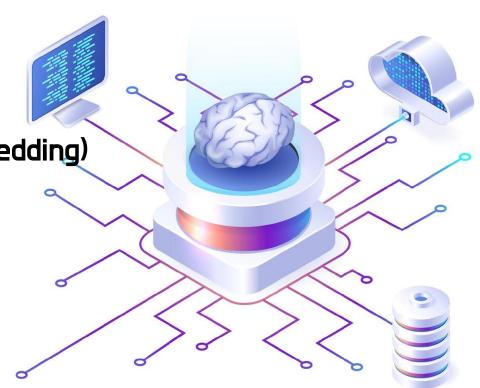
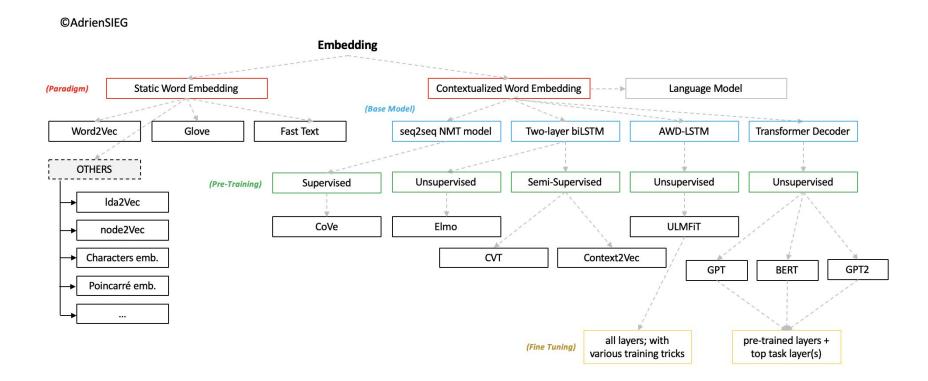


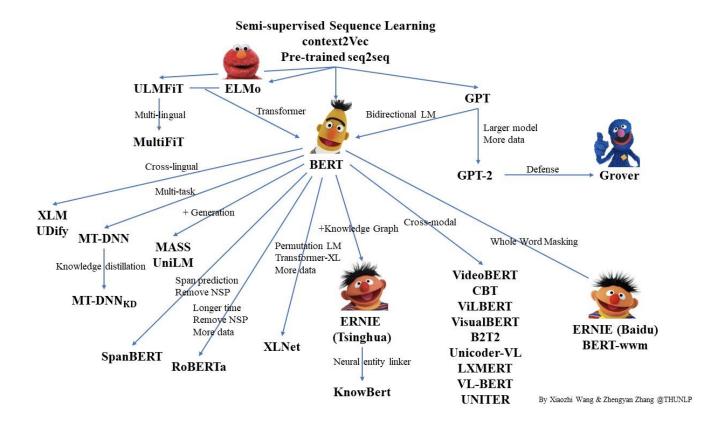


(Contextualized Word Embedding)

실무형 인공지능 자연어처리



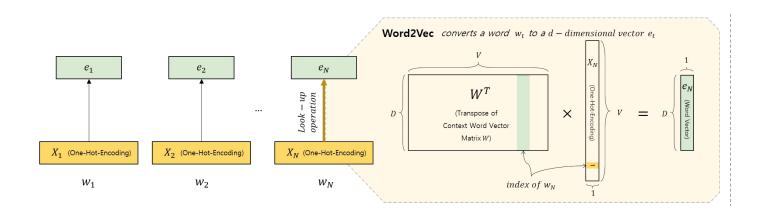




- Contextualized Word Embedding은 문맥에 따라 vector를 생성
- 같은 단어여도 문맥에 따라 다른 vector가 생성될 수 있음
- 대표적으로 ELMo, BERT, OpenAl GPT
- 이들의 특징은 같은 단어라도 문맥에 따라 다른 방식으로 표현(representation)



- Word2Vec의 Embedding은 단어 단위
- 각 단어의 one-hot-encoding vector가 W<sub>⊤</sub>와 곱해져서 Word Vector를 얻게됨

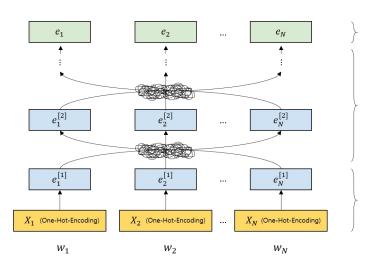




- Contextualized Word Representation의 경우 문장을 입력 받아 각 단어에 대한 representation을 산출
- '문맥'에 의존적인 '단어의 의미'를 잡아내는 feature가 산출

#### **Contextualized Word Representations**

converts a sequence of words (more precisely, tokens)  $w_1w_2...w_N$  to a series of d – dimensional vectors  $e_1e_2,...e_N$ 



Output: series of d – dimensional vectors  $e_1e_2, ... e_N$ 

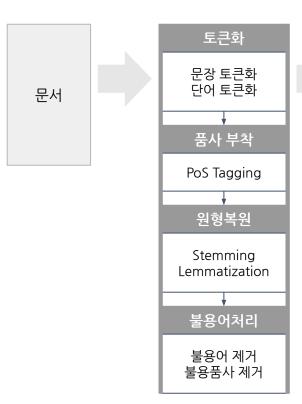
deep neural network (e.g. deep bi-LSTM or deep-self-attention Nets) for extracting features

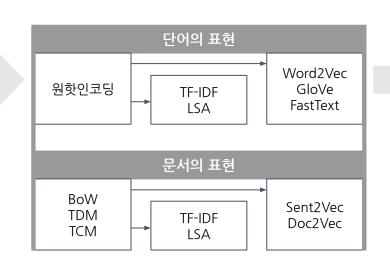
simple word (or token) embedding + (optional) positional embedding



관점	Word2Vec	Contextualized Word Representation
Input	단어 단위	문장 단위 (단어의 시퀀스)
Layer	(일반적으로) 단층	(일반적으로) 다계층
Output	해당 단어에 대한 Embedding	문장을 구성하는 각 단어에 대한 Embedding들

# 임베딩 절차





문맥적 단어 임베딩 ElMo BERT

## BERT 까지

며 H 高 Ħ 다 당 대 장 旧맹

### RNN

- 순환 구조로 시계열에 대한 고려 가능
- 장기문맥 손실 (Long dependency)

### LSTM / GRU

- grad
  vanishing/exploding
  문제 해결
- 장기문맥에 대한 고려

#### ElMo

- biLM (양방향 언어모델)
- Bidirectional, Stacking

#### CNN

- 이미지 분류에 많이 사용
- 시계열에 대한 고려 X

### Seq2Seq

- Encoder Decoder 구조
- Many to many
- 고정 문맥 벡터로 표현하여 정보 손실 발생
- Teacher forcing

# Seq2Seq with Attention

- 집중해야할 정보를 고려 = attention
- LSTM 기반 attention
- Attention mechanism

### Transformer

(Attention is all you need)

- LSTM을 제거. 속도향상
- Self-attention
- Multihead attention
- Positional encoding

### **BERT**

- Wordpiece tokenizer
- Masked Language Model(MLM)
- Next Sentence Prediction (NSP)
- Position embedding
- Pre-trained model
- Fine tuning

## 감사합니다.

Insight campus Sesac

