

Chapter2_요약_basic_nlp.ipynb

pytorch를 이용한 자연어입문 · 2024. 9. 18. 15:41

기초적인 용어들

corpus(말뭉치): 문서를 대량으로 모아놓은것. 자연언어 연구를 위해 특정한 목적을 가지고 언어의 표본을 추출한 집합

token: 텍스트를 나누는 단위

tokenization(토큰화): 텍스트를 토큰으로 나누는 과정

n-gram

n-gram: 연속된 token 시퀀스. 길이 n의 연속된 토큰 시퀀스

unigram: 토큰 1개

bi-gram: 토큰 2개

품사태깅

품사로 분류하는것(명사, 부사, 형용사...)

공감

구독하기

'[pytorch를 이용한 자연어입문](#)' 카테고리의 다른 글

[Chapter_3_요약_Diving-Deep-into-Supervised-Training.ipynb](#) (0)

2024.09.18

[Chapter01_PyTorch_NLP_Basics_1.ipynb](#) (0)

2024.09.18

[8-2](#) (0)

2024.06.28

[8-1_\(attention_RNN을 이용한 sequence to sequence의 문제\)_NMT_No_Sampling.ipynb](#) (1)

2024.06.21

[8-0](#) (0)

2024.06.14

관련글

[관련글 더보기](#)

Chapter_3 요약 Diving-Deep-into-Supervised-Training.ipynb

Chapter01_PyTorch_NLP_Basics_1.ipynb

8-2

8-1_(attention_RNN을 이용한 sequence to sequence의 문제) ...

자연어(NLP)

네이쳐2024님의 블로그입니다.

구독하기 +

댓글 1



익명

비밀댓글입니다.

2024. 9. 18. 15:50

⋮



이름

비밀번호

내용을 입력하세요.



로그인