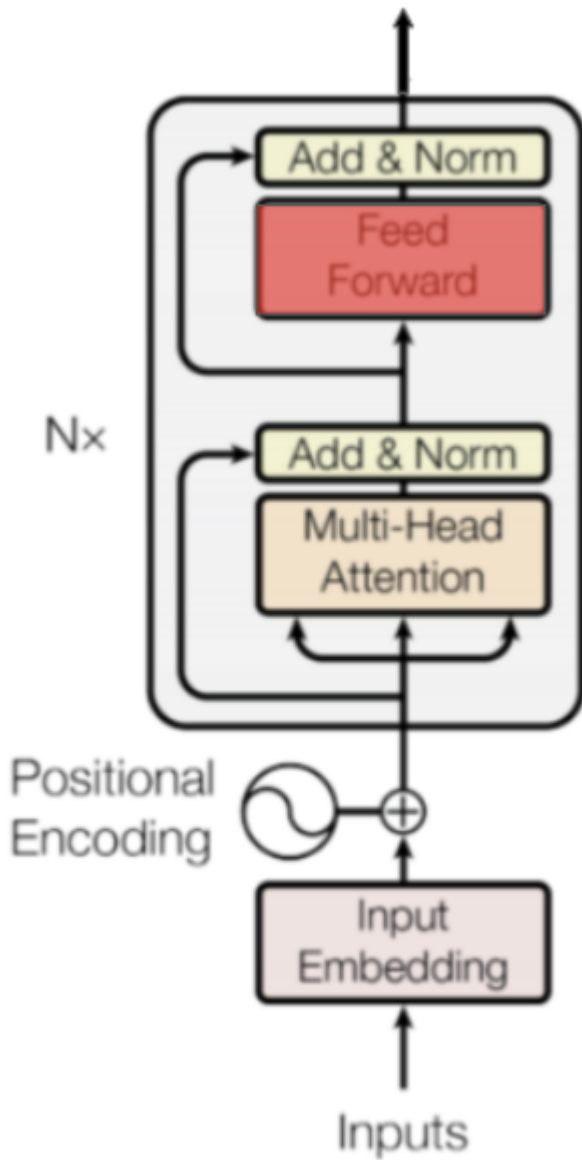


## 3.1. BERT

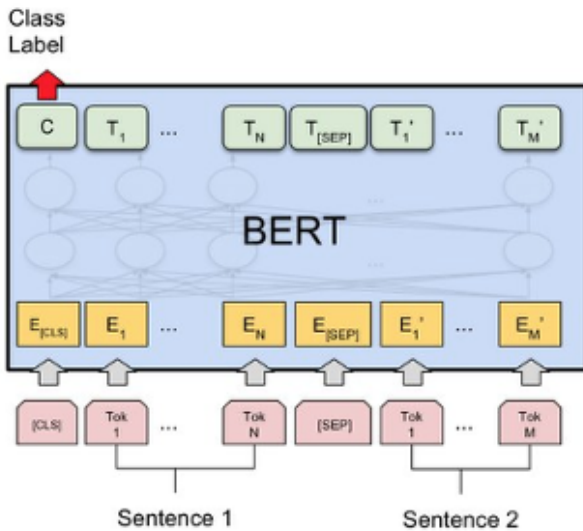
- 대규모 문서를 사용해 훈련시킨(사전에 훈련시킨) 이후 fine tuning을 거쳐 하위 NLP 분야에 사용하고 있는 범용적인 Language Understanding 모델
- Transformer 의 Encoder 부분만을 사용함
  - 단 인코더 레이어 개수(N)는 12(Base), 혹은 24(Large)를 사용
- Transformer의 positional encoding 대신 positional embedding 을 사용함
  - 위치 + 여러 segment(문단?) 로 나뉘어 있을 경우엔 segment 순서
- Position-wise feed forward 할 때 ReLU 대신 GELU 를 사용함



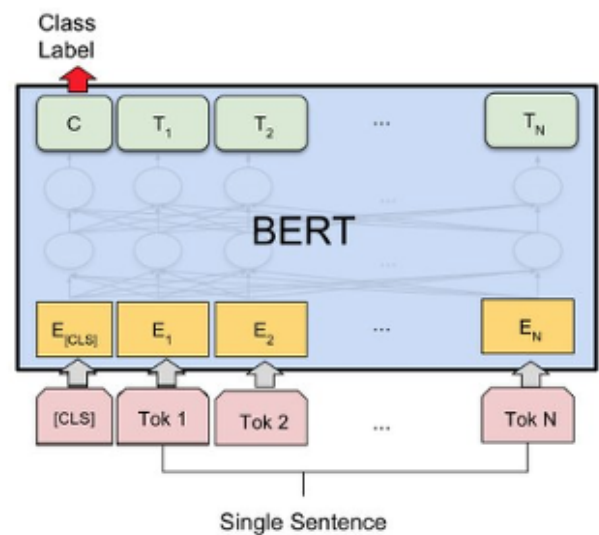
- 학습
  - Masked Language Model 및 Next Sentence Prediction을 함께 수행
    - Input 의 전체 단어에서 일부(15%) 정도를 masking 하고 이를 예측하기
    - Input sequence 및 문장을 이용해 corpus 내에서 앞 문장의 다음 문장인지를 예측하기

# BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

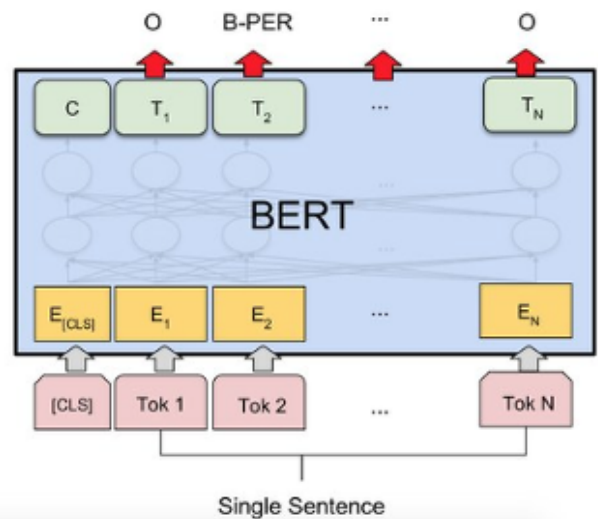
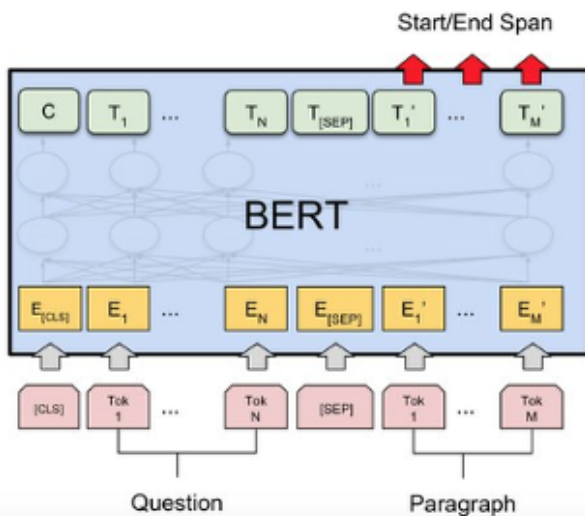
Jacob Devlin   Ming-Wei Chang   Kenton Lee   Kristina Toutanova  
Google AI Language



(a) Sentence Pair Classification Tasks:  
MNLI, QQP, QNLI, STS-B, MRPC,  
RTE, SWAG



(b) Single Sentence Classification Tasks:  
SST-2, CoLA



## Fine-Tuning

- Pre-train BERT 모델을 활용해서 4가지 형태로 Task 해결
  - 두 문장을 넣어 <CLS>토큰으로 유사한 문장인지 비교
  - 하나의 문장을 넣어 <CLS>토큰으로 Classification
  - Question과 문서를 입력해서 문서의 토큰들을 바탕으로 Question과 유사한 토큰 추출
  - 하나의 문장을 넣어 토큰별 의미분석
- Fine-Tuning 과정은 BERT모델 끝에 단순한 Fully Connected Layer 1~2개를 쌓아 각 Task를 해결
- Fine-Tuning 과정에서 BERT모델의 파라미터들을 재학습하기 때문에 Task 별로 사용하는 BERT의 파라미터가 다름