basic RAG는 사용자의 질문에 해당하는 문서의 의미에 포커싱을 한다. parent document , self query, .. 등을 해도 특정 문서의 의미에 집중하는 것은 다르지 않다.

우리가 RAG를 사용하는건 그만큼 질문자의 말을 context로 하는 latent space와 데이터셋이 임베딩된 공간의 단어들간의 혹은 문장들(sentencetokenizer)간의 거리가(euclidean/metric/topological..) 서로 다르기 때문일거다.

graph rag는 그런 점에서 basic rag보다 나은 점을 보인다. 단어들간의 관계 자체를 embedding하니 특정 맥락에 맞는 embedding이 되는걸로 볼수 있다.

하지만, 맥락은 두 이미지를 interpolation을 하듯이 입력데이터의 문장에 약간의 변형을 가하면 입력데이터의 중심맥락과 외부의 어떤 데이터의 맥락이 뒤죽박죽 섞일텐데 이걸 해결하기에는 어려워보인다.

cauchy sequence를 빌리면,

a라는 코시수열이 있을때,
e.g) a_1 = 나는 밥을 먹었다 a_2 나는 어제 밥을 먹었다 a_3 나는 어제 관광지에서 밥을 먹었다 a_4 나는 어제 매우 유명한 관광지에서 밥을 먹었다.

단어의 수가 단조적으로 증가하고 의미가 단조적으로 심화되서 i= 1부터 어떤 충분히 큰 n보다 크다고 볼수도 있으나 갑자기 a_4부터 맥락이 달라지면 수렴이 되지 않듯이, embedding space의 embedding값도 맥락값과 괴리가 생긴다.