

전국 명소 총 망라! 당신의 휴가로 여기 어때요?

#BERT #Retrieval #TF-IDF #웹API

NLP-11
(AI-ESG)

목차

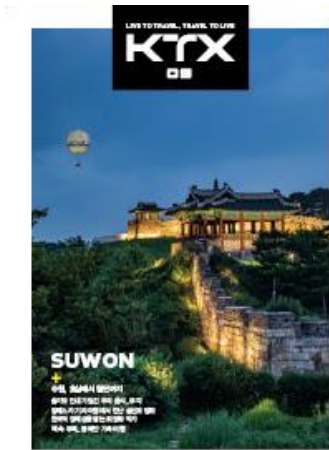
1. 서비스 소개
2. 구현 방법
3. 개선 사항

1. 서비스 소개

여행 갈 곳, 놀러 갈 곳, 어떻게 결정될까?

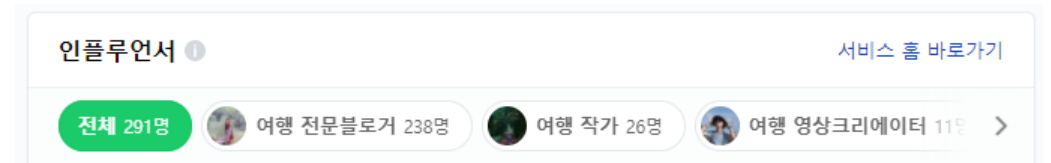
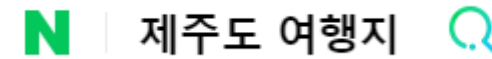
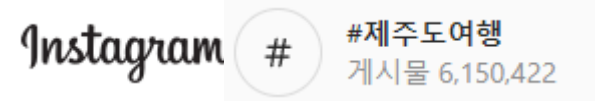
- 우연히 발견

지인 추천, 미디어

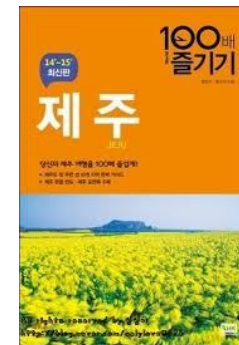


- 직접 찾기

특정 지역 명소의 후기 간접 체험

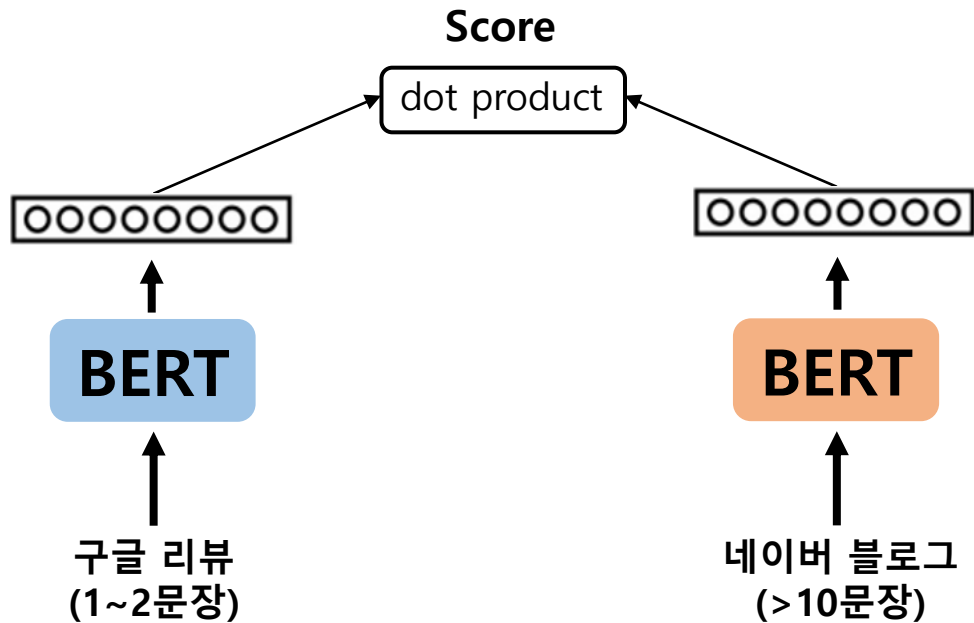


다른 방법은 없을까?



다른 방법은 없을까?


블로그, 리뷰의 글을 활용, 입력 문장에 대응되는 여행지 구축




- **Score** 최대화 훈련
- 사용자의 입력 문장에 대응하는 장소를 탐색하는 모델
장소 목록, 정보 : tour api
훈련 pair : 구글 리뷰, 네이버 블로그 (크롤링)

Dataset

 **TourAPI^{3.0}** : 장소 목록

 **blog** : 블로그 문장 (train, inference)

 **Google** : 리뷰 문장 (train)

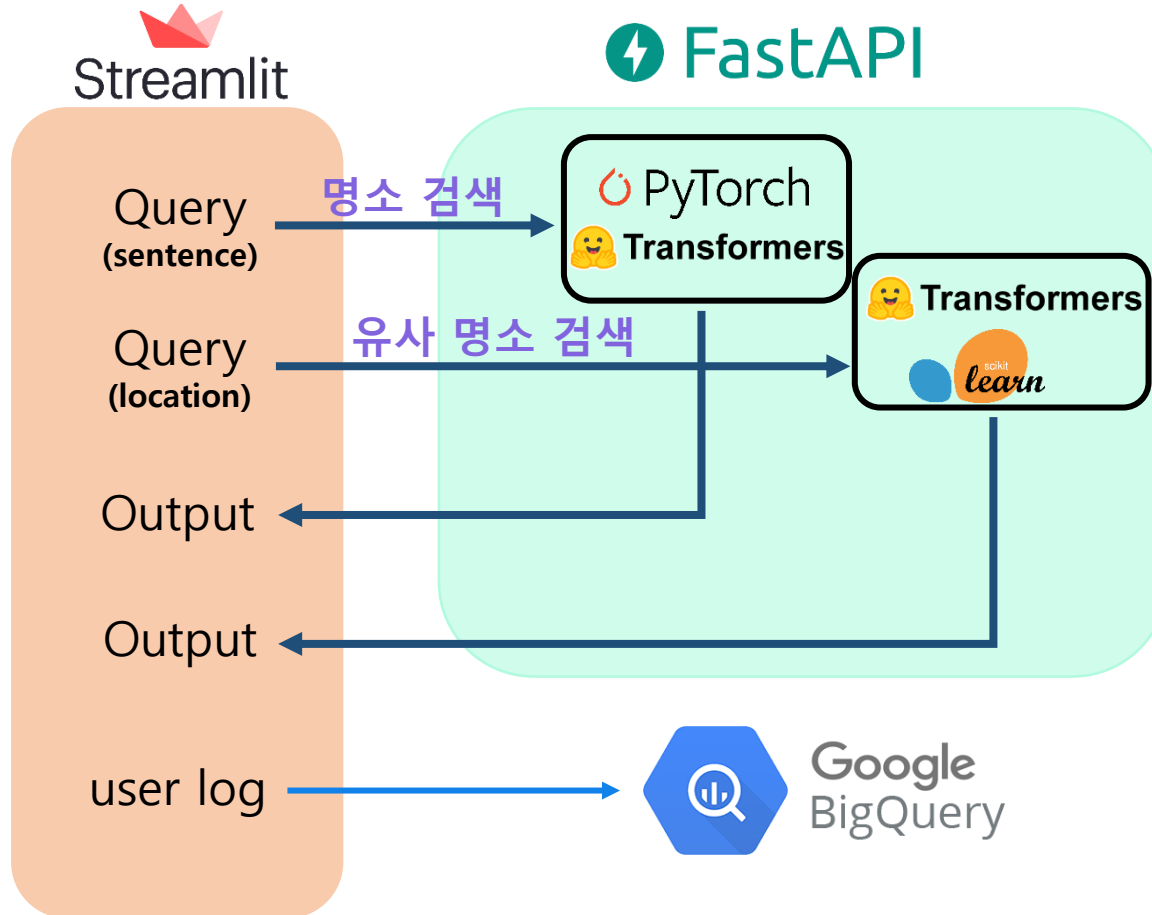
서비스 데모



저기 어때

2. 구현 방법

Service Flow



Latency

명소 검색 :

- ~560 ms (GPU)
- ~743 ms (CPU)

유사 명소 검색 :

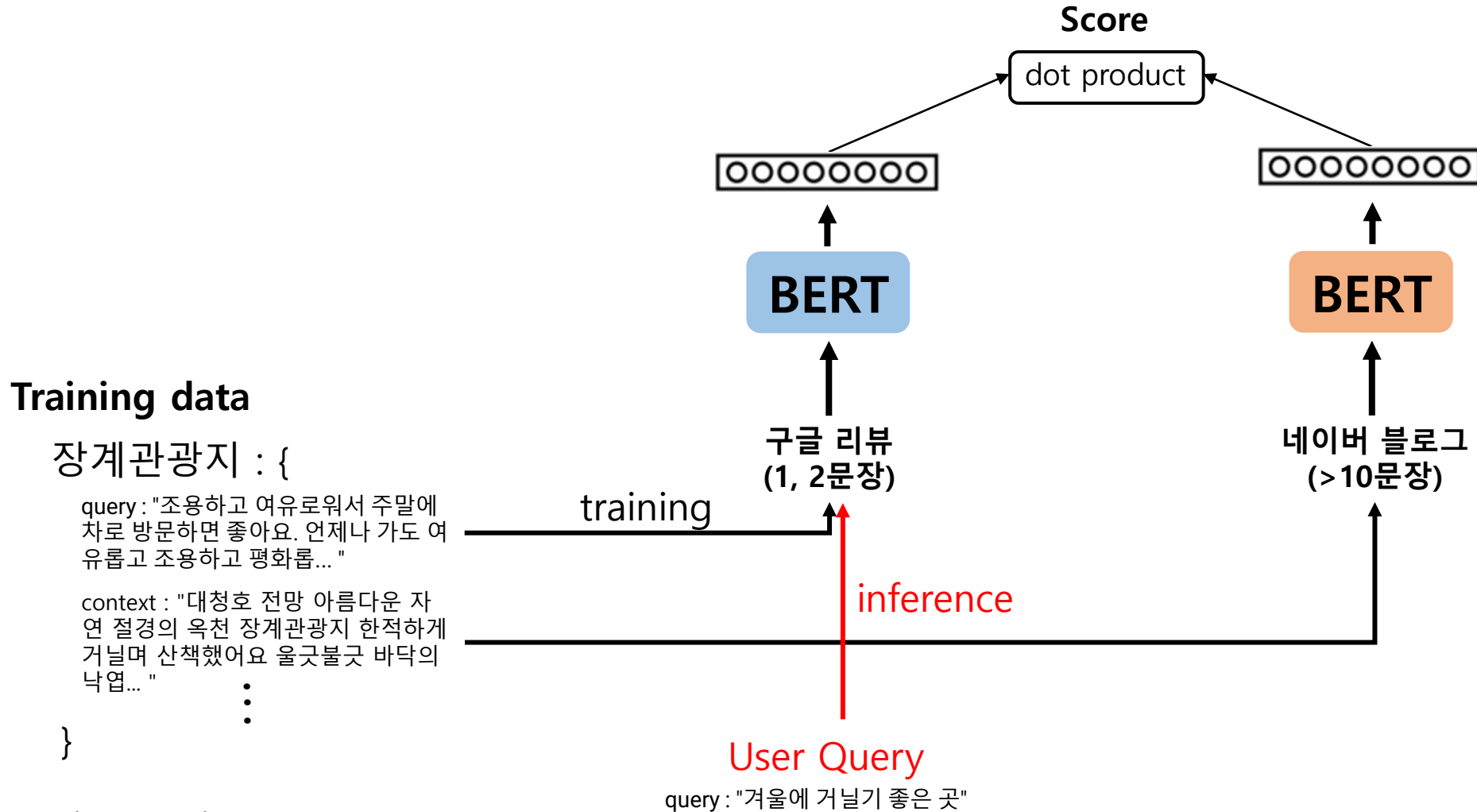
- ~980 ms (CPU)

데이터셋 수집, 구축



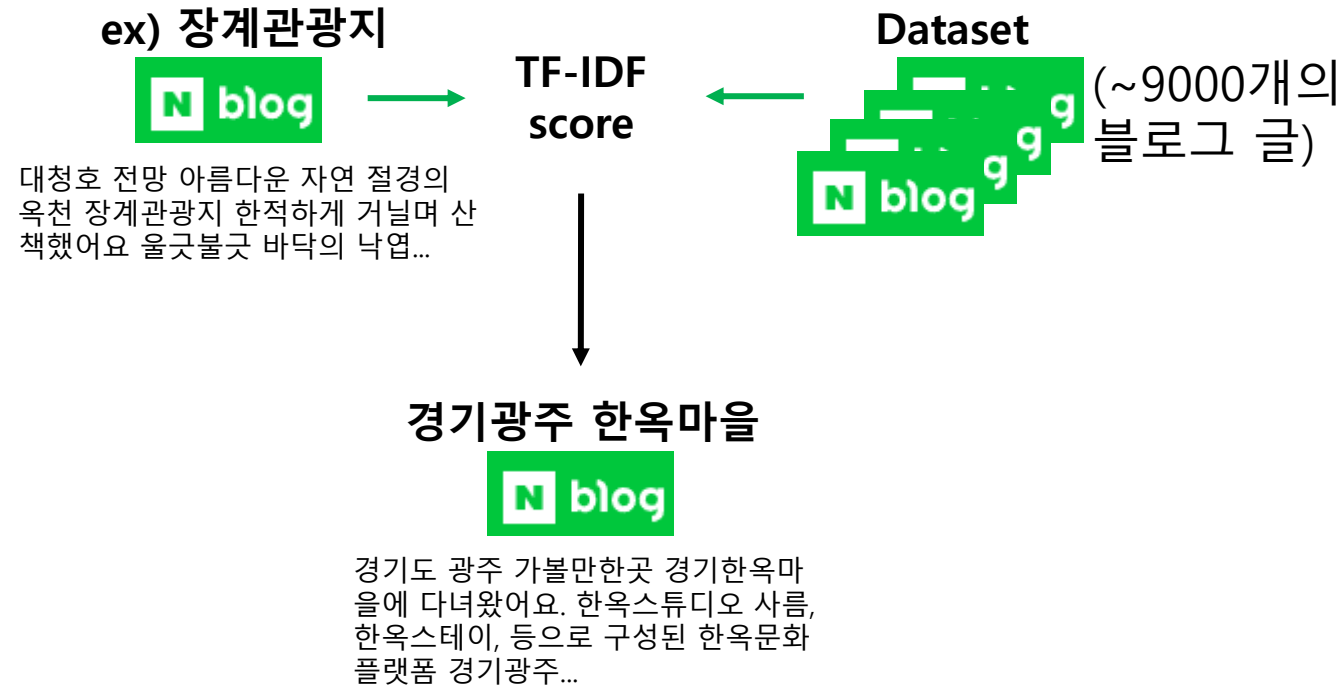
명소 검색 - 모델

리뷰, 블로그 pair의 score 최대화 훈련



유사 명소 검색 - 모델

TF-IDF score 기준 top-k 블로그 탐색, 해당 장소 추천



3. 문제점 및 개선사항

문제점

유사명소 검색

반복적으로 등장하는 장소 존재 (ex:사계해변)
여행지와 무관한 내용 때문으로 추정

데이터셋

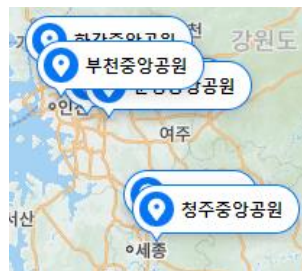
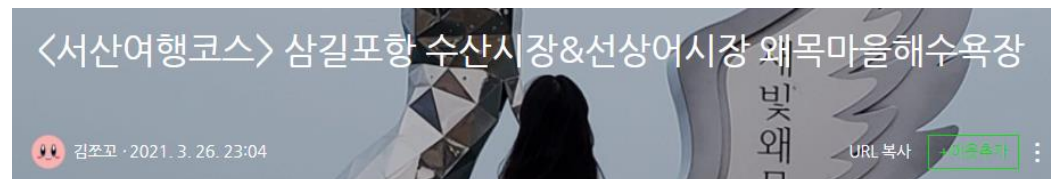
일기형 블로그 글

여행지와 관련 없는 내용 제거 필요

다른 지역, 동일한 장소명

여행지와 관련 없는 내용 제거 필요

수작업에 pair당 1~2분 소요



개선사항

최적화

- sort 알고리즘 개선
- 유사명소검색 - 전처리된 데이터 저장

데이터가 더 많아지면?

- torch.matmul
- 병렬 sort

완성도 높이기

- 추천된 장소 근처 음식점 제공 (인스타그램)
- API에 없는 사진 추가
- 유저 로그 훈련에 반영
 - 유저 query 데이터셋에 추가

해보고 싶은 것

- 문장 + 이미지 검색

Q&A