

# 9월 6일

또또 에러가 났다... 서버가 터져버렸다...

train data에 대한 cycle\_triple을 만드는 과정을 계속해서 진행하다보니 터짐, 어느순간 렉이 걸리고 느려짐 그러다 팽!!...

오늘 진행할 것은 서버 복구 후, train\_cycle\_triple을 만드는 것을 실행해보고

BM25와 DPR에서 doc가 없는 경우를 어떻게 해야할지 생각한다.

1. doc가 없다는 것은 사이클이 없다는 것이다. 해당 subgraph에는 사이클이 존재하지 않는다는 것이다.
2. 이럴 경우에는 cycle 정보를 전혀 고려할 수 없다는 것이다. 노드 개수가 200개라도 대부분이 isolate이고 그래프 형태도 ego graph(ego는 context node) 형태인 것임
3. 따라서 ranking 과정, reranking과정 모두를 생략한다.

이렇게 해서 진행하면, 사이클이 없는 그래프에 대해서는 사이클의 장점?, 영향을 이용할 수 없다.

그러면 만약 그 subgraph가 정답에 해당하는 subgraph이면 어쩌냐..

그래서 빈 리스트에 해당하는것이 정답에 해당하는 subgraph인지 아닌지 확인해본다.

self.train\_qids에서 inhouse\_train\_qids.txt에 있는 qids의 위치(인덱스), inhouse\_train\_qids.txt에 없는 qids의 위치(인덱스)를 저장한다. - > 해결

원래는 cycle triple 생성까지 전처리로 돌렸음

하지만 해보니 아래 네 과정을 모두 전처리로 해야함 슈발

1. Cycle triple → cycle passage
  - concept\_ids, cycle\_triple 데이터 필요
2. cycle passage - > BM25, DPR
  - qa\_context, cycle\_passage 데이터 필요

3. Reranking

4. selected\_cycle\_triple

- cycle\_triple, top\_N\_rerankindex 데이터 필요