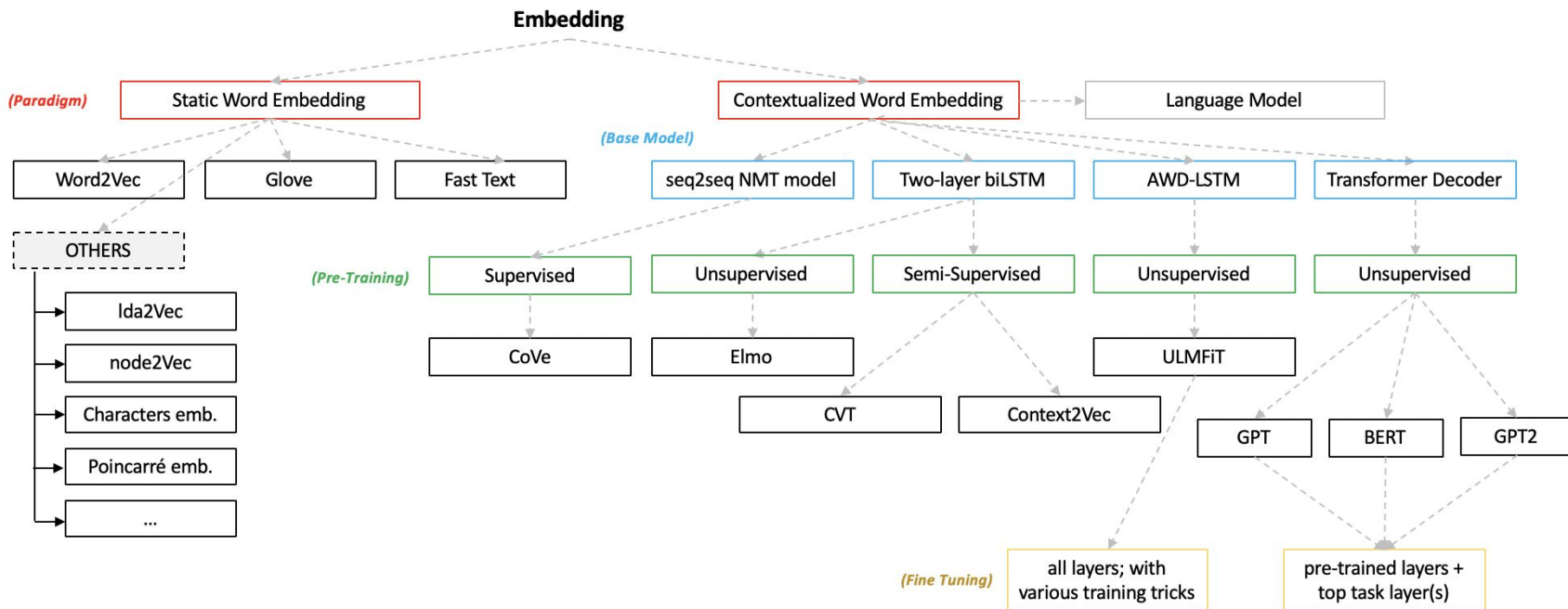


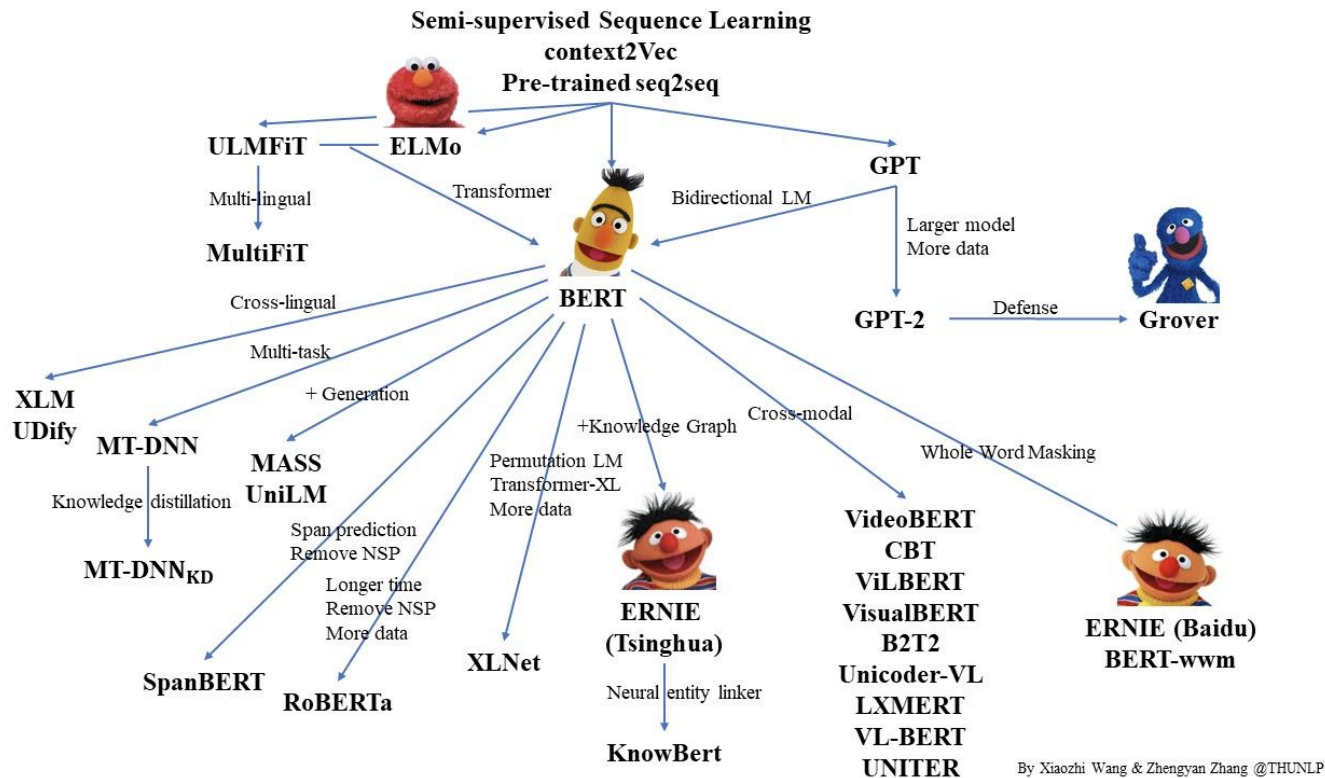
문맥적 단어 임베딩 (Contextualized Word Embedding)

실무형 인공지능 자연어처리



©AdrienSIEG



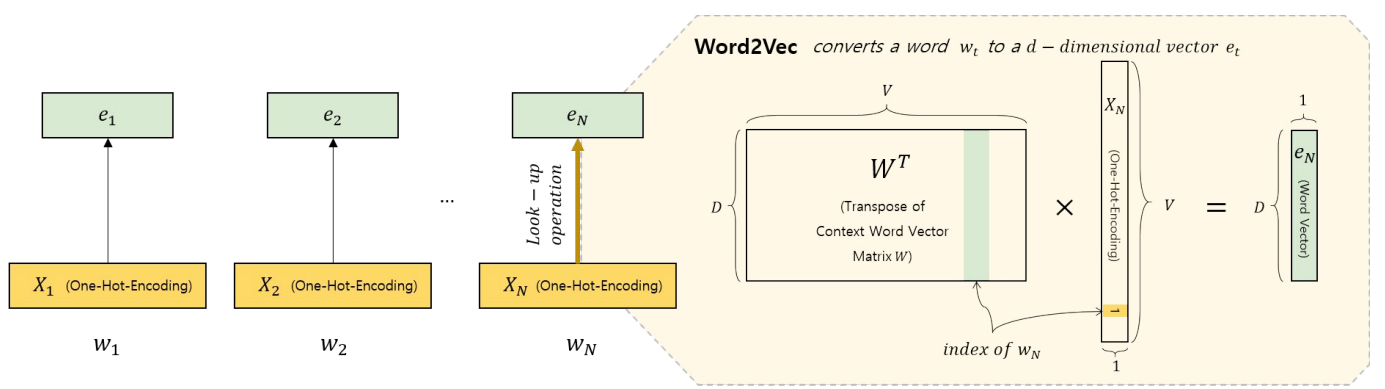


문맥적 단어 임베딩 (Contextualized Word Embedding)

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

문맥적 단어 임베딩 (Contextualized Word Embedding)

- Word2Vec의 Embedding은 단어 단위
- 각 단어의 one-hot-encoding vector가 W_T 와 곱해져서 Word Vector를 얻게됨

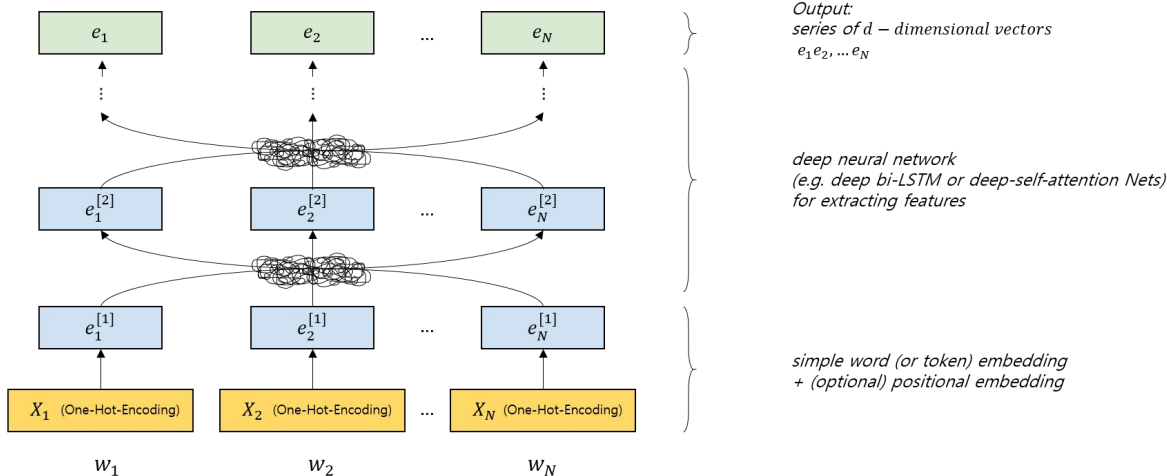


문맥적 단어 임베딩 (Contextualized Word Embedding)

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

Contextualized Word Representations

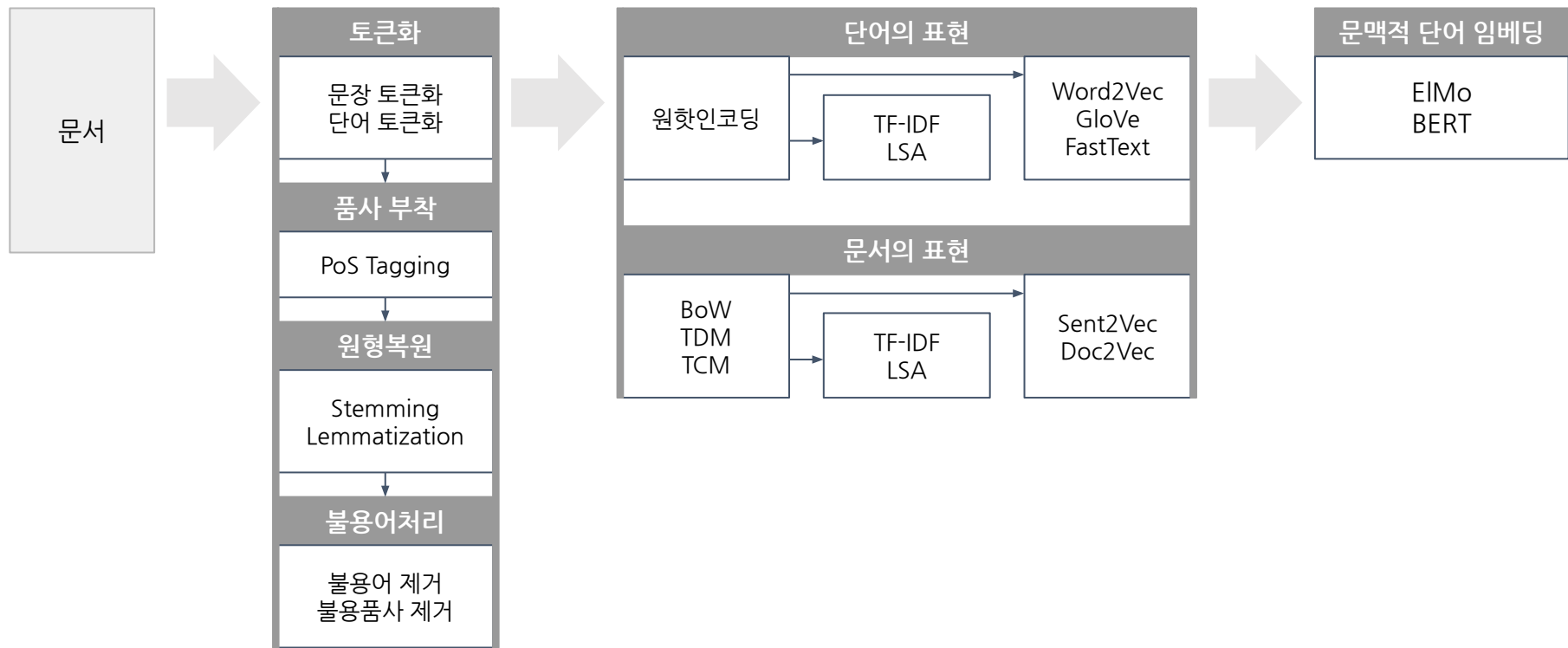
converts a sequence of words (more precisely, tokens) $w_1 w_2 \dots w_N$ to a series of d -dimensional vectors $e_1 e_2, \dots e_N$



문맥적 단어 임베딩 (Contextualized Word Embedding)

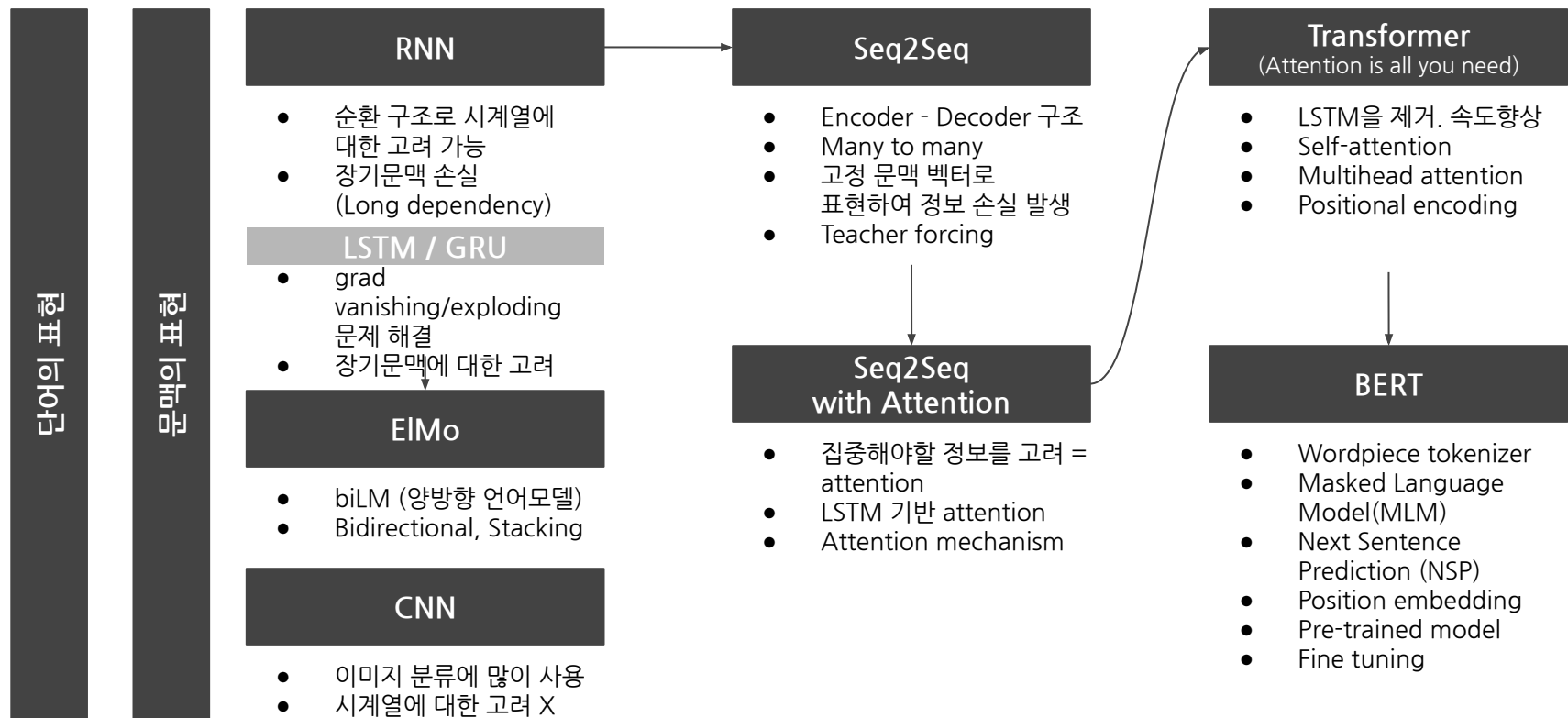
관점	Word2Vec	Contextualized Word Representation
Input	단어 단위	문장 단위 (단어의 시퀀스)
Layer	(일반적으로) 단층	(일반적으로) 다계층
Output	해당 단어에 대한 Embedding	문장을 구성하는 각 단어에 대한 Embedding들

임베딩 절차



처리 의존도를 고려하여 성능을 높이는 작업을 진행할 수 있음

BERT 까지



감사합니다.

Insight⁺campus

