**Weekly report**

날짜 : **2023-11-02**

연구원 : **최윤석**

* **이전 수행계획**
  + **관계형 테이블 임베딩 :**
    - CNE-join의 성능을 개선할 방법 고민
      * 조인하는 테이블의 개수를 3개 이상으로 확장해보기, 최적의 테이블 개수는 무엇인가? 이를 확인할 척도에 대해서도 고민
        + CNE-join의 결과로 나온 2개의 조인 가능한 테이블쌍 중에서 A-B, B-C 같이 테이블이 겹치는 쌍을 활용해 n-join-pair를 생성하는 알고리즘 및 이에 대한 융합도 산정 방안 고민 및 **소스코드 작성**(이전 CNE-join과의 일관성이 중요함)
      * 시간 단축 문제는 가장 나중으로 진행
    - 박찬훈 연구원과 DataMap 개발(Type A, B)
    - Measuring and Predicting the Quality of a Join for Data Discovery 논문이 참고한 문헌을 읽고 n-join-pair 알고리즘 아이디어 얻기
    - 박찬훈 연구원과 분담하여 빅데이터자료구조 중간고사 채점

* **수행결과**
  + **관계형 테이블 임베딩 :**
  + The Next 5 Years: What Opportunities Should the Database Community Seize to Maximize its Impact? 논문
    - n-join-pair와 CNE-join과 관련이 없는 논문
  + Discovering Related Data At Scale
    - CNE-join과 같이 조인 가능한 컬럼쌍을 찾는데 초점을 둔 논문
    - 조인 가능한 컬럼쌍인지 알기 위해 컬럼명이 유사한지 파악하고 컬럼값의 자카드 유사도 계산 등을 진행함.
    - 모델을 평가할 때 조인 가능한 컬럼쌍의 정답을 사용자들이 SQL에서 조인 쿼리로 입력한 컬럼쌍에서 얻었다는 점이 흥미로웠음.
* **빅데이터 자료구조 중간고사 채점 :**
  + 박찬훈 연구원과 분담하여 중간고사 채점중
* **수행계획**
  + **관계형 테이블 임베딩 :**
    - CNE-join의 성능을 개선할 방법 고민
      * 조인하는 테이블의 개수를 3개 이상으로 확장해보기, 최적의 테이블 개수는 무엇인가? 이를 확인할 척도에 대해서도 고민
        + CNE-join의 결과로 나온 2개의 조인 가능한 테이블쌍 중에서 A-B, B-C 같이 테이블이 겹치는 쌍을 활용해 n-join-pair를 생성하는 알고리즘 및 이에 대한 융합도 산정 방안 고민 및 **소스코드 작성**(이전 CNE-join과의 일관성이 중요함)
      * 시간 단축 문제는 가장 나중으로 진행
    - 박찬훈 연구원과 DataMap 개발(Type A, B)
    - Measuring and Predicting the Quality of a Join for Data Discovery 논문이 참고한 문헌을 읽고 n-join-pair 알고리즘 아이디어 얻기
    - 박찬훈 연구원과 분담하여 빅데이터자료구조 중간고사 채점
    - 조인 테이블 개수를 3,4개 이하로 테이블을 임베딩한 벡터들을 2차원 공간에 시각화하기
* **기타사항**