

토큰 (token) : 영어일 경우 단어 1개, 하나의 샘플은 여러 개의 토큰으로 이루어져있고,
1개의 토큰이 하나의 타임스텝에 해당, 단어하나를 정수로 매핑 (Che \rightarrow 10, Cat \rightarrow 11)

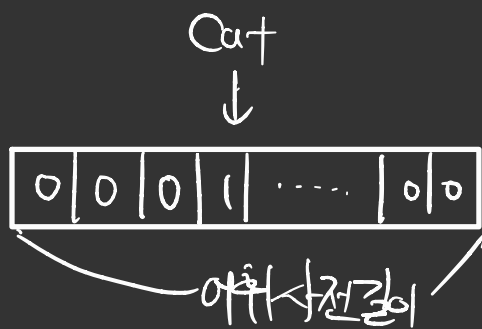
여취사전 : 훈련세트에서 고정한 단어를 뽑아 만든 목록, 없는 단어는 0로 변환해 보델에 주입

pad-sequence : 토큰의 개수 (타임스텝)를 맞추기 위해. maxlen 보다 짧으면 0으로 채워,
길다면 앞부분을 자름.

one-hot encoding : 토큰을 하나의 정수로 매핑할때 그 정수가 어떤 중요도를 나타내는지
아니기 때문에 one-hot encoding을 통해 크기 속성을 없앴듯이 input으로 전달해야 함
(데이터가 매우 커짐)

Word embedding : 각 단어를 고정된 크기의 실수 벡터로 바꿔줌.
원래 인코딩 벡터보다 훨씬 의미있는 값으로 채워져있으므로 자연어처리에서 좋은 성능
(단어사이의 관계 표현 가능)

One-hot encoding.



Word embedding.

