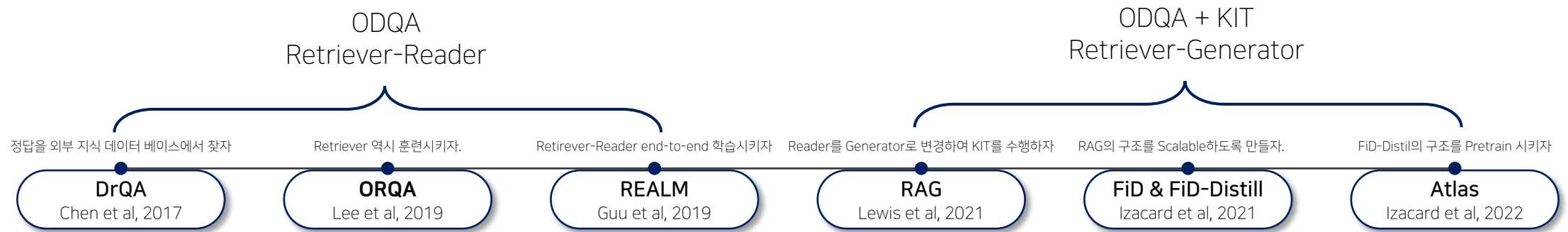
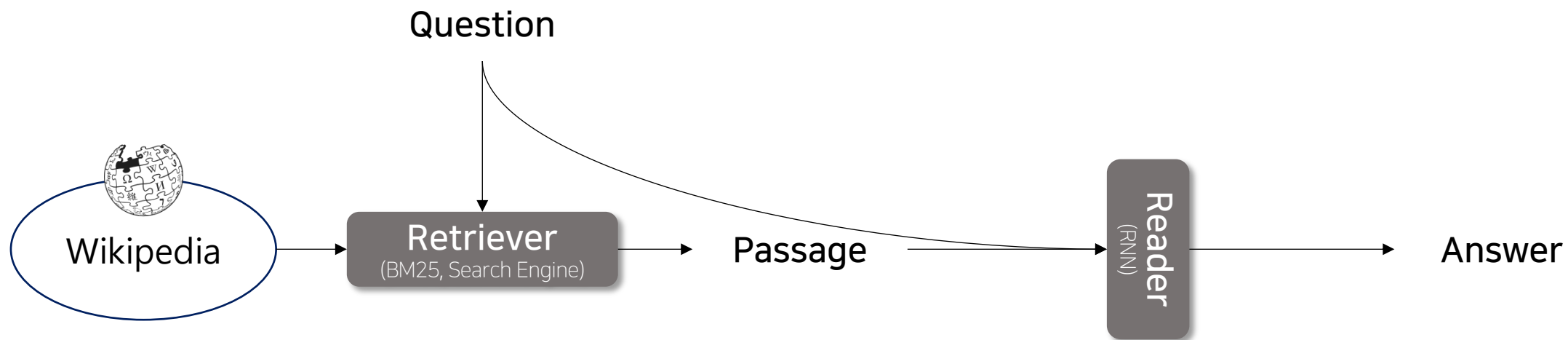


01 From ODQA to KIT

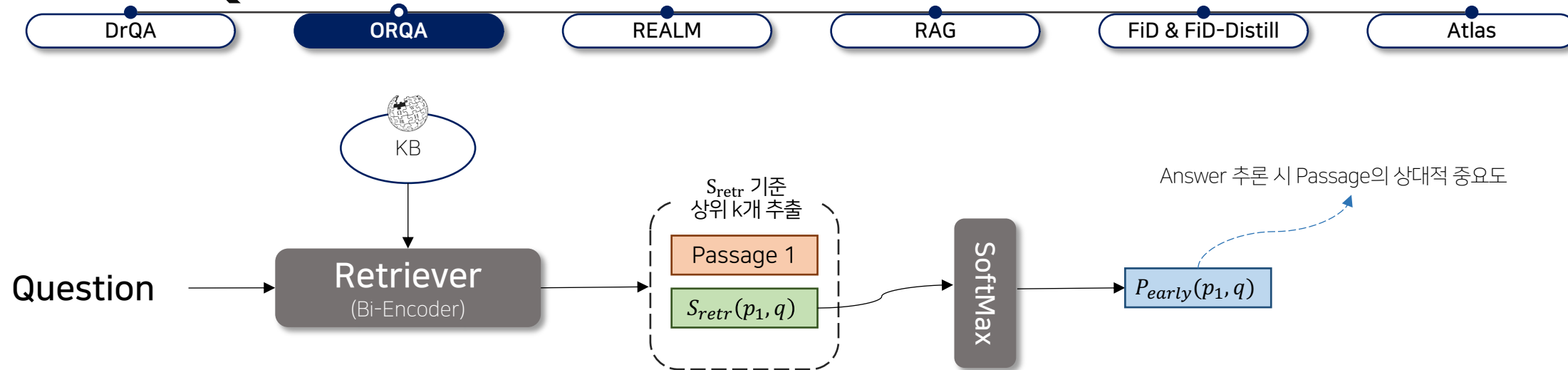


- Open-Domain Question Answering 태스크의 시작부터 최근 Knowledge Intensive Task까지 각 연구의 흐름
- Knowledge Intensive Task 발전 과정에서 각 논문의 주요 Contribution
 - RAG는 이전 ODQA 연구를 연결하여 Knowledge Intensive Task로 확장
 - RAG 이후 연구인 FiD ~ Atlas는 RAG를 기반으로 삼고 있음



- 고정된 Passage에 대해 Reader 학습을 통해 Answer 산출
- Knowledge Base에서 BM25, TF-IDF, Wikipedia Search Engine 등을 통해 빠르게 Passage를 탐색
- Reader가 Passage에서 Answer Span을 추론

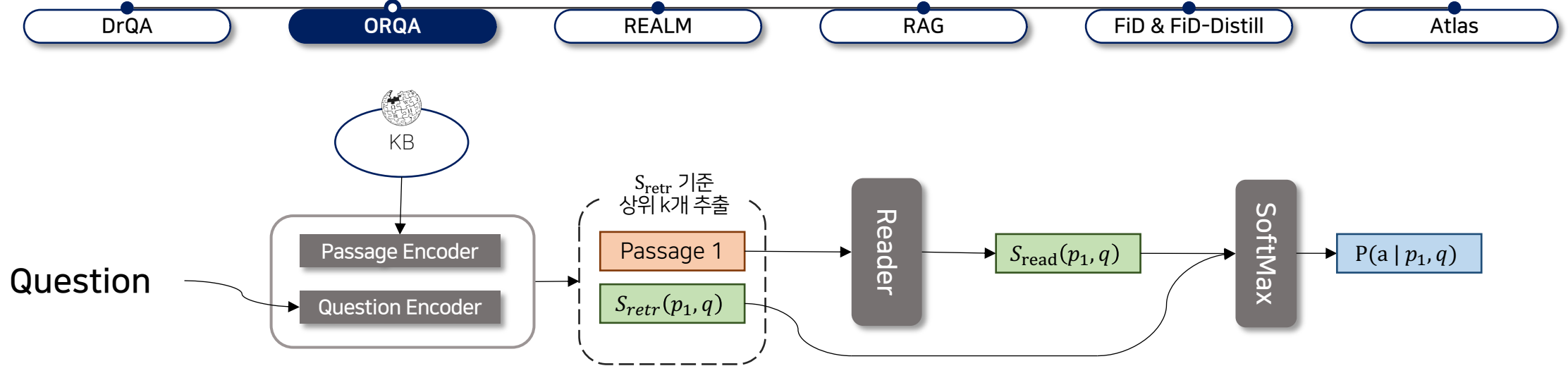
Retriever에 대한 훈련이 불가하여 Passage 개선이 불가능



- Retriever 학습에 필요한 신호(signal)는 ORQA에 의해 처음 정의
- Retriever는 두가지 Encoder로 구성
 - Passage Encoder : Pretrain(ICT) 을 통해 학습
 - Question Encoder : Pretrain과 Finetune을 통해 학습
- Finetune 시 두가지 방식으로 Retriever를 훈련
 - Passage Retrieval 과정 (P_{early})
 - Answer Span이 포함된 Passage의 likelihood 최대화 = $\max P(p|q) = \frac{\exp(S_{\text{retr}}(p, q))}{\sum_k \exp(S_{\text{retr}}(p_k, q))}$

01

From ODQA to KIT



- Retriever 학습에 필요한 신호(signal)는 ORQA에 의해 처음 정의
- Retriever는 두가지 Encoder로 구성
 - Passage Encoder : Pretrain(ICT) 을 통해 학습
 - Question Encoder : Pretrain과 Finetune을 통해 학습
- Finetune 시 두가지 방식으로 Retriever를 훈련

➢ Answer Span Prediction과정 (P)

➢ Answer Span의 likelihood 최대화 = $P(a|p, q) = \frac{\exp(S_{\text{retr}}(p, q) + S_{\text{read}}(a, p, q))}{\sum_k \exp(S_{\text{retr}}(p_k, q) + S_{\text{read}}(a, p_k, q))}$

01

From ODQA to KIT

DrQA

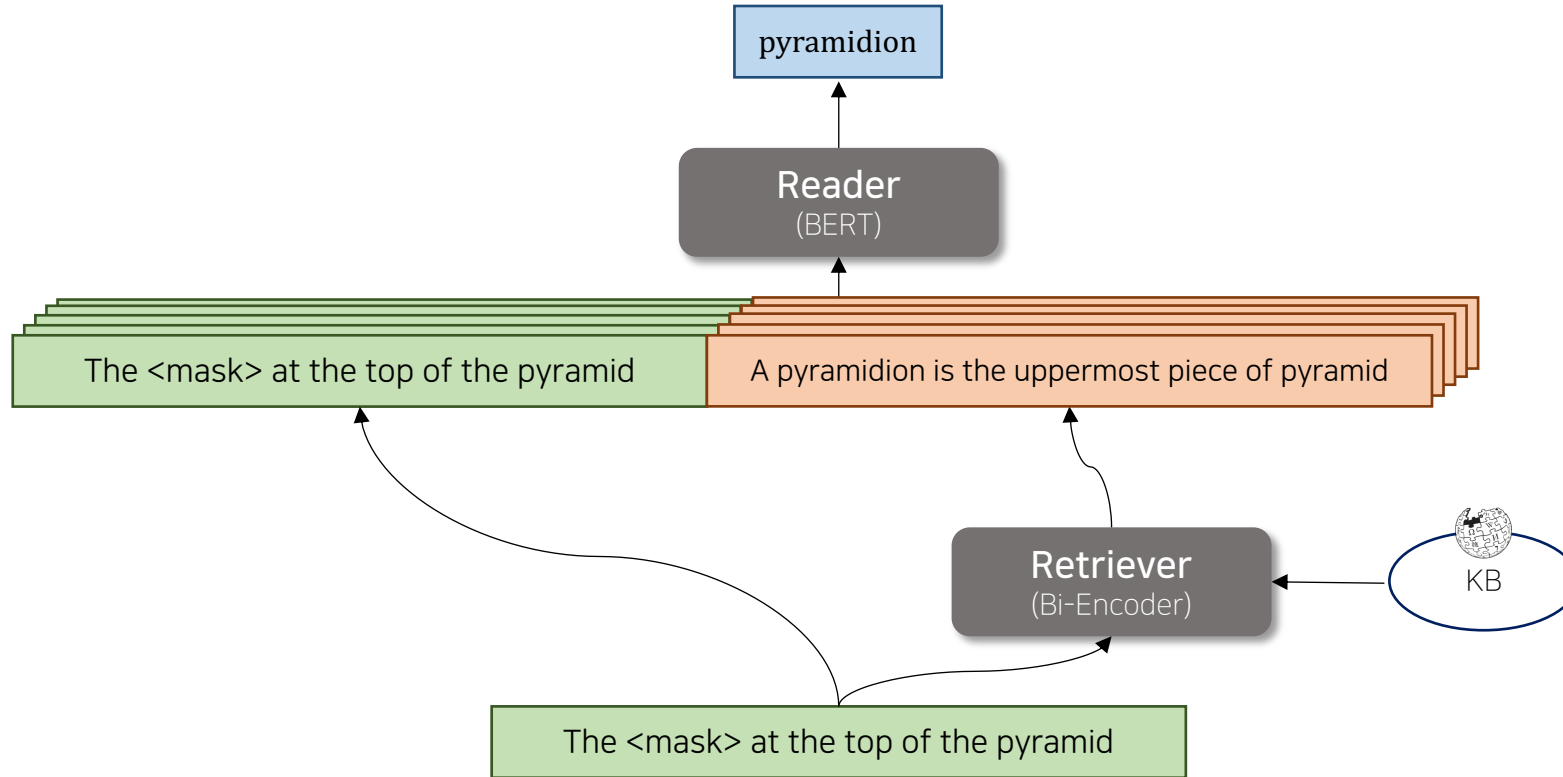
ORQA

REALM

RAG

FiD & FiD-Distill

Atlas



- REALM은 MLM을 Pretrain 태스크로 사용
- ORQA와 다르게 Retrieval 시 별도의 Annotation 사용 X
- Answer로 사용될 가능성이 높은 entity(인명, 지명 등)을 masking

01 From ODQA to KIT

DrQA

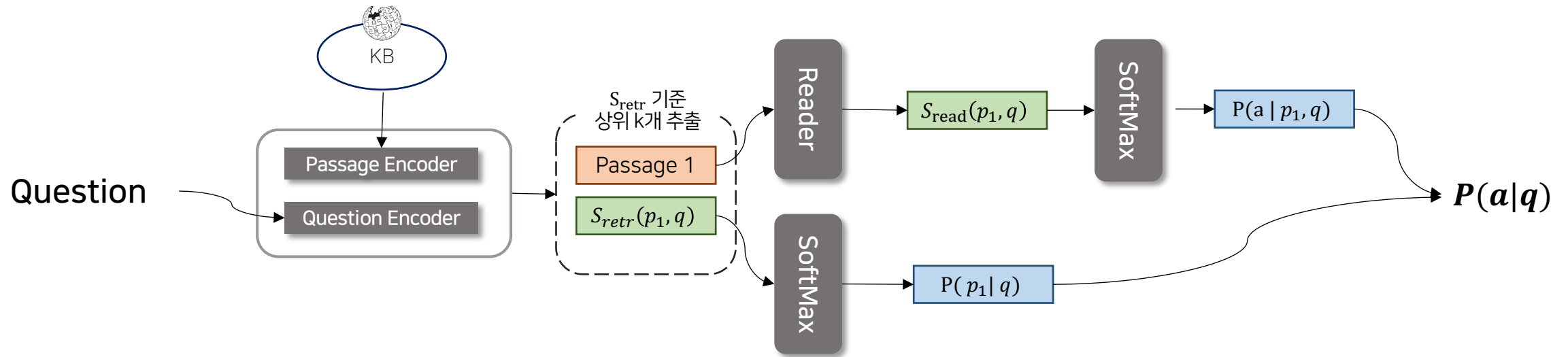
ORQA

REALM

RAG

FiD & FiD-Distill

Atlas



- REALM은 Pretrain 시 MLM을 이용하여 Passage에 대한 Annotation 사용 X
- Reader와 Retriever에 의해 산출되는 likelihood 를 이용하여 Answer Likelihood 계산
 - Answer Likelihood : $P(a|q) = \sum_k P(a|p_k, q) P(p_k|q)$

Passage의 확률분포와 Answer에 대한 Passage의 likelihood로 분해
- Finetune 시에는 Passage에 대한 Annotation을 이용

01

From ODQA to KIT

DrQA

ORQA

REALM

RAG

FiD & FiD-Distill

Atlas

- REALM은 Pretrain 시 MLM을 이용하여 Passage에 대한 Annotation 사용 X
- Reader가 MLM 수행 시 직접 Retriever에게 Signal을 전달하도록 변경

➤ Answer Likelihood :

$$P(a|q) = \sum_k P(a|p_k, q)P(p_k|q)$$

➤ Answer log likelihood의 Retriever에 대한 편미분

$$\nabla \log p(a | q) = \sum_{p \in \mathcal{P}} r(p) \nabla f(q, p)$$

$$r(p) = \left[\frac{p(a | p, q)}{p(a | q)} - 1 \right] p(p | q)$$

- $p(a|p, q)$: 특정 passage가 사용됐을 때, answer에 대한 likelihood
- $p(a|q)$: 랜덤한 passage를 latent variable로 사용했을 때, answer에 대한 likelihood

$$p(a|p, q) > p(a|q) \quad \Rightarrow \quad \frac{p(a|p, q)}{p(a|q)} - 1 > 0 \quad \Rightarrow \quad r(p) > 0$$

$$p(a|p, q) < p(a|q) \quad \Rightarrow \quad \frac{p(a|p, q)}{p(a|q)} - 1 < 0 \quad \Rightarrow \quad r(p) < 0$$

01

From ODQA to KIT

DrQA

ORQA

REALM

RAG

FiD & FiD-Distill

Atlas

- REALM은 Pretrain 시 MLM을 이용하여 Passage에 대한 Annotation 사용 X
- Reader가 MLM 수행 시 직접 Retriever에게 Signal을 전달하도록 변경

➤ Answer Likelihood :

$$P(a|q) = \sum_k P(a|p_k, q)P(p_k|q)$$

➤ Answer log likelihood의 Retriever에 대한 편미분

$$\nabla \log p(a | q) = \sum_{p \in \mathcal{P}} r(p) \nabla f(q, p)$$

$$r(p) = \left[\frac{p(a | p, q)}{p(a | q)} - 1 \right] p(p | q)$$

랜덤한 Passage 사용 대비 특정 Passage 사용 시 얻을 수 있는 likelihood 개선 비율

query 에 대해 passage에서 얻을 수 있는 likelihood 개선 기대값

Retriever 업데이트 시 필요한 정보는 “Answer 추론 시 Passage가 도움이 되는 정도”

01 From ODQA to KIT

DrQA

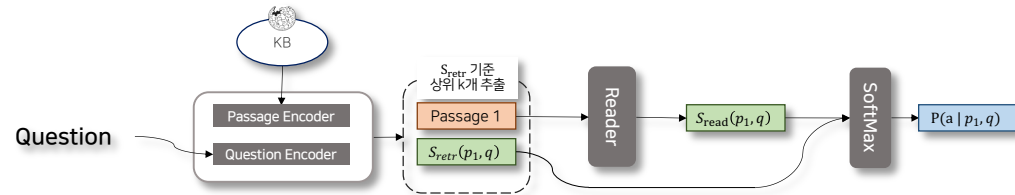
ORQA

REALM

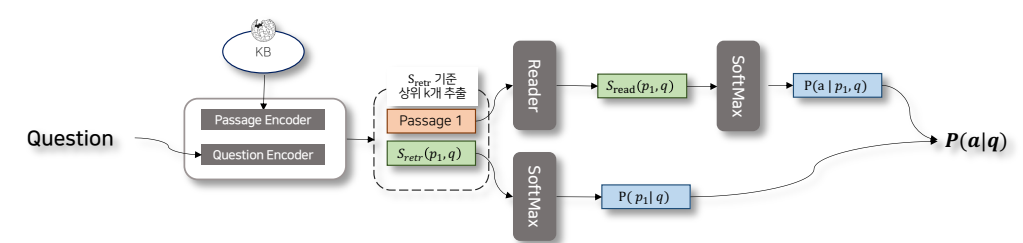
RAG

FiD & FiD-Distill

Atlas

ORQA


$$P(a|p, q) = \frac{\exp(S_{\text{retr}}(p, q) + S_{\text{read}}(a, p, q))}{\sum_k \exp(S_{\text{retr}}(p_k, q) + S_{\text{read}}(a, p_k, q))}$$

REALM


$$P(a|q) = \sum_k P(a|p_k, q)P(p_k|q)$$

$$\begin{aligned} \max P(a | q) &= \max P(a, p | q) \\ &= \max P(a | p, q)P(p | q) \end{aligned}$$

REALM을 통해 Passage를 Latent Variable로 도입한 모델링이 가능