

< 텍스트 인베딩 변환 >

• 2.2.2

[토큰화] → **토큰 인베딩** → **워드 인베딩** ⇒ **단어 인베딩** 이어져.

2.2.1 토큰화.

- 어떤 토큰이 어떤 단어인지 알기 + 해석하기
- 토큰 단위에 따른 장단점 → 등장 빈도에 따른 토큰화 단위 정함.
⇒ **서브워드 토큰화**

2.2.3 워드 인코딩.

- RNN vs 트랜스포머 기작은 차이 + 이점점.
- ⇒ 이점점 리징에서 단어 정해나서있고. + But, 텍스트에서 중요한 정보

(2.3) 이점점 이야기하기

2.3.1 사람의 글 읽는 방법과 이점점.

- 눈은 큰 흐름, But, 어떤 글은 자주 멈추면서, 단어간의 연결성을 문장 안에서 볼 수 있는 것.

OO 파리 OOOO → 고출입시, 4H인리
 저기 파리 날라간다. → 문장이네
 ⇒ 문장 단어 + 맥락을 파악할 수 있는 것.

키: 자물. 키: 열림. 값: 문장. 이점점.

Q: 열림

파리 →

| 키 | 값 |
|------|-------------|
| 저기 | 0.2, 0.3... |
| 파리 | 0.3, 0.6... |
| 날라간다 | 0.1, 0.9... |

⇒

| |
|------|
| 파리 |
| 날라간다 |

- 파리를 키로 검색했을 때 문장을 찾는다.
 나열된 것을 찾는다.