한국어 임베딩

- 데이터를 토큰화 처리한 뒤 버트 모델에 적용시켜 훈련.
- 은닉 상태의 정보는 다음과 같음.
  - 계층 개수 : 13
  - 배치 개수 : 1
  - 토큰 개수 : 33
  - 은닉층의 유닛 개수 : 768
- 네 개의 계층을 연결하여 단어 벡터를 생성
- 각 토큰의 두 번째에서 마지막 은닉 계층을 평균화하는 것으로 전체 문장에 대한 단일 벡터를 구함.
- 코사인 유사도를 구해 단어 간 유사도를 확인.