

Masking

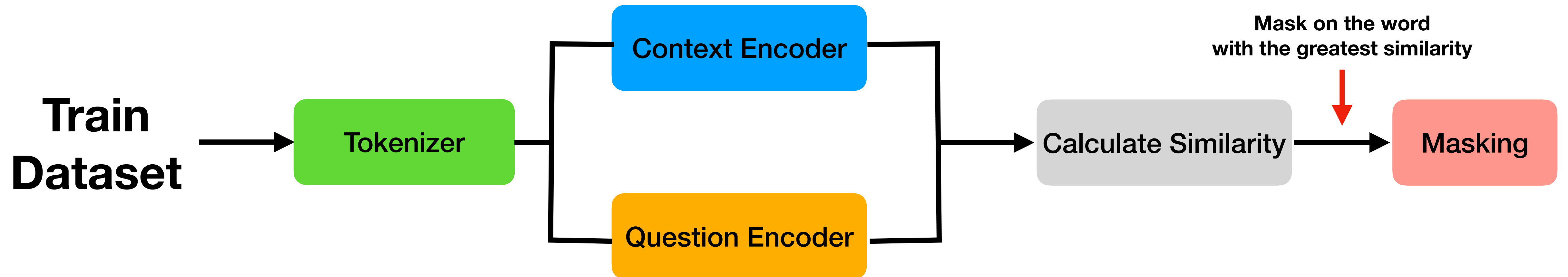
개념부터 코드까지 핫이슈!

Team KiYOUNG2, Jeangyu Heo

Idea

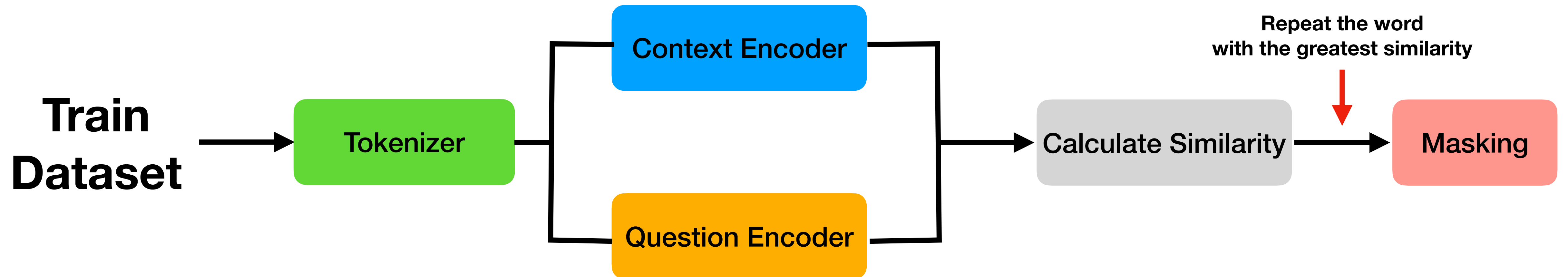
- Masking을 임의로 해주는 것보다 의미 있게 해주는게 좋지 않을까??
- Masking을 answer를 제외한 Question과의 유사도가 높은 Token이 포함된 단어에 해주면 조금 더 답에 잘 수렴하지 않을까?
- 그렇다면, 유사도가 높은 Token이 포함된 단어를 반복해주면 더 수렴하기 어렵지 않을까?
- 이 데이터셋으로 Curriculum Learning을 할 수 있지 않을까??

Masking Flow Chart



⌘ Encoder means pretrained Model




Adding Flow Chart



⌘ Encoder means pretrained Model

Curriculum Learning

	Tokens	이순신	##은	조선	중기	무신	##이다.
 Word-level High Score Masking		이순신	##은	조선	중기	[MASK]	
 Basic		이순신	##은	조선	중기	무신	##이다.
 Word-level High Score Adding		이순신	##은	조선	중기	무신무신이다.	

-  Easy
-  Normal
-  Hard

실습

- 코드를 통해서 어떻게 구현되었는지 확인해보자!
- https://drive.google.com/file/d/11z1S77y4vfXHdmab49ijcEDiaC-hTSp_/view?usp=sharing