|  |  |  |
| --- | --- | --- |
| 연구주제 | 형태소합성 네트워크에 맞는 데이터 정제 | |
| 연구목적 | Encoder 와 Decoder 에 들어갈 입력과 출력 형태에 맞는 데이터 쌍을 제작한다. | |
| * 데이터 분석   + 세종 코퍼스의 자료형태를 분석한다.   + <p></p> 사이의 단어들. 이들은 모두 하나의 문장임. * Network 상에 들어갈 수 있는 형태로 변환.   + 데이터 받아오기.     - 먼저, 데이터를 ReadFromFile.py 파일의 getData 함수를 통해서 받아온다.     - 이 데이터는 [이루어져, [(이루어지, W), (어, EC)]] 와 같이 이루어져 있다.   + 받아온 데이터를 index 로 수정.     - Make\_ExamplePair.py 파일에서 MakePair 함수를 호출한다.     - 위 함수는 input\_sent 와 output\_sent 및 두 개를 한 instance 당 한 개씩 list 묶은 것의 list인 pairs 를 return 한다.   + Variable 로 감싸기     - Make\_ExamplePair.py 파일에서 variableFromPair 함수를 통해서 Variable로 감싼다.     - 위 함수는 한 sentence 씩 변환하고 return 해준다.     - 따라서, main에서 모든 instance를 for loop 을 돌면서 variableFromPair 함수를 호출하도록 한다. | | |
| 작성자 : 홍태석 | 일자 : 2018년 3월 2일 |  |
| 점검자 | 일자 |