**Weekly report**

날짜 : **2023-12-27**

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
    - Measuring and Predicting the Quality of a Join for Data Discovery 논문이 참고한 문헌을 읽고 n-join-pair 알고리즘 아이디어 얻기
    - CNE-join 결과로 나온 조인 가능한 테이블쌍 정보를 활용해 최대 4개의 테이블을 조인한 테이블을 벡터로 임베딩
      * 조인한 테이블의 임베딩 벡터와 원천 테이블의 임베딩 벡터간의 거리를 활용해 n-join-pair에 관한 insight를 얻기
        + 3개의 테이블을 조인한 테이블의 임베딩 벡터와 원천 테이블의 임베딩 벡터간의 거리의 평균을 조인 결과의 순위를 나누는 기준으로 사용하고 있는데, 이것이 유효한지 성능평가를 진행
        + 추가로 3개의 테이블을 조인한 테이블의 임베딩 벡터와 원천 테이블의 임베딩 벡터간의 거리 중 가장 먼 거리를 조인 결과의 순위를 나누는 기준으로 사용하여 평균과의 성능 비교를 진행
    - 박찬훈 연구원과 DataMap 개발(Type A, B)
      * 박찬훈 연구원한테 DataMap 코드 인수인계 받기
    - 자동 융합 시나리오에 대한 명세를 작성해보기(use case diagram)
      * PPT로 작성
      * 융합을 할 때 Join을 할 지 Union을 할 지 확인하는 방법 – 나는 Join 파트
    - Table2Vec 논문을 참고하여 n-join-pair에 대한 아이디어 얻기
    - WITC 2024에 제출할 논문 작성하기(~1/15)
* **수행결과**
  + **관계형 테이블 임베딩 :**
    - 조인한 테이블의 임베딩 벡터와 조인에 사용된 테이블의 임베딩 벡터간의 거리를 활용해 n-join-pair에 관한 insight를 얻기
      * 3개의 테이블을 조인한 테이블의 임베딩 벡터와 원천 테이블의 임베딩 벡터간의 거리의 평균을 조인 결과의 순위를 나누는 기준으로 사용하는 방식의 성능평가 진행
        + Containment를 사용해 성능평가를 진행한 결과 1티어 조인에서 5티어 조인으로 가면서 containment 값이 점점 높아지다가 다시 9티어 조인까지 containment 값이 낮아지는 경향을 보임
* WITC 2024에 제출할 논문 작성중
  + 현재 초록과 서론까지 작성완료
* **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
    - Measuring and Predicting the Quality of a Join for Data Discovery 논문이 참고한 문헌을 읽고 n-join-pair 알고리즘 아이디어 얻기
    - CNE-join 결과로 나온 조인 가능한 테이블쌍 정보를 활용해 최대 4개의 테이블을 조인한 테이블을 벡터로 임베딩
      * 조인한 테이블의 임베딩 벡터와 원천 테이블의 임베딩 벡터간의 거리를 활용해 n-join-pair에 관한 insight를 얻기
        + 3개의 테이블을 조인한 테이블의 임베딩 벡터와 원천 테이블의 임베딩 벡터간의 거리의 평균을 조인 결과의 순위를 나누는 기준으로 사용하고 있는데, 이것이 유효한지 성능평가를 진행
        + 추가로 3개의 테이블을 조인한 테이블의 임베딩 벡터와 원천 테이블의 임베딩 벡터간의 거리 중 가장 먼 거리를 조인 결과의 순위를 나누는 기준으로 사용하여 평균과의 성능 비교를 진행
    - 박찬훈 연구원과 DataMap 개발(Type A, B)
      * 박찬훈 연구원한테 DataMap 코드 인수인계 받기
    - 자동 융합 시나리오에 대한 명세를 작성해보기(use case diagram)
      * PPT로 작성
      * 융합을 할 때 Join을 할 지 Union을 할 지 확인하는 방법 – 나는 Join 파트
    - Table2Vec 논문을 참고하여 n-join-pair에 대한 아이디어 얻기
    - WITC 2024에 제출할 논문 작성하기(~1/15)
      * 초록을 2페이지로 늘린 논문 접수를 위한 버전을 먼저 작성
        + 대표 그림 하나, 대표 성능 평가 그래프 하나 추가하여 작성
      * Full paper는 초록 버전 이후로 작성
* **기타사항**