

pytorch를 이용한 자연어...

## Chapter01\_ PyTorch\_NLP\_Basics\_1.ipynb

### TF-IDF 표현

Term-Frequency-Inverse-Document-Frequency

TF = Term-Frequency 단어의 등장횟수

= the, a -> 자주등장해도 의미가 없어서

IDF = Inverse-Document-Frequency

그 수치값을 깎자는 의미

여러 문서에서 공통적으로 나오는 단어들을

수치를 낮추기 위해서 쓰는것

N : 전체 문서 개수

Nw: 단어 w를 포함한 문서의 개수

```
N_the = 10
IDF(the) = log(11/11) + 1
N_apple = 2
IDF(apple) = log(11/3) + 1
```

$$IDF(w) = \log \left( \frac{N + 1}{N_w + 1} \right) + 1$$

모든 문서에 등장 (즉,  $N_w=N$ )이면  $IDF(w)=0$

미리보기

맞춤법

글감검색