

# BERT



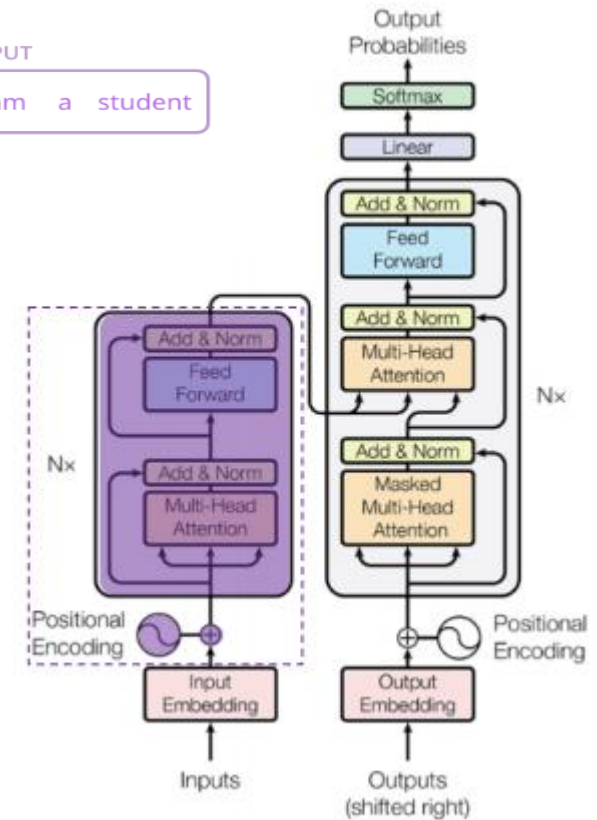
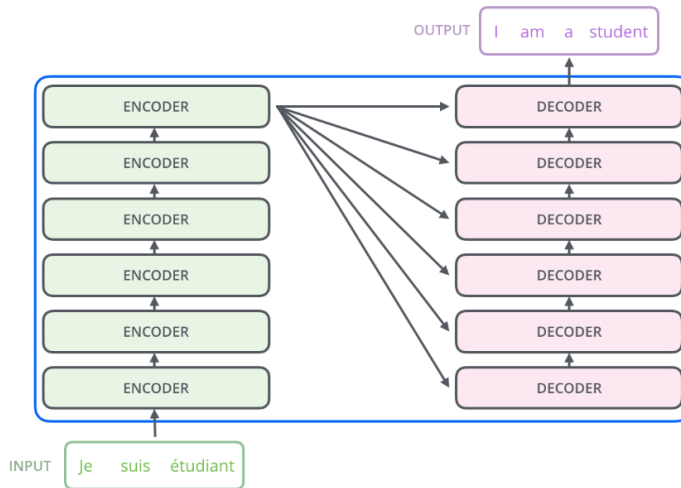
# [ BERT ]

- ① **셀프 어텐션**
- ② **Google BERT의 구조**
- ③ **BERT 활용 모델**
- ④ **BERT 활용 기술**

# 트랜스포머

셀프 어텐션으로 구성된 인코더

RNN,CNN을 사용하지 않고 기존 RNN으로 구성된 seq2seq의 한계를 넘어서는 성능을 보여 줌

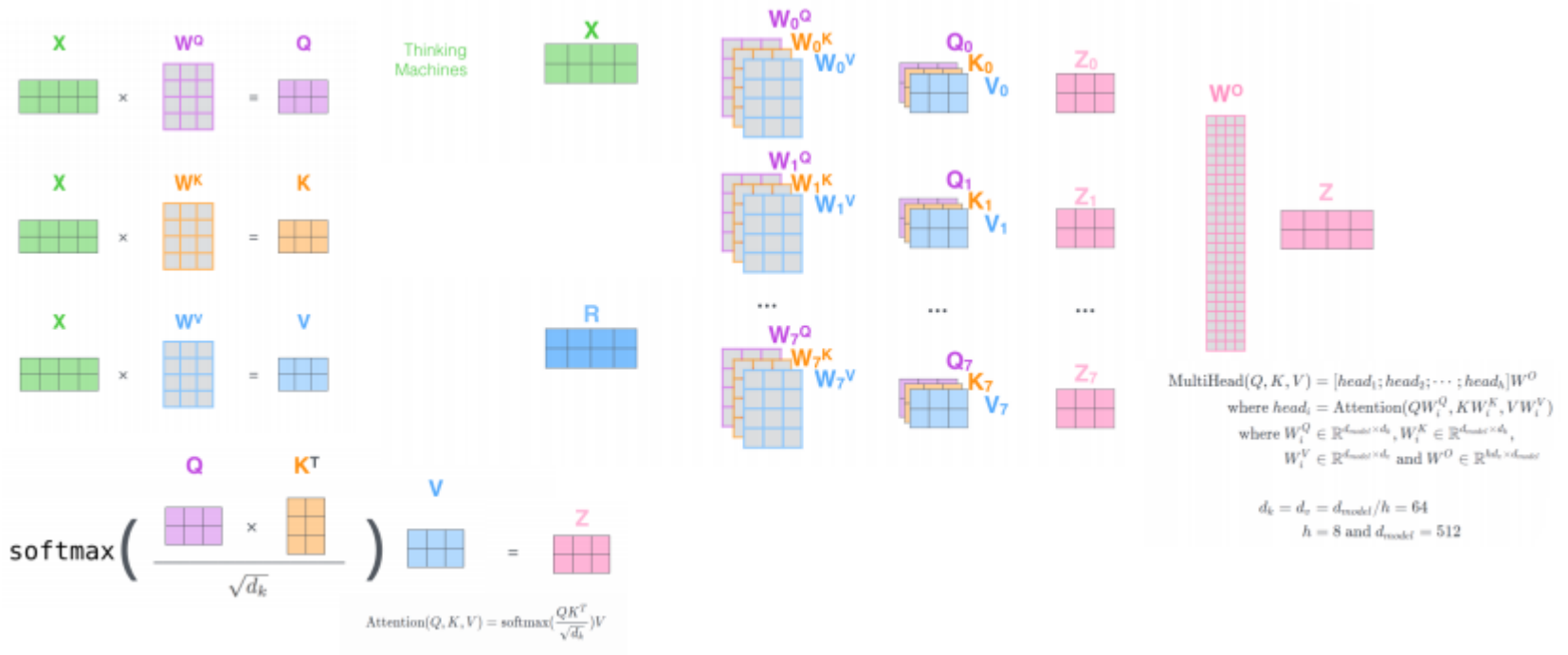


# 셀프 어텐션의 구조

입력을 각 가중치로 연산하여 Q,K,V를 구함.

이때,각 가중치는 멀티 헤드의 크기만큼 세트로 구성된 스케일 닷 어텐션에서 Q와 K의 트랜스포즈한 정보를 연산하고 스케일링하여 어텐션 스코어를 구함. 어텐션 스코어를 V에 적용하여 Z를 구함.

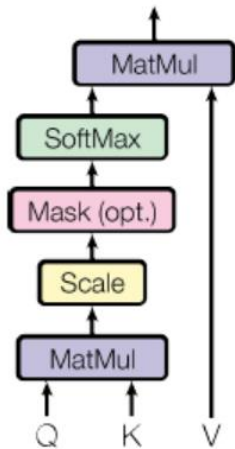
이 과정을 헤드 수만큼 피처를 나누어 스케일 닷 어텐션한 후 모두 합쳐 통합 Z를 구함.



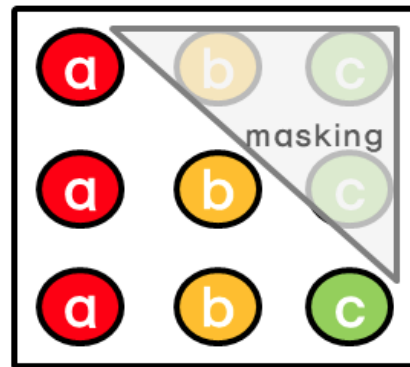
# 셀프 어텐션 마스크

셀프 어텐션에서 마스크는 두 가지 형태로 쓰임

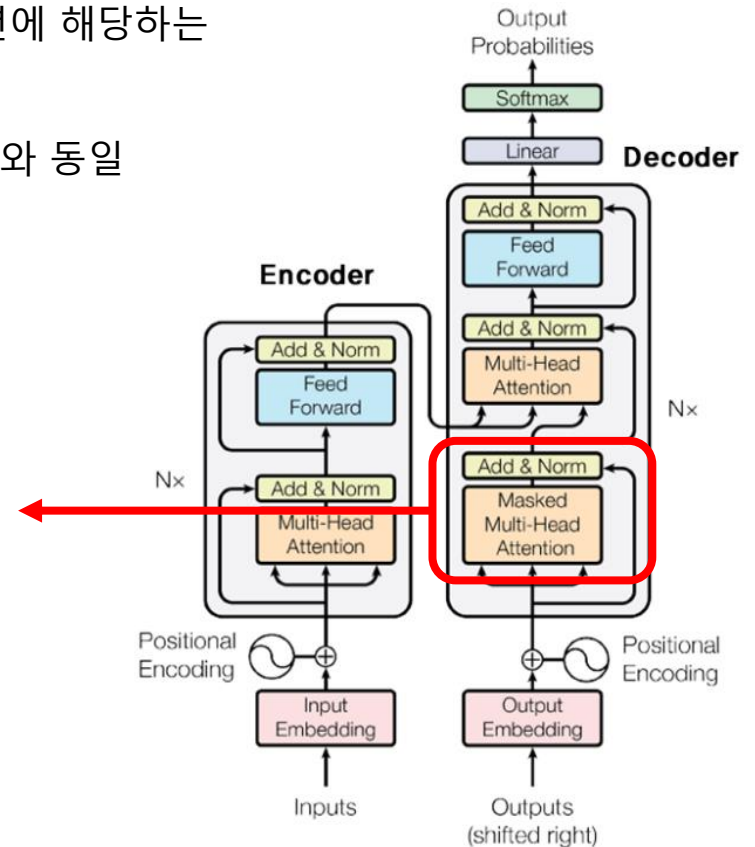
1. 인코더의 마스크 : 선택 사항으로 값이 없는 값(제로 패딩)에 페널티를 부과하여 어텐션 점수를 받지 못하게 함
2. 디코더의 마스크 : 다음 단어 예측을 위해 미래 포지션에 해당하는 정보가 어텐션 점수를 받지 못하게 함
3. 버트 인코더의 마스크는 트랜스포머 인코더의 마스크와 동일



인코더의 마스크



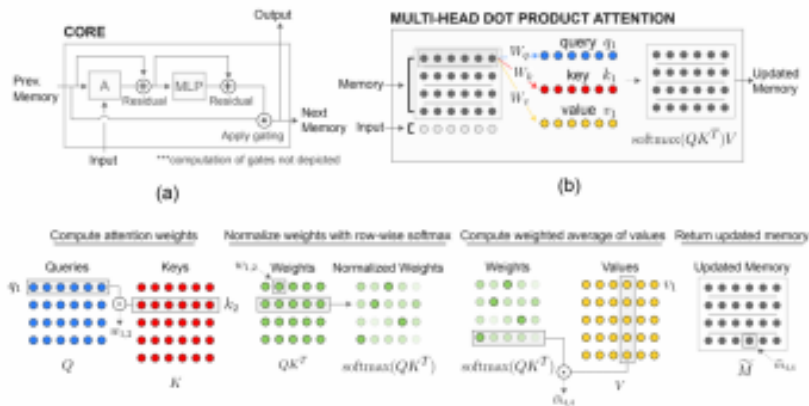
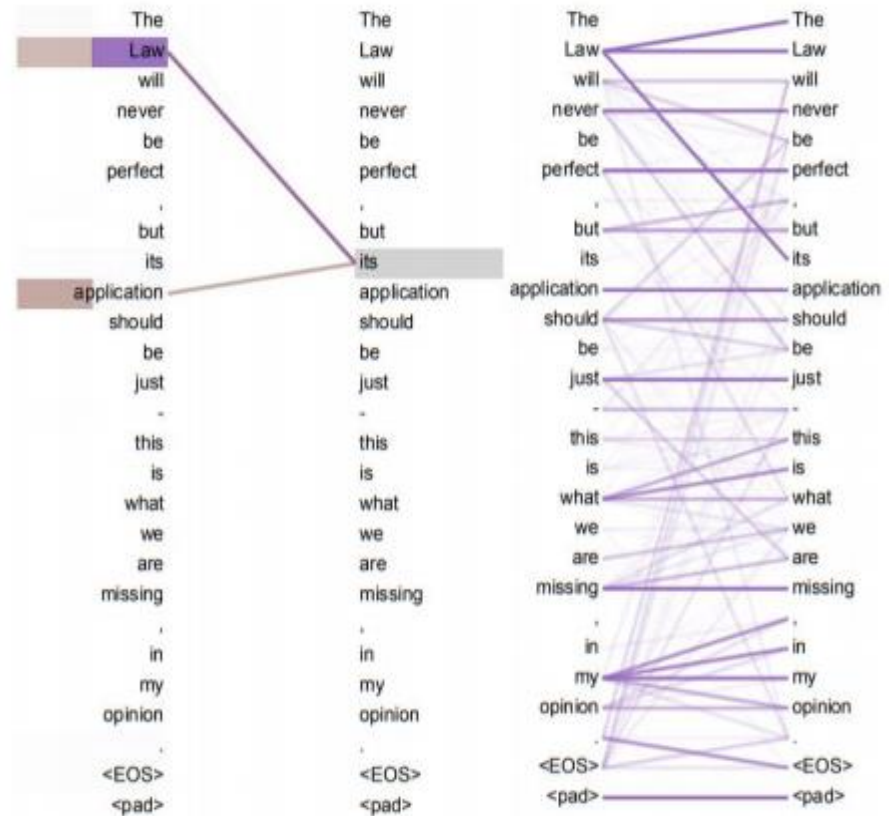
디코더의 마스크



# 셀프 어텐션의 효과

단어 간의 연관 관계 정보를 얻을 수 있음.

단어의 특징 정보를 가지고 있는 피처를 몇 개의 조각으로 나누어 셀프 어텐션하면 좀 더 효과적인 단어 간의 연관 관계 정보를 얻을 수 있음



# Drop-out, Normalization

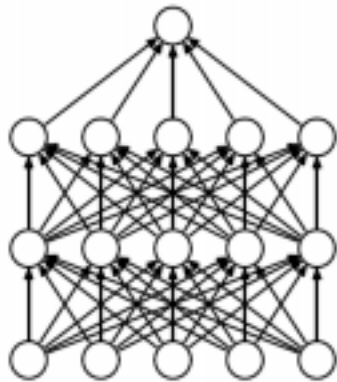
과적합(overfitting)을 줄이는 방법

## Drop-out :

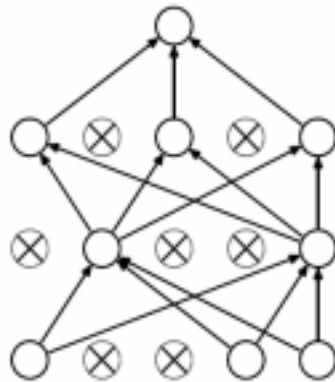
일부 노드의 연결을 임의로 삭제하는 것, 테스트 할 때는 모든 노드 사용

## Normalization(정규화) :

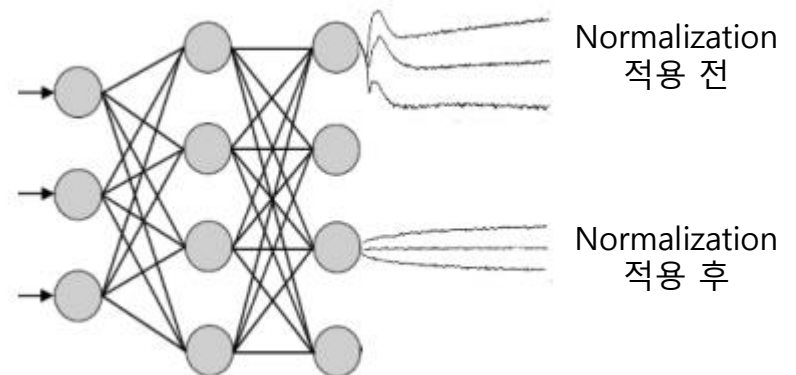
값이 원하는 범위를 벗어나지 않도록 제한함 값 범위를 제한하므로 가중치  
초깃값 선택의 의존성이 적어짐



일반적 뉴럴넷



Drop-out 적용 후



# [ BERT ]

- ① 셀프 어텐션
- ② Google BERT의 구조
- ③ BERT 활용 모델
- ④ BERT 활용 기술



# Google BERT 개요

트랜스포머로 구성된 양방향 언어 표현(Bidirectional Encoder Representations from Transformers)

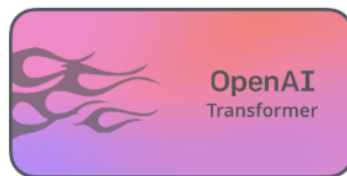
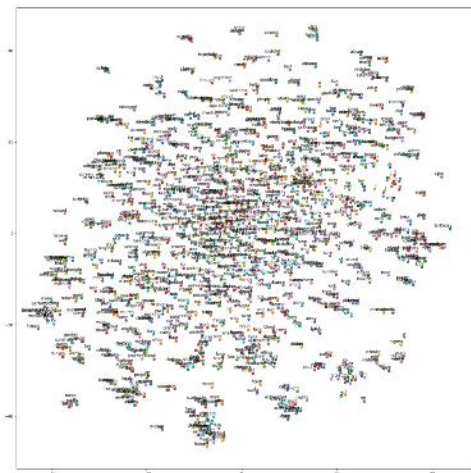
## 피쳐 기반 접근법

미리 학습된 가중치의 단어 별 특징(피쳐)을  
뽑아 학습에 사용하는 방식

이미지를 분류하여 카테고리 단어를 뽑아내는  
종류 에서는 높은 성능 향상을 나타내기도 함  
Word2Vec, GloVe, FASTTEXT, CoVe, ELMo 등

## 파인 튜닝 접근법

다른 분야의 Task에서 미리 학습시킨 가중치를  
불러와 미세 조정(파인 튜닝)하여 사용하는 방식  
영상 처리 분야에서 많이 사용되어 옴  
자연어 처리 분야에서는 2018년 들어 많이 나타나기  
시작한 기법 ULMFit, GLoMo GPT, BERT 등



# Google BERT 구조

트랜스포머로 구성된 양방향 언어 표현(Bidirectional Encoder Representations from Transformers)

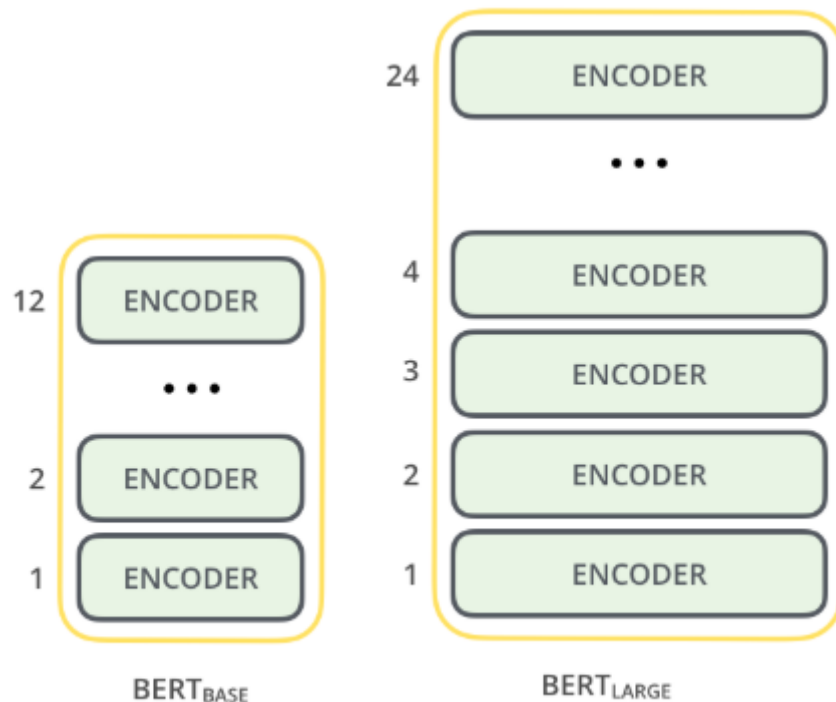
버트는 베이스 모델이 110M 파라미터가 넘고 라지 모델은 340M파라미터가 넘는 거대 인코더로 구성  
거대 인코더가 입력 문장들을 부호화하여 언어를 모델링하면 이를 파인 튜닝하여 여러 자연어 처리 Task를 수행할 수 있음

## BERT BASE

트랜스포머 레이어 수 : 12개,  
특징을 담는 피쳐 사이즈 : 768  
멀티 헤드 어텐션의 헤드 수 : 12개

## BERT LARGE

트랜스포머 레이어 수 : 24개,  
특징을 담는 피쳐 사이즈 : 1024  
멀티 헤드 어텐션의 헤드 수 : 16개



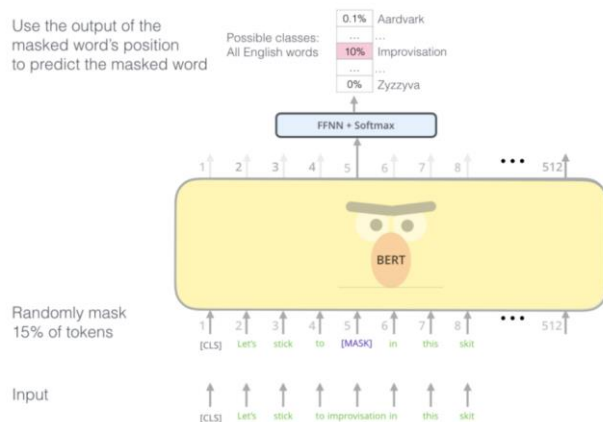
# Google BERT 학습 방법

## Masked 언어 모델

양방향 맥락을 고려하여 언어모델링 되도록 문장에서 정해진 비율(15%)의 무작위 단어 마스크.

마스킹된 단어를 맞추도록 모델을 구성함.

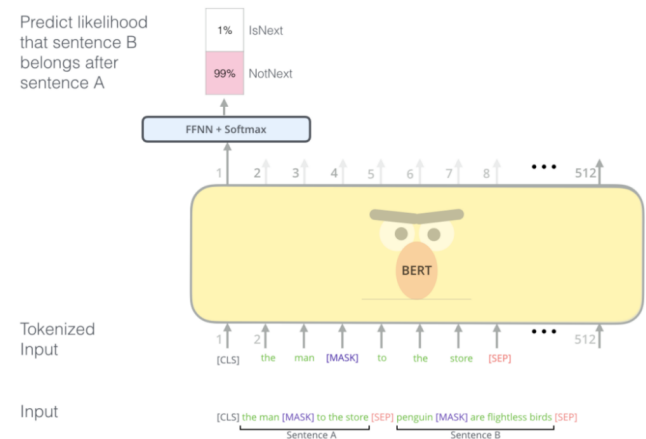
미세조정 시 올바른 예측을 돕도록 마스킹에 노이즈를 섞음(10%는 랜덤단어, 10%정상단어 그대로 둠)



## 다음 문장 예측

두 문장 간 관련이 고려되어야 하는 NLI와 QA의 파인 튜닝을 위해 마스크 된 단어를 맞추는 Task와 동시에 두 개의 문장이 연관이 있는지 맞추도록 학습함.

50%연관 문장, 50%연관되지 않는 문장을 무작위 추출하여 학습.

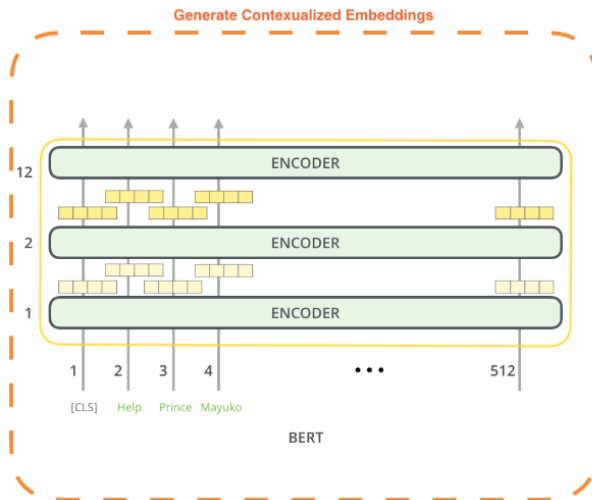


# [ BERT ]

- ① 셀프 어텐션
- ② Google BERT의 구조
- ③ **BERT 활용 모델**
- ④ BERT 활용 기술

# 피쳐 추출로 이용하는 방법

미리 학습된 버트를 사용하여 문맥이 고려된 워드 임베딩으로 사용 가능  
개체명 인식 같은 테스트에서는 파인 튜닝 방법과 유사한 성능을 내었음  
가장 좋은 성과는 마지막 4개의 정보를 합쳐 사용하는 것이 가장 좋은 결과를 냄  
(카카오에서 이를 응용해서 문장별 단어 벡터의 가중합의 거리를 계산하여  
관련 있는 문장의 유사도를 판별할 수 있는 서비스를 시도함)



The output of each encoder layer along each token's path can be used as a feature representing that token.



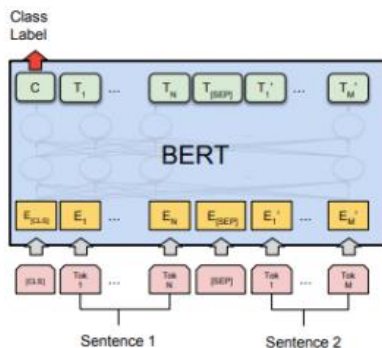
But which one should we use?

What is the best contextualized embedding for "Help" in that context?  
For named-entity recognition task CoNLL-2003 NER

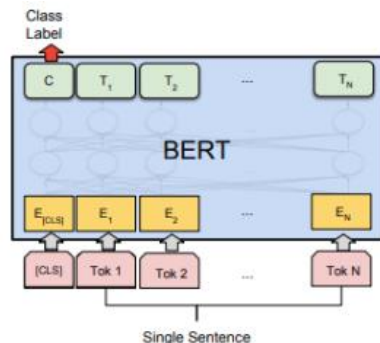
		Dev F1 Score
First Layer	Embedding	91.0
Last Hidden Layer		94.9
Sum All 12 Layers		95.5
Second-to-Last Hidden Layer		95.6
Sum Last Four Hidden		95.9
Concat Last Four Hidden		96.1

# 파인 튜닝 이용법

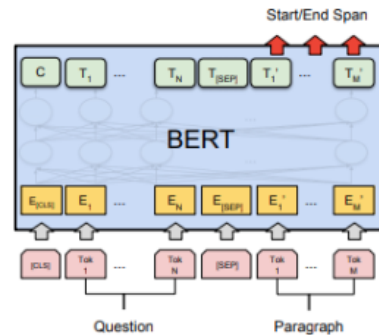
1. 프리트레이닝 가중치를 모델에 불러옴
2. 입력 문장이 두 개인 Task는 마스크 없이 문장 구분자(SEP)를 넣어 아래와 같이 입력함
3. 입력 문장이 한 개인 Task는 마스크와 문장 구분자(SEP) 없이 아래와 같이 입력함
4. 분류 문제는 클래스 라벨을, 정답 구간을 예측하는 문제는 시작 지점과 끝 지점을, 개체명 태깅 문제는 태깅 문장을 정답으로 넣음
5. 입력 형태는 바뀌지 않고 마지막 단어에 있는 분류자 부분만 다양한 자연어 처리 Task에 맞게 바꿔 BERT 가중치를 여러 분야에 활용할 수 있음



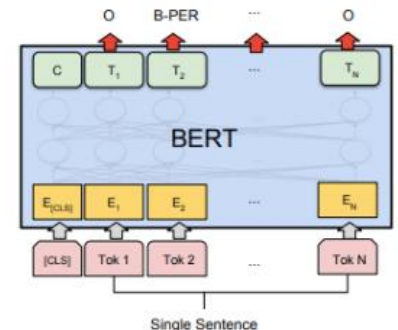
(a) Sentence Pair Classification Tasks:  
MNLI, QQP, QNLI, STS-B, MRPC,  
RTE, SWAG



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



(c) Question Answering Tasks:  
SQuAD v1.1



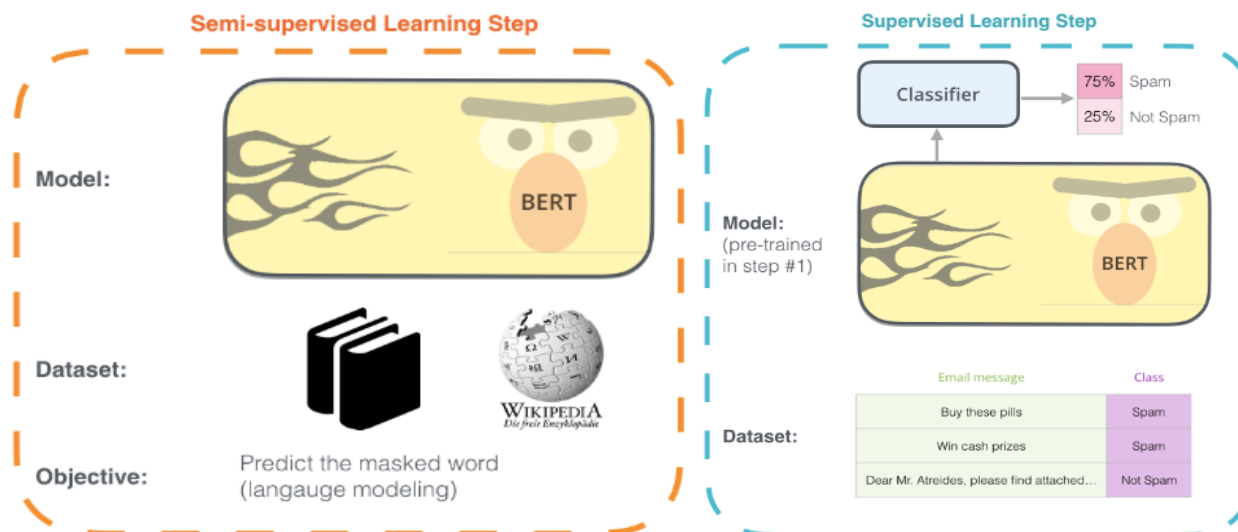
(d) Single Sentence Tagging Tasks:  
CoNLL-2003 NER

# 프리 트레인 모델

현재 보유하고 있는 한국어 BERT Pre-Train 모델은 한국어 뉴스 + 한국어 Wiki 데이터 약 1000만 건을 반복학습한 모델.

BERT가 학습해야 하는 파라미터 수는 약 1억 1천만 개이며, 현재 보유한 모델은 1080ti GPU x 8 을 사용하여 3개월 이상 학습.

이렇게 사전학습된 BERT 모델은 다양한 NLP TASK들을 학습하는 BASE 모델로서 역할. BERT-XDC, BERT-MRC, BERT-NER 등의 BERT 기반 모델들은 모두 위와 같이 Pre-Train을 거친 BERT를 기반으로 튜닝 및 추가 학습 됨.

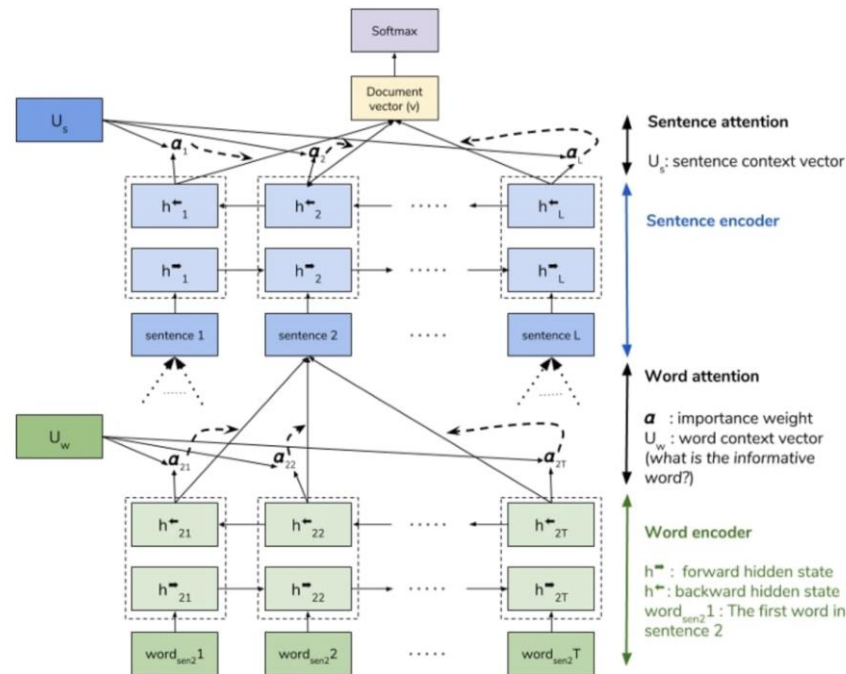


# Hierarchical Attention Networks

Bidirectional - RNN (양방향 RNN) 기반 모델

문장을 구성하는 단어 단위의 Attention과 문서 내 문장 단위의 Attention Weight를 추출 가능

Attention Weight를 기반으로 문서 분류에 있어서 어떤 단어, 어떤 문장이 높은 영향력을 가지는지 표현할 수 있는 모델



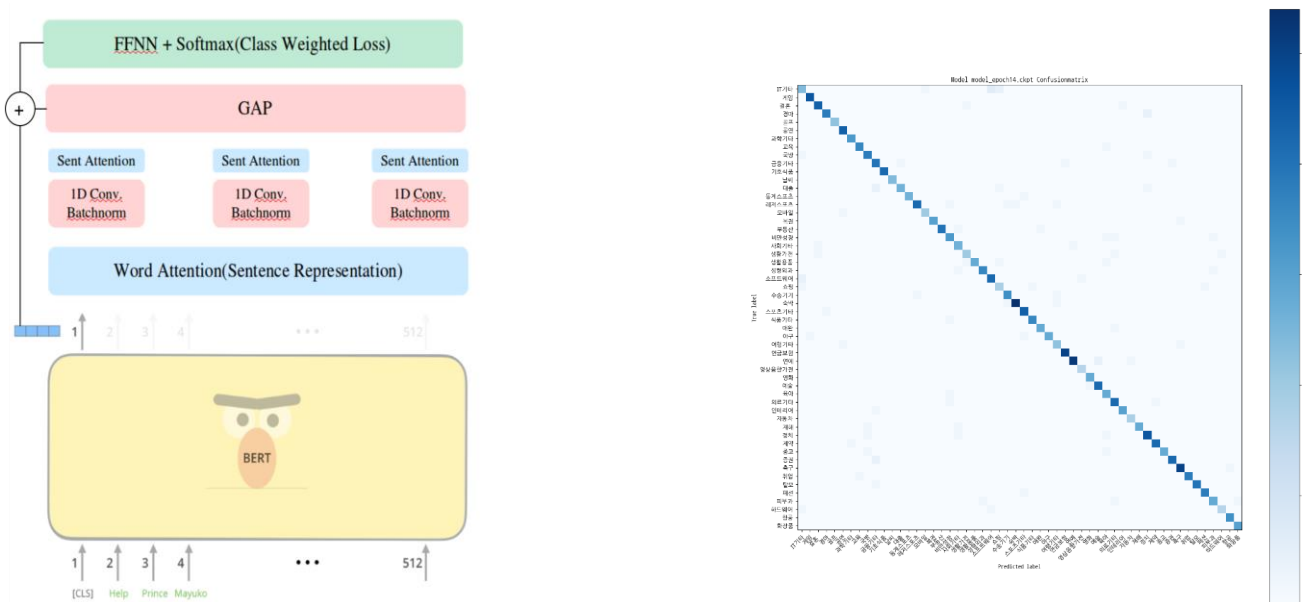


# BERT-CNN Attention Model

Pre-Train BERT Model을 기반으로 하고, CNN Based Attention Layer를 추가하여 성능을 향상시킨 모델

문장 내의 단어 단위의 Attention 및 문장 단위의 Attention Weight를 추출하고, 문서 분류에 있어서 어떤 단어, 어떤 문장이 높은 영향력을 설명할 수 있는 모델

BERT Word-Piece 방식을 사용함에 따라 ETRI - NLP 종속에서 탈피하면서도, 더 많은 단어를 커버할 수 있음



# [ BERT ]

- ① 셀프 어텐션
- ② Google BERT의 구조
- ③ BERT 활용 모델
- ④ BERT 활용 기술


# BERT 활용 기술

## MAUM-AI 플랫폼 내의 BERT-XDC

뉴스 본문을 입력하고 분석하기 버튼을 누르면  
해당 뉴스가 어떤 카테고리의 뉴스인지 분류하고,  
영향력이 높은 단어 및 문장을 Highlight하여 보여줌


정확한 텍스트 분류(XDC, eXplainable Document Classifier)를 통한 뉴스 분석

내용과 분량에 관계 없이 어떤 문서든 분류하고 분석합니다

 본문넣기

기사검색

문서 본문(뉴스기사)을 검색하거나 직접 붙여 넣기 해주세요.

 분석하기

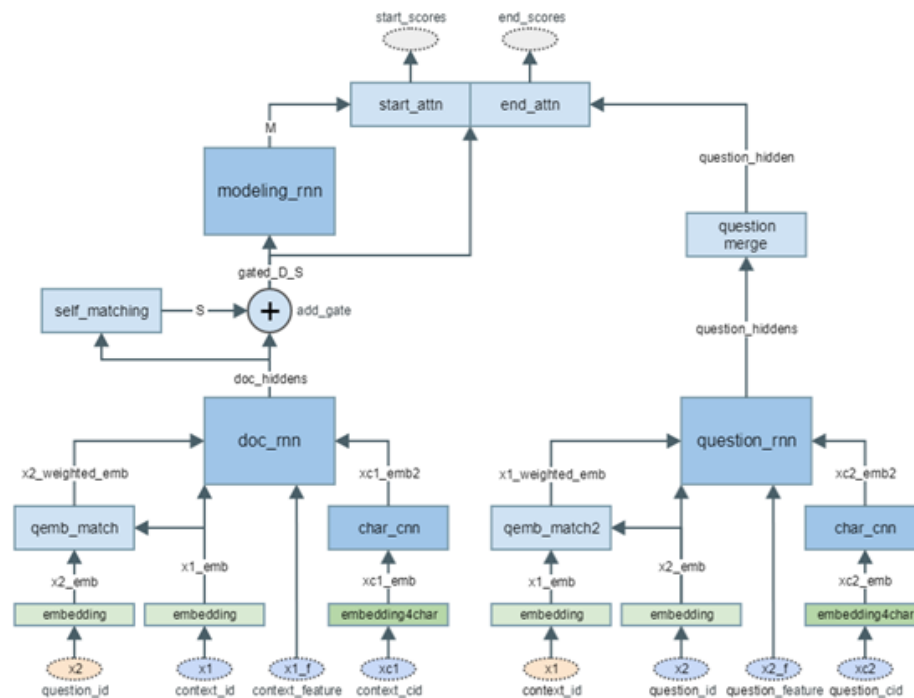
# BERT 활용 기술

기존 MRC(Machine Reading Comprehension)

RNN 기반 모델

초기 MRC 알고리즘

성능향상을 위해 많은 input feature 가 사용됨



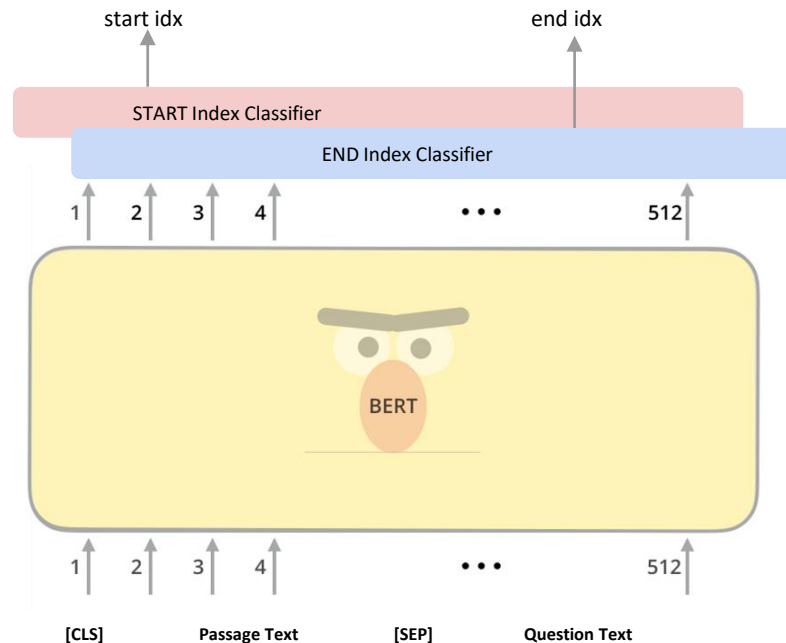
# BERT 활용 기술

## BERT MRC Model

Pre-Train BERT Model을 기반으로 하고, 정답이 있는 Start Index와 End Index를 추출하는 Dense Layer를 추가하여, Passage 내에서 질문에 대한 답변을 추출하는 모델

기존 모델에 비해 MRC 모델의 구조가 매우 단순해지면서도, BERT Model을 활용함으로써 성능이 크게 향상됨

Word-Piece 방식을 사용함으로써 ETRI - NLP 종속에서 탈피함



# BERT 활용 기술

## BERT MRC Model – 한글 MRC Model 성능 비교

### S2 Net Model vs BERT - MRC Model

전국에 장맛비가 내린 지난 24일 오전 대전 서구의 한 신호등에서 시민들이 우산을 쓰고 걷고 있다. 기상청은 29일 남부지방과 제주에 비가 내릴 것으로 예보했다. 2016.6.24/뉴스1 News1 주기철 기자차윤주 기자 = 수요일인 29일 전국이 흐리고 남부지방과 제주도는 빗방울이 떨어지겠다. 기상청은 이날 남부지방과 제주도가 장마전선의 영향을 받아 대체로 흐리고 비가 오는 곳이 있을 것으로 예보했다. 남부지방 비는 아침에 그치고 제주는 종일 내릴 것으로 보인다. 강원도 영서는 오후에 소나기가 오는 곳이 있고, 중부지방은 구름이 많이 낄 전망이다. 아침 최저기온은 17~21도, 낮 최고기온은 23~32도로 오늘과 비슷하겠다.

수요일 날씨는?

**SNet: 29일**

**BERT: 전국이 흐리고 남부지방과 제주도는 빗방울이 떨어지겠다**

두산이 한화에서 거포 내야수 신성현(27)을 영입했다. 대신 포수 최재훈을 내줬다. 최재훈은 그동안 여러 팀이 탐냈을 만큼 수비력이 뛰어난 최고의 백업 포수다. 그러나 이미 주전 포수 양의지와 또 다른 백업 포수 박세혁이 탄탄하게 자리를 잡고 있다. 결국 오랜 고민 끝에 결단을 내렸다. 신성현은 현재 두산 전력에 꼭 필요한 선수였다. 두산 관계자는 "현재 두산 내야수 엔트리는 김재호와 허경민을 제외하면 모두 좌타자들"이라며 "힘 있는 대형 우타 내야 자원이 필요한 상황이었다"고 설명했다. 트레이드 제안은 포수가 필요했던 한화가 먼저 해왔다. 두산도 이 참에 늘 눈여겨봤던 신성현을 데려오기로 했다. 두산 관계자는 "양 구단의 이해관계가 맞아 떨어져 트레이드가 성사됐다"고 덧붙였다.

최재훈이랑 트레이드 된 선수는 몇 살이야?

**SNet: 신성현**

**BERT: 27**

# BERT 활용 기술

## BERT MRC Model – MRC 2.0 Model

본문에 정답이 없을 때에도 무조건 예측 값을 추출하는 1.0 버전과 달리, 본문 내에서 답 할 수 없는 질문에 대해 정답을 추출하지 않는 모델

정답이 없는 본문에 대해 정답을 추출하지 않음으로써, 질문에 대한 정답을 담고 있는 본문을 선별 하는데 도움을 줄 수 있음

### <MRC 2.0 예시>

정부가 공식 통계 지표로 활용하는 한국감정원의 '주간 아파트 가격 동향'마저 보합 전환하며 서울 집값 오름세에 대한 우려가 커지고 있다.

불안 조짐이 계속되면 추가 대책을 내놓겠다는 정부·여당의 '경고 메시지'가 지속적으로 나오는 가운데 국토교통부가 검토하겠다고 밝힌 '민간택지 분양가 상한제 적용'이 주택 가격 안정에 악이 될지 오히려 독이 될지 시장의 관심이 쏠리고 있다.

감정원이 27일 발표한 주간 아파트 매매가격 동향에 따르면 6월 마지막 주 서울 아파트값은 전주와 같았다.

내림 폭이 줄어들다 결국 33주 만에 하락세를 멈춘 것이다.

서울 아파트 가격은 민간 기관 집계로는 이미 플러스 전환했다.

부동산 114의 경우 6월 둘째 주부터, KB부동산 리브온의 경우 지난주인 6월 셋째 주부터 상승세로 돌아섰다.

감정원은 "강남 일부 지역 재건축 및 신축 매수세로 시장 하방 경직성이 커지며 보합 전환했다"고 설명했다.

질문1 : 집값 얼마나 올랐어?

답변1 : ''

질문2 : 정부가 활용한 통계지표가 뭐야?

답변2 : '주간 아파트 가격 동향'

질문3 : 아파트 값이 몇주 만에 하락세를 멈췄니?

답변3 : 33주

# BERT 활용 기술

## BERT MRC Model – maum.ai 플랫폼 내의 BERT MRC

뉴스 기사 등 질문 하고자 하는 예문과 질문을 입력하고 답변 찾기 버튼을 누르면,  
예문 내에서 답변의 위치를 Highlight하여 보여줌

정확하게 읽고 답하는 AI 독해 (MRC, Machine Reading Comprehension)

내용을 읽고 분석하여 질문에 맞는 답을 정확하게 찾습니다

본문넣기

언어 한국어 ▼

예문

기사

인공지능(AI) 전문기업 마인즈랩이 캐나다 3대 인공지능 연구기관 에이미(AMII: Alberta Machine Intelligence Institute)에 합류하고, 구글 딥마인드를 비롯한 세계 최고 수준의 AI 연구진과 함께 딥러닝 분야의 연구를 함께한다. 마인즈랩은 에이미의 회원사로서 에이미 소속의 연구진들과 함께 공동 연구를 추진한다. 인공지능 강국으로 평가받는 캐나다에서 엘리먼트AI(Element AI), 벡터 연구소(Vector Institute)와 함께 3대 인공지능 연구기관 중 하나로 꼽히는 에이미는 강 화학습의 창시자 리차드 서튼 교수의 주도 아래 알파고를 개발한 구글 딥마인드와 앨버타 대학 등이 모여 각 분야에서의 딥러닝 연구를 활발하게 이끌고 있다.

질문하기

마인즈랩은 어디에 합류했나요?

버튼 클릭

답변찾기