 이상한 나라의 앨리스 책과 비슷한 책 찾아요!!! 이상한 나라의 앨리스같은 판타지? 그런 신비한 책을 좋아하는데요 찾아봐도 그런책은 안보이더라구요ㅠㅜ ...

A 만물상점

1:1

저도 판타지 장르를 정말 좋아해요!! 추천해드릴게요. 1. 달러구트 꿈 백화점 <달러구트 꿈 백화점>은 '무의식에서만 존재하는 꿈을 정말 사고 팔 수 있을까?'라는 기발한 질문에 답을 찾아가며, 꿈을 만드...
2021.01.26.

Q 책 추천(판타지?모험?) 음 약간 뭔가 도깨비같은 그런 신비한 세상에서 주인공이 모험을 펼치는? 이상한 나라의 앨리스같은 느낌.. 책 추천해주세요용

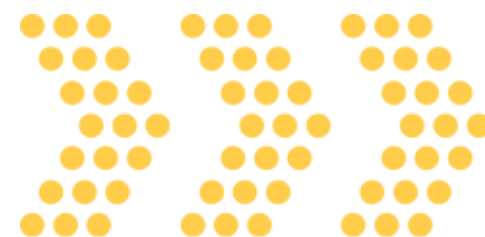
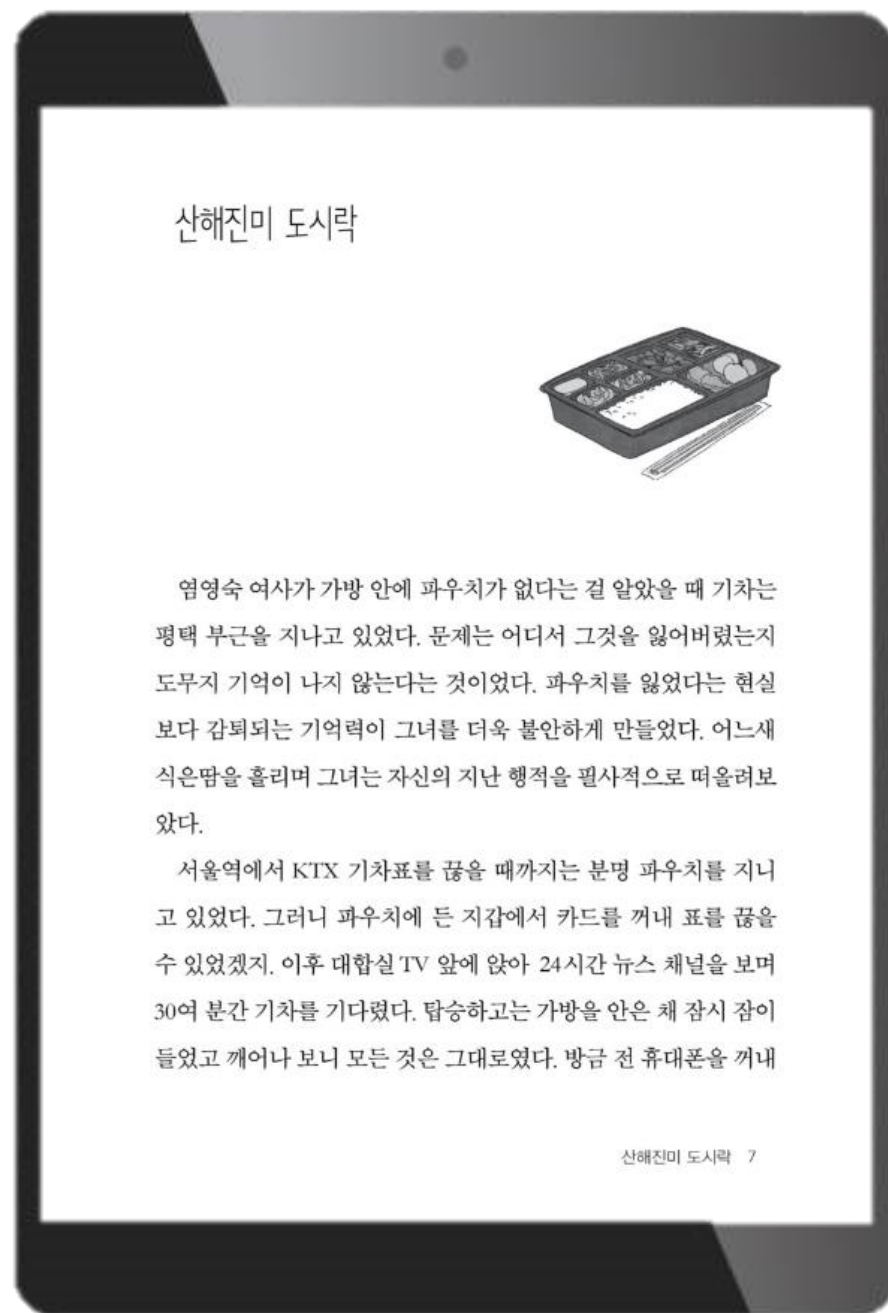
2021.03.03.

관련 + 신비 + 판타지

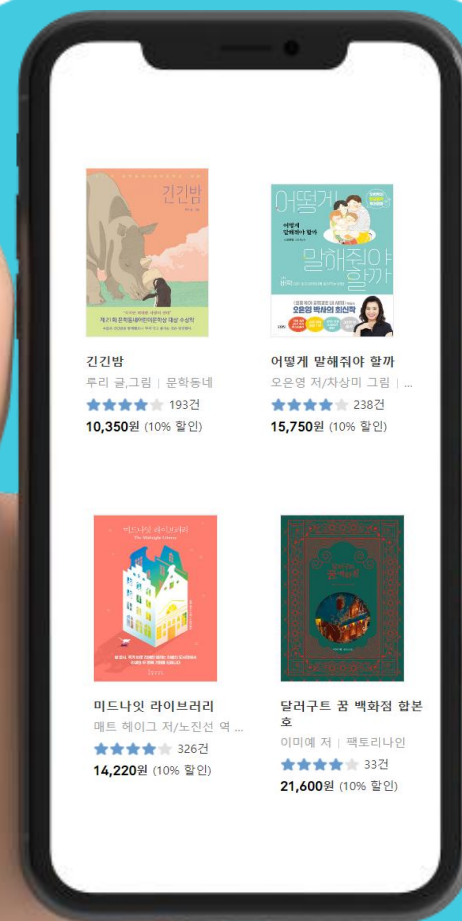


BOODLE
SERVICE

1. 프로젝트 배경



미리보기를 통한
책 내용 유사도 검증으로
원하는 작품을 추천



차별화 핵심 전략

미리보기

책 내용 유사도 검증



BOODLE
SERVICE

2. 프로젝트 과정 및 결과

개발환경



Streamlit





BOODLE
SERVICE

서비스 도식화

2. 프로젝트 과정 및 결과

텍스트 데이터 수집

YES24 ebook 소설 카테고리
베스트 셀러 + 월별 베스트 셀러
제목, 미리보기 텍스트 데이터 수집
Json 으로 저장

데이터셋 전처리

- 결측치 제거
- 정규 표현식
- 1000자 미만 데이터 제거

형태소 분석기

Okt / Mecab

한국어 임베딩

TF-IDF : 빈도수 기반 단어-문서 행렬
Doc2Vec : 문장 수준 임베딩

Cosine similarity
기반 유사도 평가

도서 제목을 입력하면
입력한 도서와 유사한
상위 10개 도서의 제목이
추천되는 알고리즘 구현

Streamlit을
통한 웹 구현

연산속도

Okt ≪ Mecab

도서 검색

TF-IDF
Doc2Vec

코사인 유사도
계산

유사도 상위 10개
도서 추천



BOODLE
SERVICE

서비스 도식화

2. 프로젝트 과정 및 결과

텍스트 데이터 수집

YES24 ebook 소설 카테고리
베스트 셀러 + 월별 베스트 셀러
제목, 미리보기 텍스트 데이터 수집
Json 으로 저장

데이터셋 전처리

- 결측치 제거
- 정규 표현식
- 1000자 미만 데이터 제거

형태소 분석기

Mecab

한국어 임베딩

TF-IDF : 빈도수 기반 단어-문서 행렬
Doc2Vec : 문장 수준 임베딩

Cosine similarity
기반 유사도 평가

도서 제목을 입력하면
입력한 도서와 유사한
상위 10개 도서의 제목이
추천되는 알고리즘 구현

Streamlit을
통한 웹 구현

연산속도

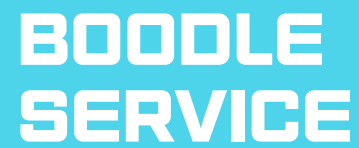
Okt <<< Mecab

도서 검색

TF-IDF
Doc2Vec

코사인 유사도
계산

유사도 상위 10개
도서 추천



2. 프로젝트 과정 및 결과

이|계|로 통하는 차원 문

"호으."

강성호는 기지개를 켜며 일어났다. 오늘도 지겨운 하루가 시작된다. 먹고살기 위해 꼭 해야 하는, 그러나 잘 되지는 않는 그런 일.

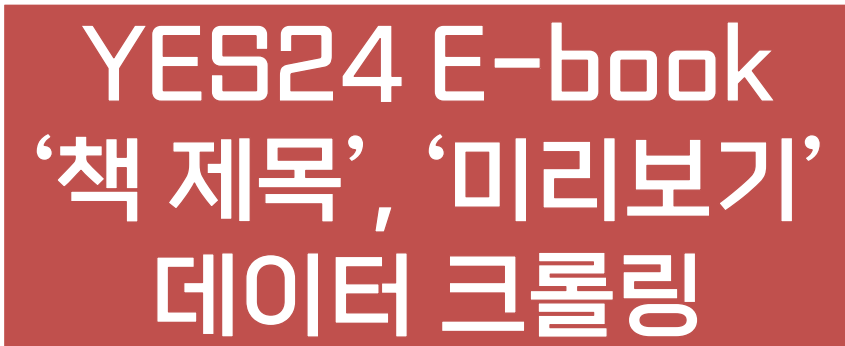
"아이고, 두야..."

이불을 정돈하고 단칸방 바닥에 떨어진 머리카락을 물티슈로 대충 치운다. 그의 방은 가게와 붙어 있다. 무슨 가게냐 하면, 분식집이다.

부산 동아여중, 동아여고, 동아대 앞이라는 환상의 입지를 가졌지만 손님은 시원치 않다. 그의 가게가 학교 길 건너편에 있기 때문이다. 튀김이나 김밥 따위를 길을 건너면서까지 사 먹는 사람은 많지 않다. 학교 앞에도 분식집 두어 군데가 있다면 더더욱.

"하아암."

하품을 하며 얼굴을 대강 씻는다.





BOODLE
SERVICE

데이터 수집

YES24 E-book
‘책 제목’, ‘미리보기’
데이터 크롤링

2. 프로젝트 과정 및 결과

1 페이지 크롤링중...

100%

20/20 [00:58<00:00, 2.61s/it]

미리보기 없음: [100%페이백][대여] 순백의 비명
미리보기 없음: 가꾸는 이의 즐거움
미리보기 없음: 그램머블 제로 Grammable zero
미리보기 없음: 아몬드
미리보기 없음: 동급생
미리보기 없음: 소년이 온다
2 페이지 크롤링중...

100%

20/20 [01:07<00:00, 3.37s/it]

미리보기 없음: 구의 증명 [할인]
미리보기 없음: 완전한 행복
미리보기 없음: 1984 (한글판) 019
미리보기 없음: 우리가 빛의 속도로 갈 수 없다면
3 페이지 크롤링중...

	title	preview
0	[100%페이백][대여] 용서받지 못한 밤	None
1	기적의 분식집	1◆\n이게로 통하는 차원 문\n\n"흐으욱."\n강성호는 기지개를 켜며 일어났다....
2	유리의 살의	\n\n* 이 작품은 픽션입니다.\n실제 인물, 단체와는 일절 관계가 없습니다.\n...
3	작별인사	\n\n\n머지않아 너는 모든 것을 잊게 될 것이고,\n머지않아 모두가 너를 잊게 ...
4	가꾸는 이의 즐거움	None

...

2. 프로젝트 과정 및 결과

결측치 제거

불용어 처리

연속 공백 제거

데이터 전처리

	title	preview
0	불편한 편의점	염영숙 여사가 가방 안에 파우치가 없다는 걸 알았을 때 기차는 평택 부근을 지나고 ...
1	달려구트 꿈 백화점 : 잠들어야만 입장 가능합니다	목차 작가의 말 프롤로그 번째 제자의 유서 깊은 가게 주문하신 꿈은 매진입니다 한밤...
2	미드나잇 라이브러리	지은이 매트 헤이그 강렬한 존재감과 위대한 재능을 가진 소설가 로 평가받는 영국의 ...
3	기적의 분식집	이계로 통하는 차원 문 흐으옥 강성호는 기지개를 켜며 일어났다 오늘도 지겨운 하루가...
4	유리의 살의	이 작품은 픽션입니다 실제 인물 단체와는 일절 관계가 없습니다 이 책은 주 한국저작...
...
1280	문어의 아홉 번째 다리	디르크 로스만지음 서경홍옮김 이 책의 한국어판 저작권은 모모 에이전시를 통해사와의...
1281	백은의 잭	히가시노 게이고 일본 최고의 베스트셀러 작가 년 오사카 출생 오사카 부립대학 졸업 ...
1282	빅티켓	테릴 리 랭크포드를 위하여 우리는 우리의 삶을 우화처럼 기억한다 작자 불명 본 도서...
1283	나쁜 토끼	와카타케 나나미 일본 코지 미스터리의 여왕 년 도쿄에서 태어나 릿쿄 대학 문학부를 ...
1284	욕조가 놓인 방	이승우 년 전남 장흥에서 태어났다 년 한국문학 신인상을 받으며 작품 활동을 시작했다...

1285 rows × 2 columns

전처리 후
총 1285개 도서에 대한 데이터 확보

2. 프로젝트 과정 및 결과

형태소 분석

Ex. 불편한 편의점(index: 0) 토큰화 결과

```
tagged_corpus_list[0]
```

```
TaggedDocument(words=['염', '영숙', '여사', '가', '가방', '안', '에', '파우', '치가', '없', '다는', '걸', '알', '았', '을', '때', '기  
차', '는', '평택', '부근', '을', '지나', '고', '있', '었', '다', '문제', '는', '어디', '서', '그것', '을', '잃어버렸', '는지', '도무  
지', '기억', '이', '나', '지', '않', '는다는', '것', '이', '었', '다', '파우', '치', '를', '잃', '었', '다는', '현실', '보다', '감퇴',  
'되', '는', '기억력', '이', '그녀', '를', '더욱', '불안', '하', '게', '만들', '었', '다', '어느새', '식은땀', '을', '흘리', '며', '그  
녀', '는', '자신', '의', '지난', '행적', '을', '필사', '적', '으로', '떠올려', '보', '았', '다', '서울', '역', '에서', '기차표', '를',  
'끊', '을', '때', '까지', '는', '분명', '파우', '치', '를', '지니', '고', '있', '었', '다', '그러', '니', '파우', '치', '에', '든',  
'지갑', '에서', '카드', '를', '꺼내', '표', '를', '끊', '을', '수', '있', '었', '겟', '지', '이후', '대합실', '앞', '에', '았', '아',  
'시간', '뉴스', '채널', '을', '보', '며', '여', '분간', '기차', '를', '기다렸', '다', '탑승', '하', '고', '는', '가방', '을', '안',  
'은', '채', '잠시', '잠', '이', '들', '었', '고', '깨어나', '보', '니', '모든', '것', '은', '그대로', '였', '다', '방금', '전', '휴대  
폰', '을', '꺼내', '려고', '가방', '을', '열', '었', '을', '때', '안', '에', '있', '어야', '할', '파우', '치가', '없', '는', '것',  
'에', '소스라치', '게', '놀랐', '을', '뿐', '이', '었', '다', '지갑', '통장', '수첩', '등', '자신', '의', '가장', '중요', '한', '것',  
'들', '이', '담긴', '파우', '치가', '없', '다는', '것', '에', '그녀', '는', '숨', '이', '막힐', '지경', '이', '었', '다', '염', '여  
사', '는', '자신', '이', '타', '고', '있', '는', '기차', '의', '속력', '에', '뒤지', '지', '않', '게', '두뇌', '를', '가동', '해야',  
'했', '다', '차창', '밖', '으로', '빠르', '게', '지나가', '버리', '는', '풍경', '을', '되돌리', '기', '라도', '할', '듯', '기억',  
'을', '리', '와인', '드', '했', '다', '혼잣말', '에다', '다리', '를', '떨', '며', '골몰', '해', '있', '는', '그녀', '의', '행동',  
'에', '옆', '자리', '중년', '사내', '가', '헛기침', '을', '했', '다', '그녀', '의', '몰입', '을', '방해', '한', '건', '옆', '자리',  
'사내', '의', '헛', '기침', '이', '아니', '라', '가방', '안', '에서', '울린', '휴대폰', '착신', '음악', '이', '었', '다', '아바',  
'의', '노래', '였', '는데', '곡명', '이', '떠오르', '지', '않', '았', '다', '치키', '티', '타', '였', '나', '댄싱', '퀸', '이', '었',  
'나', '아이', '고', '준', '희', '야', '할머니', '가', '진짜', '치매', '오', '려나', '보다', '염', '여사', '는', '떨리', '는', '손',  
'으로', '가방', '안', '휴대폰', '을', '꺼내', '고', '나', '서', '야', '곡명', '이', '땡큐', '포', '더', '뮤직', '이', '라는', '것',  
'을', '기억', '해냈', '다', '동시', '에', '낯선', '번호', '로', '전화', '가', '오', '는', '것', '을', '확인', '했', '다', '그녀',  
'는', '심호흡', '을', '한', '뒤', '전화', '를', '받', '았', '다', '여보세요', '상대방', '은', '대답', '이', '없', '었', '다', '다만',  
'주변', '의', '소음', '이', '공공장소', '임', '을', '짐작', '케', '했', '다', '누구', '세요', '염', '영숙', '이', '에요', '사람',  
'의', '목소리', '라기', '엔', '너무나', '거칠', '고', '불분명', '했', '다', '마치', '겨울잠', '을', '끝낸', '곰', '이', '동굴', '에  
서', '나와', '처음', '입', '을', '열', '면', '나올', '법', '한', '소리', '였', '다', '예', '그런데', '요', '지갑', '요', '맞', '아요',
```




BOODLE
SERVICE

추천 시스템 모델

Tf-idf 사용

단어의 빈도와 역 문서 빈도를 사용하여 DTM(문서단어행렬) 내에 각 단어들마다 중요한 정도를 가중치를 부여하는 Bag-of-Words 기반의 방식

2. 프로젝트 과정 및 결과

학습의의

Tf-idf Doc2vec

두 알고리즘
사용하여 출력값
비교, 각각의
장단점 파악

TF-IDF 임베딩

```
In [10]: from sklearn.feature_extraction.text import TfidfVectorizer

tfidf = TfidfVectorizer().fit(analyzed_data)
tfidf_matrix = tfidf.transform(analyzed_data)
```

코사인 유사도

```
In [11]: from sklearn.metrics.pairwise import cosine_similarity

cosine_sim = cosine_similarity(tfidf_matrix, tfidf_matrix)
```

```
In [12]: title_to_index = dict(zip(df['title'], df.index))
```

Doc2vec 사용

Word2Vec의 확장으로 문장, 문단, 문서의 단위로 vector를 생성하는 방식

Mecab 임베딩

```
In [5]: from konlpy.tag import Mecab
        from tqdm.auto import tqdm
        from gensim.models.doc2vec import TaggedDocument

        mecab = Mecab('C:\#mecab\#mecab-ko-dic')

        tagged_corpus_list = []

        for index, row in tqdm(df.iterrows(), total=len(df)):
            text = row['preview']
            tag = row['title']
            tagged_corpus_list.append(TaggedDocument(tags=[tag], words=mecab.morphs(text)))
```

100% 1308/1308 [00:46<00:00, 34.33it/s]

모델링

```
In [7]: from gensim.models import doc2vec

        model = doc2vec.Doc2Vec(vector_size=300, alpha=0.025, min_alpha=0.025, workers=8, window=8)

        # Vocabulary 빌드
        model.build_vocab(tagged_corpus_list)

        # Doc2Vec 학습
        model.train(tagged_corpus_list, total_examples=model.corpus_count, epochs=50)

        # 모델 저장
        model.save('dart.doc2vec')
```

2. 프로젝트 과정 및 결과

학습의의

Tf-idf
Doc2vec

두 알고리즘
사용하여 출력값
비교, 각각의
장단점 파악



BOODLE
SERVICE

2. 프로젝트 과정 및 결과

학습의의

Tf-idf
Doc2vec

두 알고리즘
사용하여 출력값
비교, 각각의
장단점 파악

결과 비교

추천 시스템 모델

[TF-IDF] 이상한 나라의 앨리스 유사도 확인

```
In [21]: get_recommendations('이상한 나라의 앨리스')
```

Out[21]:

	제목	score
515	초판본 이상한 나라의 앨리스	93.8
728	이상한 나라의 앨리스 - 허밍버드 클래식 01	87.8
157	거울나라의 앨리스 (한글판) 102	77.7
1141	스틸앨리스	72.8
557	초판본 거울나라의 앨리스	72.7
433	앨리스 죽이기	38.6
1253	티핑 더 벨벳	27.3
736	매혹당한 사람들	23.4
872	오후도 서점 이야기	19.1
786	체체파리의 비법	13.5

Out[20]:

[D2V] 이상한 나라의 앨리스 유사도 확인

```
In [20]: similar_doc = model.docvecs.most_similar('이상한 나라의 앨리스')
pd.DataFrame(similar_doc, columns=["제목", "유사도"])

C:\Users\kissd\AppData\Local\Temp\ipykernel_6564\3453630975.py:1: DeprecationWarning:
  similar_doc = model.docvecs.most_similar('이상한 나라의 앨리스')
```

	제목	유사도
0	초판본 이상한 나라의 앨리스	0.890372
1	이상한 나라의 앨리스 - 허밍버드 클래식 01	0.673004
2	거울나라의 앨리스 (한글판) 102	0.455951
3	초판본 거울나라의 앨리스	0.436793
4	코끼리의 마음	0.410410
5	곰돌이 푸 1 (한글판) 113	0.405198
6	피터 래빗 이야기 2 (한글판) 092	0.386643
7	이솝 이야기 1 (한글판) 054	0.384174
8	곰돌이 푸 이야기 전집	0.375442
9	피터 래빗 이야기 3 (한글판) 093	0.367471



BOODLE
SERVICE

2. 프로젝트 과정 및 결과

학습의의

Tf-idf
Doc2vec

두 알고리즘
사용하여 출력값
비교, 각각의
장단점 파악

결과 비교

추천 시스템 모델

[TF-IDF] 백마산장 살인사건 유사도 확인

```
In [20]: get_recommendations('백마산장 살인사건')
```

Out[20]:

	제목	score
124	하쿠바산장 살인사건	99.7
250	백야행 2	49.1
36	노르웨이의 숲	43.4
1086	백광	28.3
87	목소리를 드릴게요	9.6
890	고 온(GO ON) 1	8.6
1184	산책자	8.4
834	절망의 구	8.1
1199	밤은 짧아 걸어 아가씨야	8.0
180	레베카 (개정판)	7.5

Out [4]:

	제목	유사도
0	하쿠바산장 살인사건	0.945348
1	속명	0.403551
2	질풍론도	0.397875
3	브루투스의 심장	0.388415
4	몽환화	0.382844
5	추리소설가의 살인사건	0.381695
6	백야행 2	0.367586
7	다임 아이	0.367503
8	눈보라 체이스	0.358259
9	작열	0.353841

[D2V] 백마산장 살인사건 유사도 확인

```
In [19]: similar_doc = model.docvecs.most_similar('백마산장 살인사건')
pd.DataFrame(similar_doc, columns=["제목", "유사도"])

C:\Users\kissd\AppData\Local\Temp\ipykernel_6564\2075191916.py:1: DeprecationWarning
rty has been renamed `dv`).
  similar_doc = model.docvecs.most_similar('백마산장 살인사건')
```


추천 시스템 모델

장단점 비교

	Tf-idf	Doc2vec
장점	비교적 간단하고 빈도 기반 방법이라 표현 모델이 학습될 필요가 없음	타겟 단어의 정보와 타겟 단어 주변 단어들을 이용하여 신경망 학습
단점	단어와 단어 사이의 개념적 유사도를 고려하지 못함	Word2Vec을 문서 단위로 확장하여 문서들 사이의 의미적 관계 보존

2. 프로젝트 과정 및 결과

학습의의

Tf-idf Doc2vec

두 알고리즘
사용하여 출력값
비교, 각각의
장단점 파악

성능적 측면에서 비교

문서내에 어떤 단어가
많이 쓰였는가를 기반으로
미리보기 유사성을 판단



문서내에 어떤 단어가
같이 쓰였는가를 기반으로
미리보기 유사성을 판단

더욱 비슷한 문맥의 도서 추천 가능

미리보기 기반 소설 추천 서비스

좋아하는 소설 제목을 입력하시오.

백마산장 살인사건

검색 대상: 백마산장 살인사건

TF-IDF:

	제목	score
0	하쿠바산장 살인사건	99.7000
1	백야행 2	49.1000
2	노르웨이의 숲	43.4000
3	백광	28.3000
4	목소리를 드릴게요	9.6000
5	고 온(GO ON) 1	8.6000
6	산책자	8.4000
7	절망의 구	8.1000
8	밤은 짧아 걸어 아가씨야	8.0000
9	레베카 (개정판)	7.5000

Doc2Vec:

	제목	유사도
0	하쿠바산장 살인사건	94.5000
1	숙명	40.4000
2	질풍론도	39.8000
3	브루투스의 심장	38.8000
4	몽환화	38.3000
5	추리소설가의 살인사건	38.2000
6	백야행 2	36.8000
7	다잉 아이	36.8000
8	눈보라 체이스	35.8000
9	작열	35.4000

기억이 잘 나지 않는 책 입력하시오.

데미

데미

▼ [

- 0 : "데미안 (한글판) 010"
- 1 : "데미안 - 세계문학전집 044"
- 2 : "데미안 (한글+영문)"
- 3 : "데미안 초판본"
- 4 : "데미안"
- 5 : "데미안"
- 6 : "데미안 - 더모던타임즈 05"

]

2. 프로젝트 과정 및 결과

도서를 입력함으로써 해당 도서의 미리보기와
문서 유사도가 비슷한 도서 제목을 유사도 스코어 값과 함께 추천



BOODLE
SERVICE

3. 기대효과

기대효과



검색의 즉각성에 의해
도서를 찾는 시간/비용 감소

지식IN이나 다른 사이트를 뒤적이지 않아도
실시간으로 도서를 추천 받을 수 있음.



내용 기반 특화 검색엔진

타 사이트의 도서 추천과 달리
미리보기 내용을 기반으로 추천하기에
독자가 원하는 책에 더 가까이
접근할 수 있음.



BOODLE
SERVICE

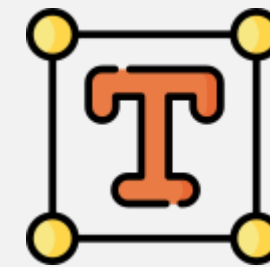
3. 기대효과

한계 및 개선사항

한계



미리보기의 한계



텍스트 추출의 한계



데이터 양의 한계



“동일한 제목의 책은 하나만 추천하는 전처리를 추가하면 중복된 책 없이 추천이 가능할 것입니다.”

ex) 이상한 나라의 앨리스 초판본, 이상한 나라의 앨리스 허밍버드 클래식 등을 [이상한 나라의 앨리스] 하나로 통일

개선사항

“책에 대한 키워드를 추출하여 사용자에게 알려주는 서비스를 추가해도 좋을 것 같아요”





BOODLE
SERVICE

세미 프로젝트

잘 만들었다

팀 프로젝트 의의

POINT 1

프로젝트
완성



같이 해보면

별거 아니다..!!!



팀 프로젝트 의의

POINT 1

프로젝트
완성

★ POINT 2 ★

팀워크

‘팀원 모두가 배워가는 프로젝트’

✓ 멘토(A)와 멘티(B)의 구성

✓ 본인이 맡은 업무 노하우 공유 시간

→ I learn from You, You learn from me.

→ 최종 프로젝트 시작 전 자신감 상승의 기회로 삼자



BOODLE
SERVICE

감사합니다



BOODLE
SERVICE

Q&A