**Weekly report**

날짜 : **2023-06-01**

연구원 : **최윤석**

* **이전 수행계획**
  + **관계형 테이블 임베딩 :**
    - Value overlap + Semantic overlap + Embedding(EmbDI) 을 사용하여 테이블 융합이 가능한지 확인하는 나만의 아키텍처 연구
    - EmbDI의 임베딩을 개선하여 더 좋게 만드는 방법에 대해 연구
      * W2V 대신에 Elmo / Bert 등의 임베딩을 사용하는 방안
    - 학술대회에 제출할 아이디어 정리 및 수업시간에 발표할 PPT 작성
      * 1단계 : 임베딩에 관한 고민
        + 데이터 융합차원에서(조인) 범주형 + 조인만 가정하고 더 잘 찾아내는 방법
        + 수치형 + 조인만 가정하고 더 잘 찾아내는 방법
        + Ex. 주소를 가지고 조인할 때 ) 값들의 포맷이 천차만별일텐데 이를 해결하는 임베딩 방법
  + Partially하게 임베딩하는 방법 -> 풀 네임을 임베딩하기 보다 부분만을 임베딩 하는
  + **객체지향프로그래밍 TA**
    - 객체지향프로그래밍 과제4 채점
    - 객체지향프로그래밍 중간고사 점수 온라인 강의실에 등록
* **수행결과**
* **관계형 테이블 임베딩** : (임베딩 / 융합 가능성 파트 분류해서 작성).
  + 학술대회에 제출할 아이디어 정리 및 PPT 작성하여 수업시간에 발표 완료
    - Column value에 Named Entity Recognition을 사용하여 각 컬럼을 대표하는 Named Entity를 알아내고, 이를 활용하여 두 범주형 컬럼간의 조인 가능성을 알아내는 방법
* **객체지향프로그래밍 TA**
  + 과제4 채점완료 및 엑셀에 점수 기입완료
  + 중간고사 점수 온라인 강의실에 등록완료
* **수행계획**
  + **관계형 테이블 임베딩 :**
    - Named Entity Recognition할 문장을 만들기 위해 Column name + is + Column value 구조에서 추가적으로 덧붙여 문장 내 단어의 수를 늘릴 방법 고민 (메타데이터 활용 등..)
    - 범주형 – 범주형 컬럼간의 조인 아이디어를 구현한 코드 작성
* **기타사항**
  + **리포트 작성 시, 논문의 내용정리 같이 남이 한 부분에 대한 설명은 자세하게 하지 말고 그 중에서 내가 얻은 아이디어나 내 연구에 관한 부분을 자세하게 설명할 것**
  + **테이블 간 조인 / 유니온 하기위한 컬럼을 찾는 문제에 관해 생각 (Elmo / Bert / Word2Vec) 등도 생각 / 내가 손으로 하는 것을 자동화하는 방법으로 생각**
  + **데이터 융합 프로세스**





혼합형 데이터셋 예시