**Weekly report**

날짜 : 2022-04-07

연구원 : 최윤석

* 수행결과

▶ 개체명 인식기 모델이 받아들일 수 있는 최대 문장길이를 품사 모델과 같은 100으로 재 학습

▶ 개체명 인식기 모델과 품사 모델을 h5파일로 변환해 저장



▶ 개체명 인식기, 품사 모델을 이용해 단어 단위로 데이터셋 만들기

텍스트이(가) 표시된 사진

자동 생성된 설명

Figure 1. 개체명 인식 + 품사모델을 이용해 단어별로 예측한 결과

* 수행계획

▶ 들어오는 문장이 100자가 넘으면, 넘어가는 단어에 대해서는 예측을 못하는데, 신문기사 데이터들은 기본적으로 다 100자를 넘기기에, 신문기사 데이터를 문장 단위로 쪼개는 함수 개발

▶ 고유명사와 개체명 인식할 때, 단어에 조사가 붙는 경우가 많은데 (ex: 최윤석은 – NNP – PER) 이를 해결할 방안찾기

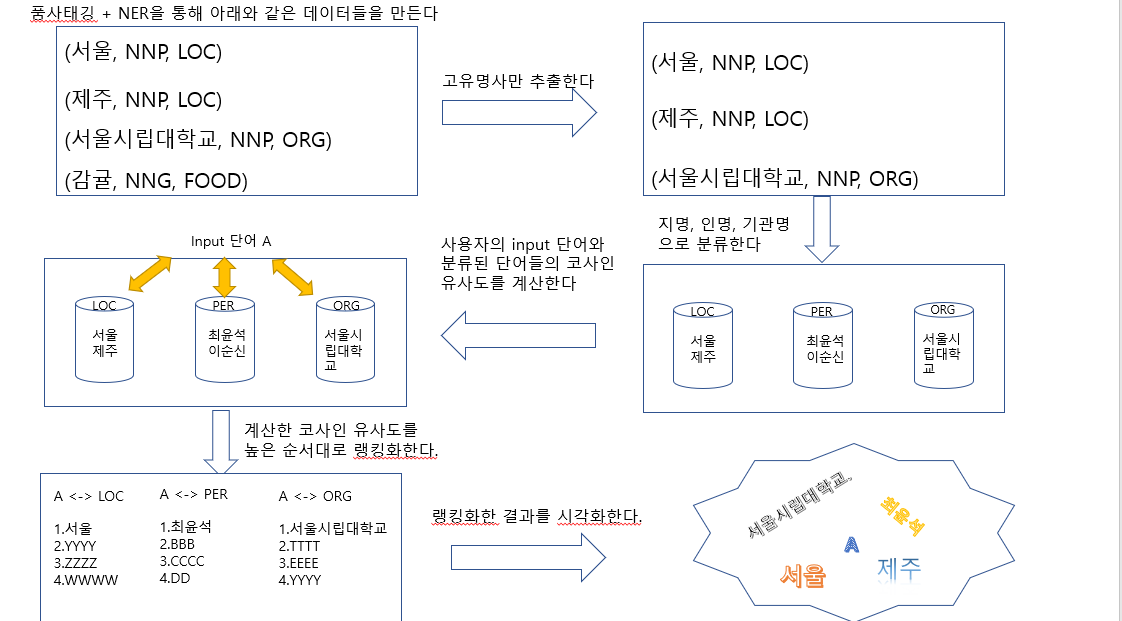


Figure 2. 앞으로의 진행 흐름

* 기타사항

▶ 카테고리별로 고유명사가 잘 나오는지 UI를 먼저 완벽하진 않지만 먼저 테스팅해보기 – 조사제거는 UI가 먼저 어느정도 완성된 후에 진행

▶ YOLO를 포함한 이미지 분류 프레임을 이용해 교수님이 보내주신 이미지 데이터 분류

▶