후속질문추천알고리즘 코드 인수인계서 최지운

1. DB 질문 답변에 대해서 gpt api 자체 제작 프롬프트 사용해서 핵심 키워드를 뽑아서 DB에 새로운 행으로 저장하는 코드 -> answer\_설명키워드추출gptAPI.py, question설명키워드추출.py
2. 들어온 타겟질문을 답변 rag가 잘되게 + 오타가 있으면 제대로된 질문으로 대체하기위한 1차적으로 DB에있는 질문으로 변환하는 코드

->1차\_문장by문장\_유사도 기반 질문추천(자동완성기능).py 🡺 타겟질문과 전체질문을 모두 fit\_transform함

->test.py 🡺 타겟질문은 놔두고 미리 fit\_transform한 vectorizer를 사용하여 계산

1. 모든 질문 / 답변 에 대해서 word2vec 학습시켜 키워드 계산할수있게하는 코드

->word2vec\_test.py

1. 변환된 타겟질문 keyword vs DB에서 이 질문과 매칭되는 답변 keyword 중 겹치는 것은 답변 keyword에서 제외시킴(이미 설명했다고 판단)

->테스트 코드 🡺define\_recomm\_keyword\_test.py

->메인 코드 🡺 2차\_usingGPTapi\_답변의 핵심키워드 추출.py 코드 중 define\_recomm\_keyword1 함수

1. 제외시킨후 답변의 키워드중 질문과 가장유사도가 높은 3개를 최종키워드로 선택

-> 테스트 코드 🡺define\_recomm\_keyword\_test.py

->메인 코드 🡺2차\_usingGPTapi\_답변의 핵심키워드 추출.py 코드 중 define\_recomm\_keyword2 함수

1. 최종 3개의 키워드에 대해 모든 답변의 키워드들과 유사도를 구해서 0.7 \* 평균 + 0.3 \*표준편차로 최종 score를 구한뒤 가장 score가 높은 최상위 3개의 답변 에 매칭되는 질문을 최종적으로 추천한다.

->2차\_usingGPTapi\_답변의 핵심키워드 추출.py 코드 중

Final\_generate\_recommend 함수

\*참고\*

코드에 대한 설명과 파라미터 설명 등 은 주석으로 달아놓았습니다.

데이터 프레임을 바꾸거나 전처리만 새롭게 해서 사용하시면됩니다.

2차 usingGPTapi\_답변의 핵심 키워드 추출.py 코드는 어디에 사용될지 몰라서, 메인 부분이 없이 함수 구현만 해놓은 상태라 플라스크에 각 함수 기능을 붙이는 용도로 사용하시면 될것같습니다.