**캡스톤디자인 주간 보고서**

|  |  |
| --- | --- |
| 팀원 | 이예준, 박기홍 |
| 주제 | 미리내 – ML based Korean Educational Reconjugator using seq2seq |
| 기간 | 2020. 11. 16. ~ 2020. 11. 22. |
| 진  행  내  용 | 1. **월요일 회의 (11.16)**  * 기존에 Training pairs를 뽑아내던 AI-hub corpus는 일부였지만 AI-hub 전부와 KAIST corpus들까지 합쳐서 CTO분께서 json파일로 새롭게 생성해주셨다. 또한 모델의 하이퍼 파라미터들을 변경해가며 모델의 성능개선을 하는 방향을 제시해주셨다.  1. **ExtractorFromJson개발 및 Back&Front for DEMO (11.16 ~ 11.18)**  * 이예준 : 대부분 Mirinae pipeline을 통해 나오는 training data들은 json파일의 형태로 전달된다. 이후로도 추가될 데이터들을 감안했을 때 이 데이터들을 추출하는 방식을 형식화할 필요성을 느꼈다. 또한 이 과정에서 조금씩 불완전한 부분의 corpus들에 한해서도 정제해 줘야하고 상대적으로 json보다 읽는 속도가 빠른 txt파일로 추출해야 했다. * 박기홍 : 미리내 프로젝트를 서비스할 back-end서버를 flask를 사용하여 구현하기로 하였다. 따라서 flask library에 대하여 스터디를 한후 서비스에 필요한 API를 설계하였다. 설계를 바탕으로 서버를 구현하였고, 구현된 API들을 테스트하였다. front-end단은 react를 사용하여 구성하기로 하였다. 다만 아직 react framework에 대한 skill이 부족하다고 판단되어 1주간의 스터디 기간을 갖기로 하였다. 서비스의 명세를 바탕으로 인터페이스를 설계하고 충분한 스터디를 진행한 뒤 구현하기로 하였다.  1. **Apply LSTM & Display Attention mechanism (11.20 ~ 11.22)**  * 이예준 : 새로운 training pairs들을 통해 테스트를 진행했지만 이전보다 눈에 띄게 나아진 점은 찾기 힘들었다. 그래서 기존에 사용하던 GRU를 LSTM으로 변경하는 작업을 거쳤다. 하지만 Attention 부분에서 에러가 발생했는데 이를 잡지 못 하여 다음주에 CTO분께 조언을 구할 예정이다. 또한 Attention mechanism과 관련하여 맨 처음 verb-phrase가 튀는 현상에 대한 원인을 잡을 수 있을 것 같다고 생각하여 Attention을 matplotlib를 이용하여 Visualizing하는 데에 성공했다. 완전하게 이상적으로 적용된 것 같지는 않지만 다음 주차에 집중적으로 R&D할 예정이다. |