

BERT

BERT.

- 정의. - BERT는 Bi-Directional Transformer로 이루어진 언어 모델.
- 문장의 context vector를 잘 전달하기 위하여 사용.

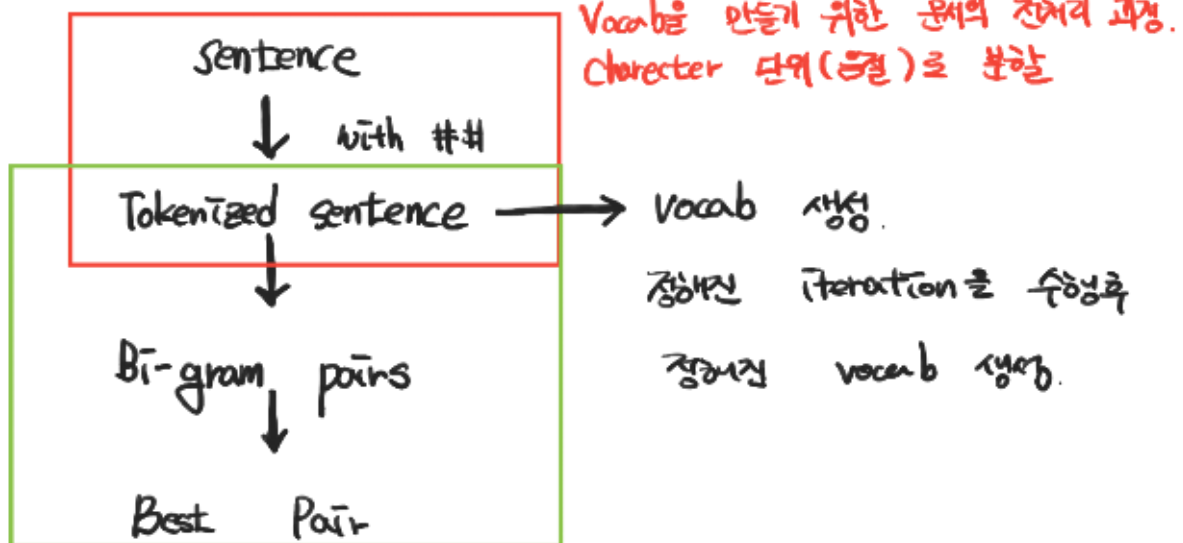
• BERT의 학습 데이터

✓ 데이터의 tokenizing.

- Word Piece Tokenizing.

→ 단어나에 기반하여 문장의 토큰 단위를 결정하는 것.
이미있는 자원 (sub Word)로.

*. 순서.



분리된 단어를 Bi-gram Pair에서
결합하여 Best Pair를 찾는 과정.

"Iteration" 만큼 진행.

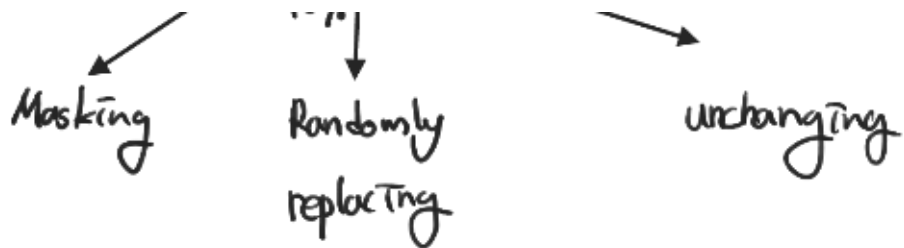
✓ MLM (Masked Language Model) : input token을 일정 확률로 Masking

ex)

<CLS> I go to school <SEP>

↳ 무작위 15%의 확률로 단어 선택.

80% / 10% / 10%



v Is Next ?

. BERT는 2개의 문장을 이어서 사용

✓ 50%는 원래. 이어진 다음 문장 50%는 랜덤으로 선택된 문장.