

3주차 2조

팀원: 황동욱, 권도혁, 송성근

```
lookup.KeyValue  
f.constant(['em  
=tf.constant([0  
ce = tf.lookup.StaticV  
init,  
num_oov_buckets=5)
```

```
lookup.StaticVocabular  
initializer,  
num_oov_buckets,  
lookup_key_dtype=None  
name=None,  
experimental_is_sparse
```

공모전 내용

LLM을 통한 과학 기반 문제의 답변을 찾기
(정답 또는 유사한 답을 최대 3개까지 예측)

CPU, GPU \leq 9h run-time

Internet access disabled

Freely & publicly available external data

데이터셋 예시

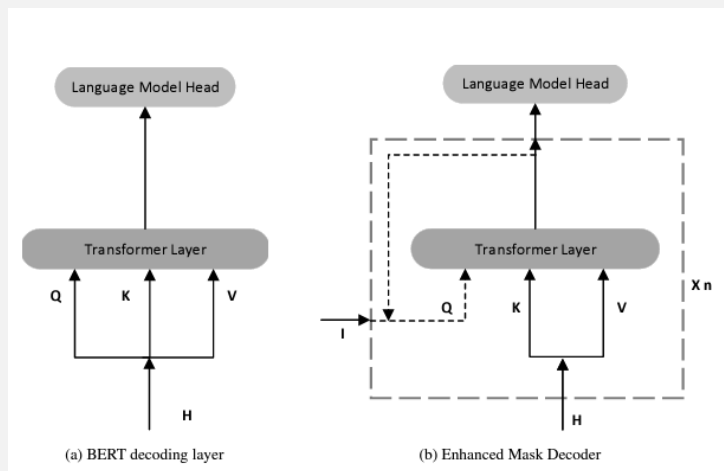
What is the term used in astrophysics to describe light-matter interactions resulting in energy shifts in the radiation field?

- A. Blueshifting
- B. Redshifting
- C. Reddening
- D. Whitening
- E. Yellowing

사용 모델

DeBERTa

기존 BERT 모델에서 2가지 방법을 이용하여 성능을 개선



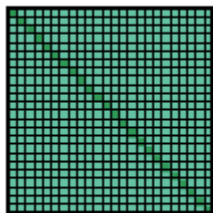
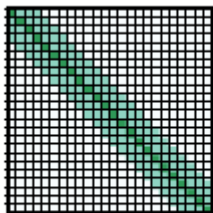
Disentangled attention
self attention layer에서 각 word들의 relative position에 따른 가중치 부여

Enhanced mask decoder
Transformation layer 이후 absolute position 정보를 주어 relative 와 absolute position 을 모두 고려

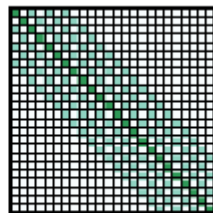
사용 모델

Longformer

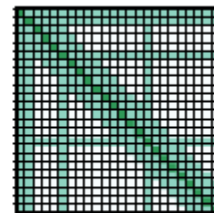
Sequence 길이에 선형적으로 비례하는 attention 메카니즘

(a) Full n^2 attention

(b) Sliding window attention



(c) Dilated sliding window



(d) Global+sliding window

Sliding window attention : 토큰 근처 w 크기의 윈도우에 대해서만 attention 진행

Dilated sliding window : d 개씩 토큰을 건너뛰며 윈도우 구성

프로토타입



llm science test - Version 8

Succeeded · 2h ago

0.534



- BERT -> DeBERTa, LongFormer
- Train with multiple dataset