

# $R^2$ FID: Joint Reranker 기반 Fusion-In-Decoder를 이용한 오픈 도메인 테이블 질의 응답

이성민, 박은환, 나승훈, 서대룡, 전동현, 강인호

cap1232@jbnu.ac.kr

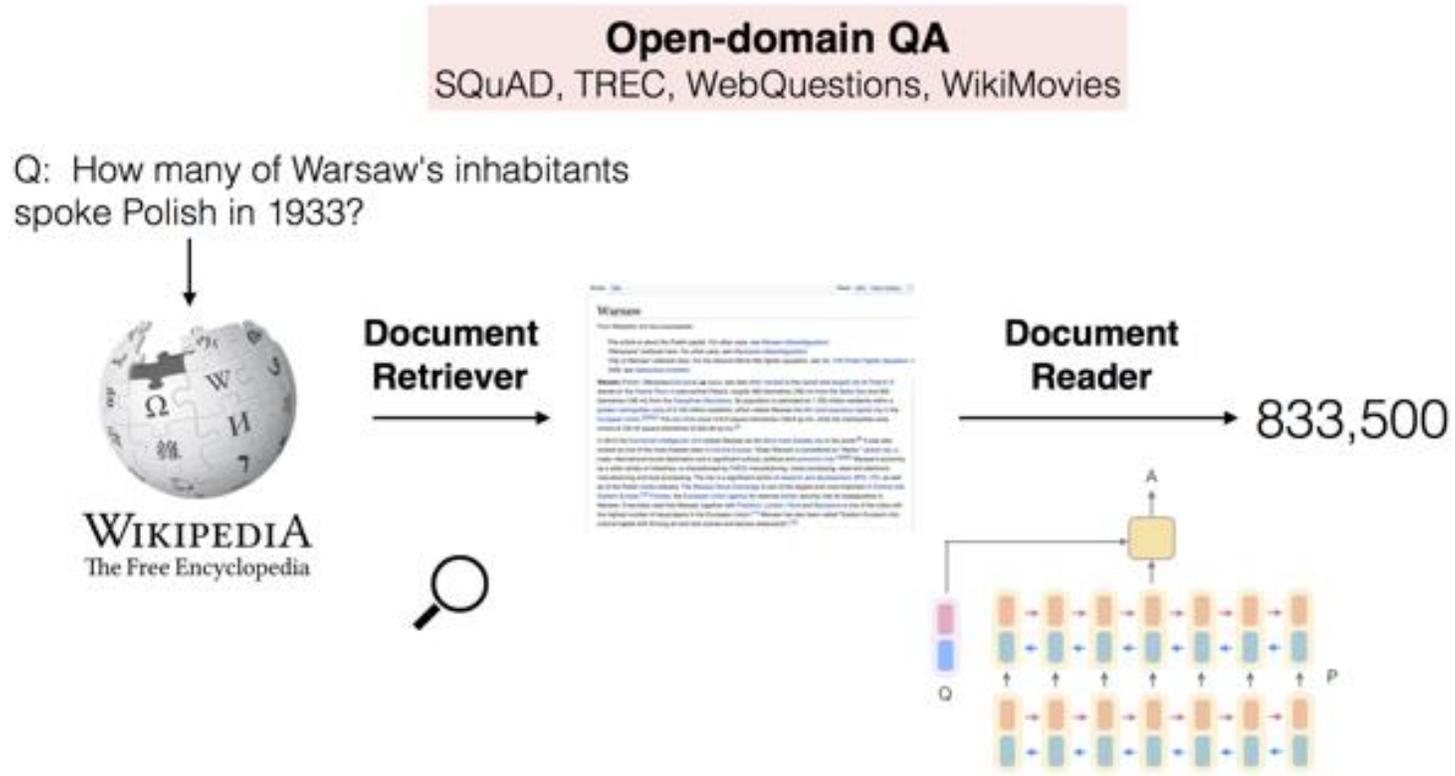


**NAVER**

# Introduction

# Introduction

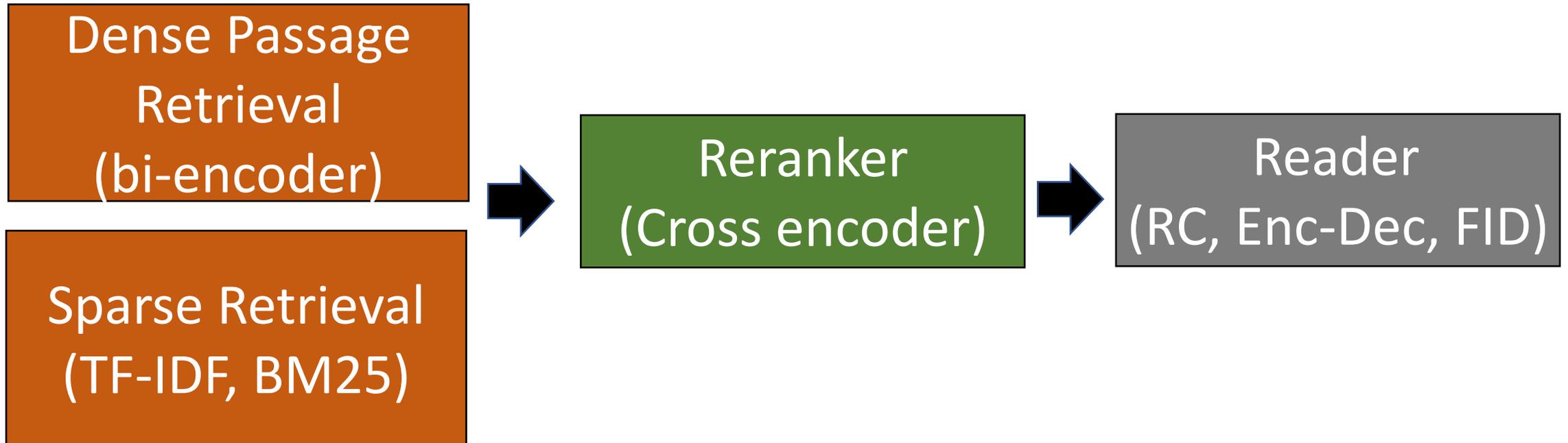
오픈 도메인 질의 응답은 명시적인 단서가 제공되지 않아 위키피디아나 웹, 지식베이스 같은 데이터 자원을 기반으로 자연어 질문에 답변하는 Task (e.g. IRQA, KBQA )



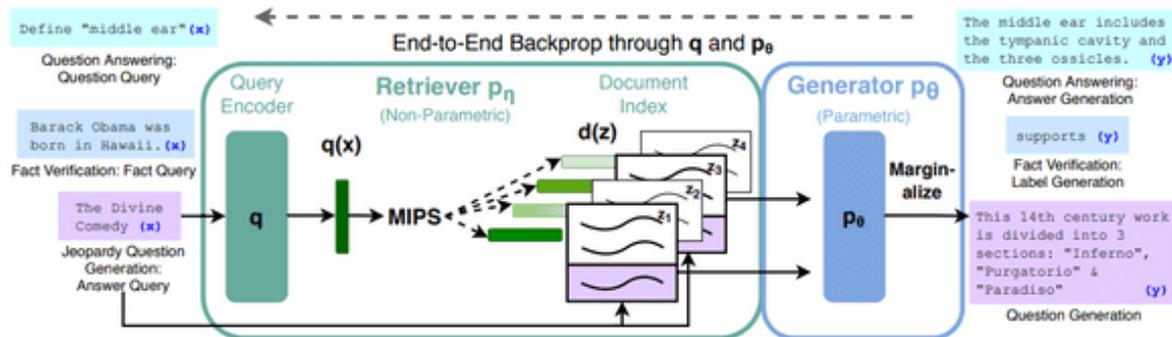
# Introduction

(검색 기반) 오픈 도메인 QA를 위한 방법들

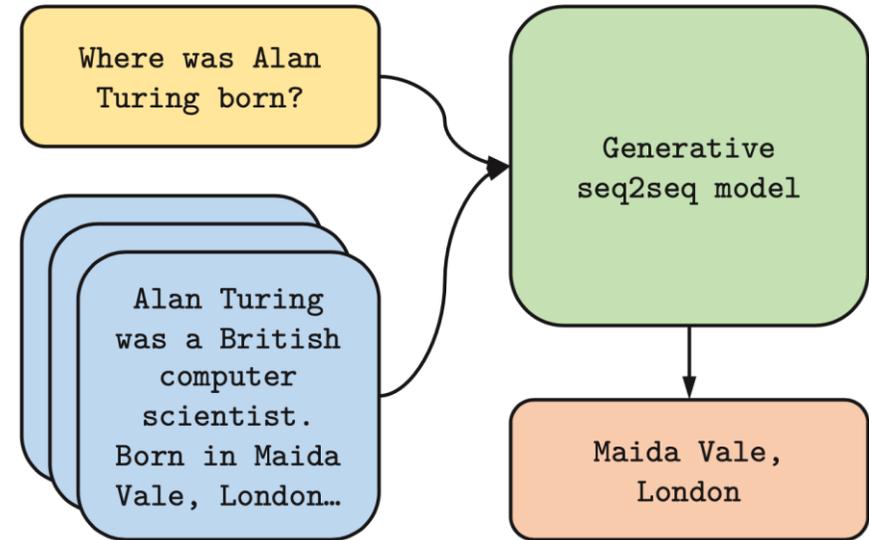
- Retrieval-Reranker-Reader



# Introduction

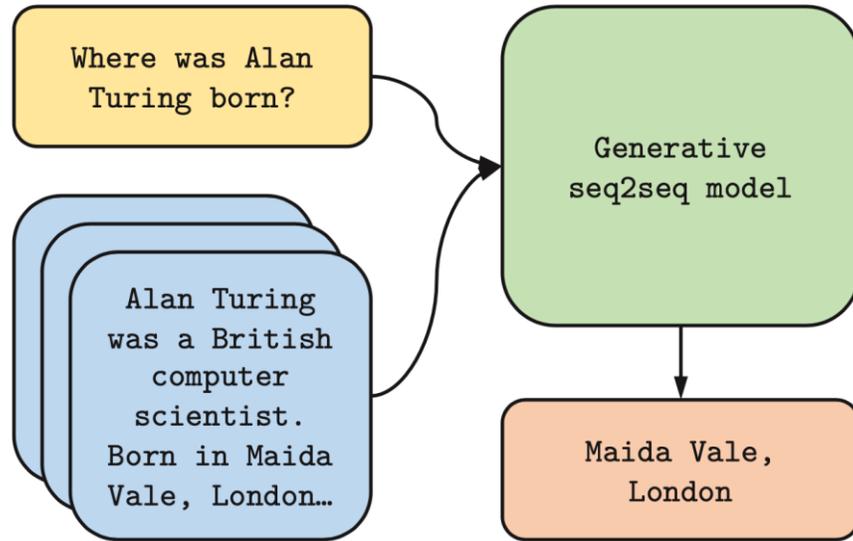


Retrieval augmented generation(RAG)



Fusion-in-Decoder

# Introduction



+

Reranker

Fusion-in-Decoder

# Our approach

# Model Architecture

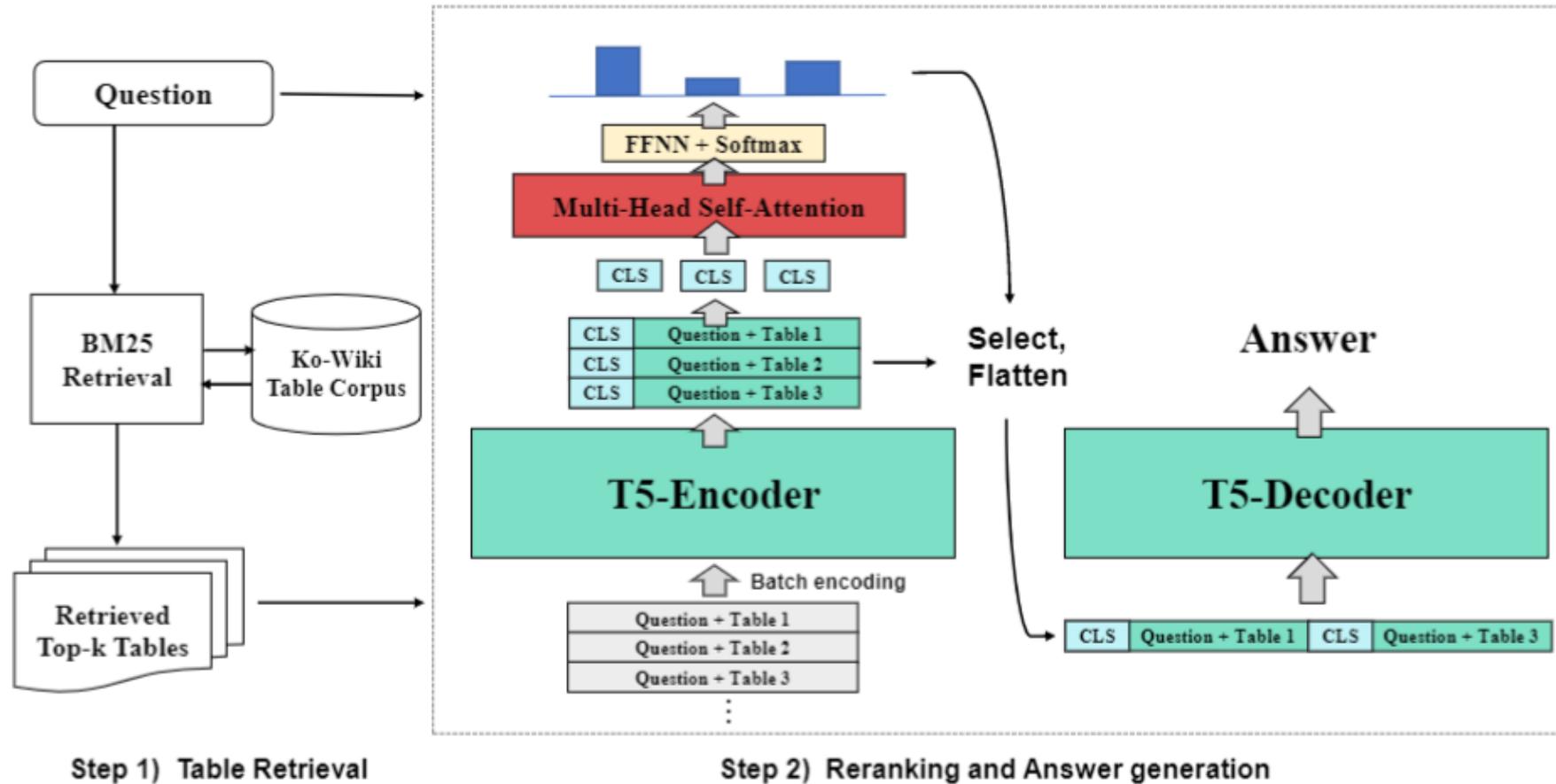


그림 1. Joint Reranker 기반 Fusion-In-Decoder를 이용한 한국어 테이블 오픈 도메인 질의 응답 시스템 구조도

# Experiment - Dataset

Korean-Specific Dataset for Table Question Answering, Jun et al. LREC 2022.

Example		
Field	Content	Note
T	축구 팀 코리아 은퇴식 2002년 11월부터 70경기 이상 대한민국 축구 국가대표팀 경기를 치른 선수들에게 은퇴식을 거행하고 있다. 박지성은 2014년 7월 K리그 올스타전에서 은퇴경기를 했다.황선홍은 폴란드전 인터뷰에서 이번경기를 마지막 경기로 삼겠다라고 밝혔다.이후 홍명보와 은퇴식을 했다.	Title with description
U	<a href="https://ko.wikipedia.org/w/index.php?title=축구_팀_코리아">https://ko.wikipedia.org/w/index.php?title=축구_팀_코리아</a>	Url
QAS	"qid": "22638_1", "question": "2010년 8월 11일 대한민국 축구 국가대표팀 은퇴식을 거행한 선수는 누구 인가요?", "answer": "이운재"	Question-answering set
TBL	[[["거행일","이름","포지션","등번호","경기","활동 기간","장소"], ["2002년 11월 20일","황선홍","공격수","18","103","1988-2002","서울월드컵경기장"], ["2002년 11월 20일","홍명보","수비수","20","136","1990-2002","서울월드컵경기장"], ["2010년 8월 11일","이운재","골키퍼","1","133","1994-2010","수원월드컵경기장"], ...]]	Flattened table of 2D list

표 1. 데이터셋 통계

데이터 셋	갯수
Train set	58221
Dev set	11771
전체 테이블 수	1266298

# Experiment - results

표 2. 검색기 성능 평가

	Top-1	Top-5	Top-20	Top-50	Top-100
Train	3.06	14.16	28.19	37.11	43.57
Dev	3.08	14.09	28.27	37.67	44.33

표 3. 질의 응답 모델 성능 평가

Model	EM	F1-Score
FID(N1-20)	17.49	22.81
R <sup>2</sup> FID(N1-50, N2-20)	20.85	26.06

# Conclusion

- 본 논문에서는 한국어 오픈 도메인 테이블 질의응답 시스템을 구성하고, 단일 테이블 검색기의 한계를 보완하기 위해 Joint Reranker 기반 Fusion-In-Decoder를 제안하고 적용하여 질의응답 Task에서 베이스라인에 대비하여 EM 3.36, F1-Score 3.25 향상을 이루어냈다.

## Future works

- 테이블 구조에 대해 Inductive Bias를 주는 구성요소(Component) 구현
- 대용량 테이블 코퍼스를 통한 테이블 Pre-trained model 개발 연구 등을 진행할 것이다.