**Weekly report**

날짜 : **2023-05-11**

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
    - Dataset Discovery in Data Lakes 논문 읽고 정리
    - Search and Join Algorithms for Tables in Data Lakes 논문 읽고 정리
    - 학술대회에 제출할 아이디어 정리 (한국전자거래학회)
  + **객체지향프로그래밍 TA**
    - 중간고사 채점완료
* **수행결과**
* **관계형 테이블 임베딩** : (임베딩 / 융합 가능성 파트 분류해서 작성).
  + Dataset Discovery in Data Lakes 논문 읽고 정리
    - 두 컬럼이 유사한 컬럼인지 알아내기 위해 Column name / value / format / embedding / domain 의 5가지에 대해 각각 유사한지 계산하고 그 결과를 가중합을 하여 최종적으로 두 컬럼이 유사한지 알아내는 것에 관한 논문. 자세한 내용은 다음주 수업시간에 PPT로 발표할 예정
* **객체지향프로그래밍 TA**
  + 중간고사 채점완료 및 엑셀에 점수 기입완료
* **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
    - Search and Join Algorithms for Tables in Data Lakes 논문 읽고 정리
    - 학술대회에 제출할 아이디어 정리 (한국전자거래학회)
    - Dataset Discovery in Data Lakes 논문에 대해 수업시간에 발표할 PPT 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