

# Dense Retrieval을 이용한 한국어 FAQ Retrieval

강동찬<sup>01</sup>, 나승훈<sup>1</sup>, 김태형<sup>2</sup>, 최윤수<sup>2</sup>, 장두성<sup>2</sup>

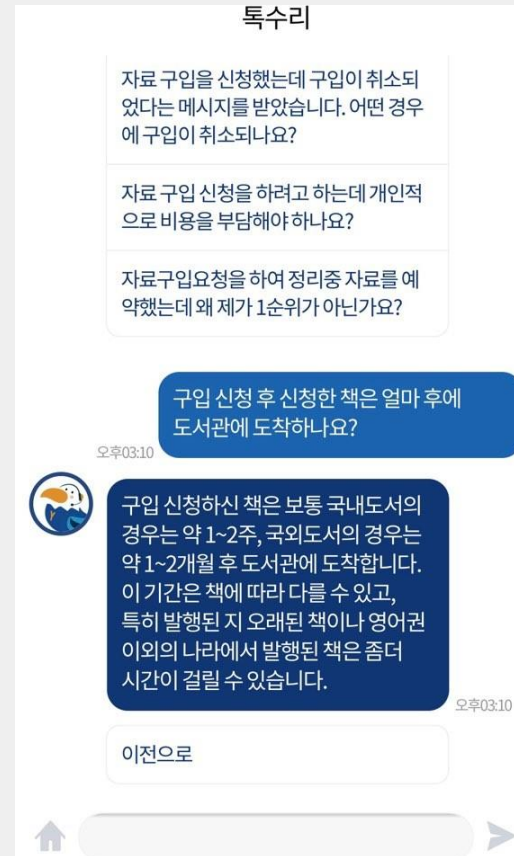
<sup>1</sup>전북대학교, <sup>2</sup>KT융합기술원

- 1. Introduction**
- 2. Related Works**
- 3. Proposed Model**
- 4. Experiment**
- 5. Conclusion**

# 1. Introduction

## ■ FAQ: Frequently Asked Questions

- 특정한 주제의 유저 쿼리에 대해 미리 준비된 질문, 답변 쌍을 이용하여 대답하는 기술
- 기업, 웹사이트, 상담/문의 봇, AI 스피커 등에서 질의응답 세부기술로 활용되고 있다.



# 1. Introduction

## ■ Advantages

- 빈도가 잦은 유저의 쿼리에 대해 빠르게 대응하여 서비스 비용을 절감할 수 있다.

AI 봇을 도입하여 상담원이 단순 문의를 해결하는 경우를 줄이고 전문화 된 질의는 상담원이 진행

- 관리가 용이하며 신뢰도가 높은 응답을 유저에게 제공해줄 수 있다.

서비스 약관처럼 빠르게 반영되어야 하는 정보를 업데이트하기 쉽다.

## ■ Challenge

- 서비스 제공 시에는 유저 쿼리에 나타나는 어휘들과 FAQ 컬렉션에 나타나는 어휘 사이에는 차이가 있다.

때문에 질 높은 자동화 FAQ 서비스 제공을 위해서는 검색 시스템에 자연어 이해, 의미론적인 텍스트 매칭에 대한 능력이 요구된다.

# 1. Introduction

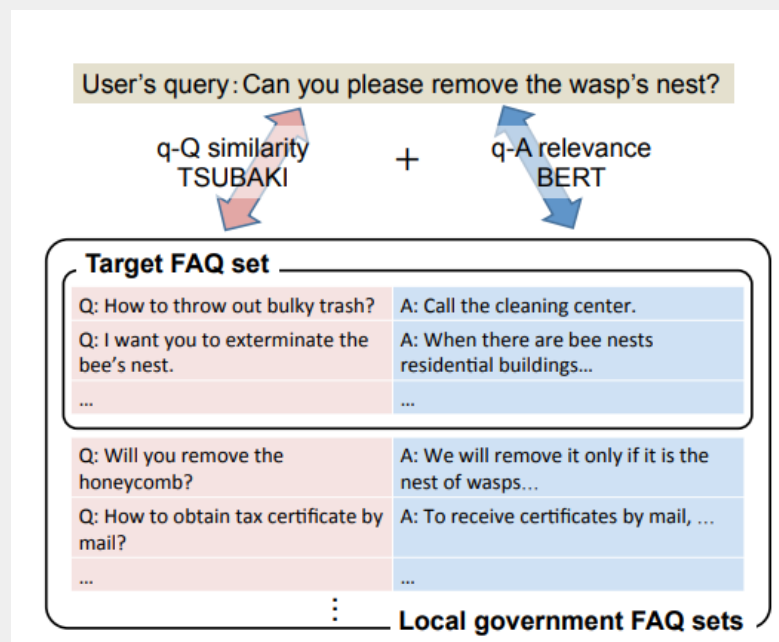
FAQ Retrieval은 이러한 어려움을 해결하고자 하는 자연어 처리 태스크

FAQ Retrieval에서는 유저 쿼리와 관련 있는 (질문,답) 쌍을 FAQ 컬렉션에서 검색하여 상위 검색결과로 가져오는 것을 목표로 한다.

## 2. Related Works

### ■ Related Works

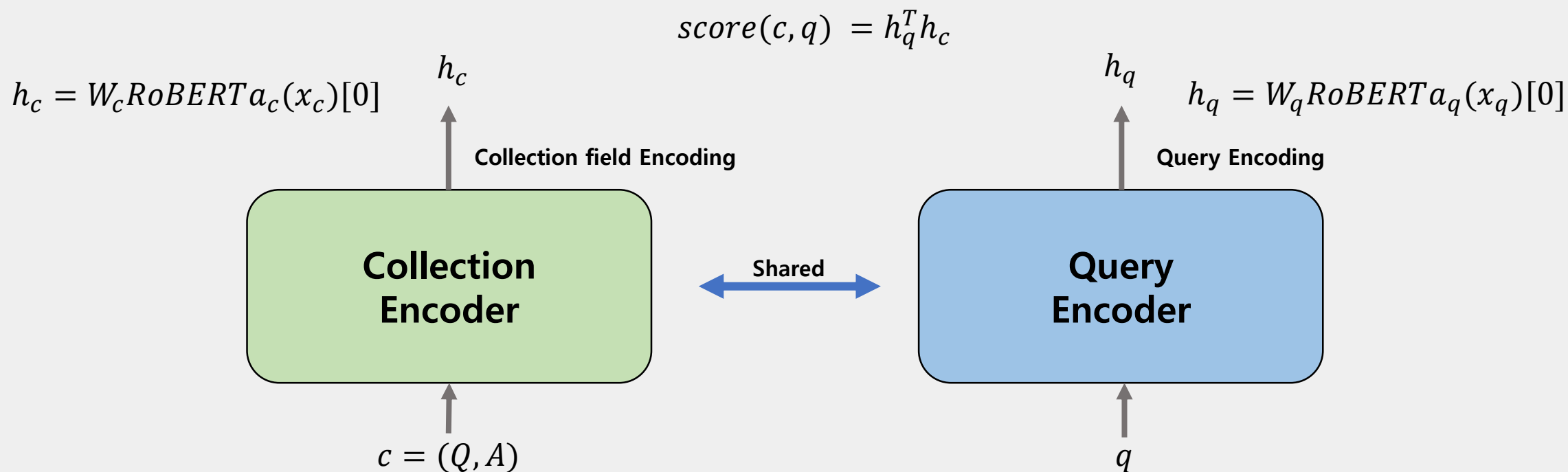
- 미리 수집되어 있는 FAQ 컬렉션 집합(C)에 속한 (질문,답변) 쌍들은 질문(Q), 답변(A) 이라는 두가지 필드를 가진 인스턴스, 유저 쿼리는 (q)로 볼 수 있다.
- 기존의 연구에서는 유저 쿼리 q에 대해 인스턴스의 어떤 필드(Q, A, QA)에 매칭 시킬 것인가에 따라 q-Q, q-A, q-QA의 필드 매칭을 사용했으며, 매칭 네트워크에 대한 연구들이 진행됨.
- 대표적인 연구로 *Sakata*는 q-Q BM25를 통해 소규모 후보군을 가져오고, q-A BERT reranker를 이용해 재정렬했다.



### 3. Proposed Model

#### ■ Proposed Model

- 컬렉션 인코더는 Q, A, QA 필드 중 하나를 인코딩 하여 벡터로 표현하며, 쿼리 인코더는 쿼리를 인코딩하여 벡터로 표현. 이후 두 벡터의 내적이 유사도 점수로 계산되는 모델.
- 인코더는 서로의 파라미터를 공유



# 3. Proposed Model

## ■ Proposed Model

### FAQ Collection

Q: 특정 번호 수신 차단 알려줘

A: 휴대폰 특정번호 수신차단 서비스는 받기 싫은 특정 전화번호의 통화, 문자 차단 ...

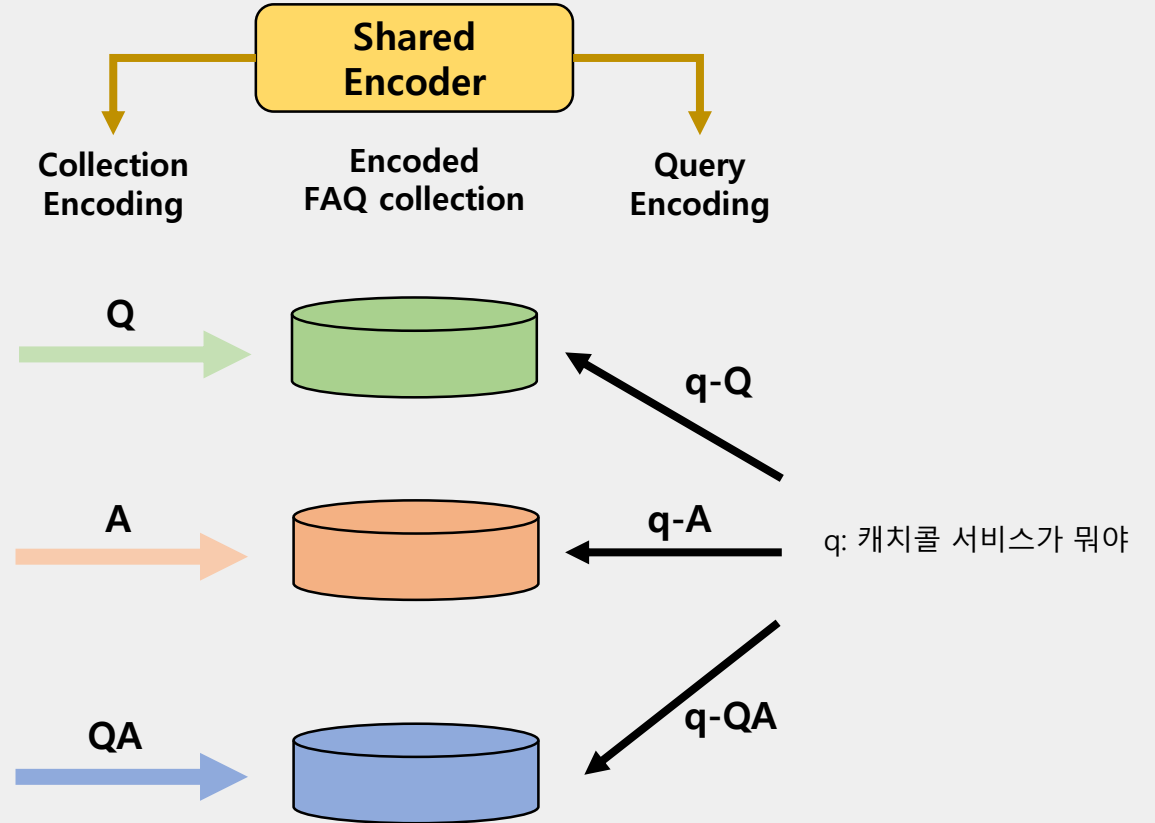
Q: 내가 전화했을 때 친구 전화가 꺼져 있었는데 전화를 켜면 나한테 전화 왔었다고 문자 가는 서비스 있나요?

A: 통화요구알리미 서비스는 상대방과 통화 연결이 안되는 경우에 상대방에게 ...

Q: 통화 중에 어디서 전화 왔는지 문자로 보내줬던거 한번에 보고 싶은데 어디서 볼 수 있어요?

A: 캐치콜 이력은 KT닷컴 홈페이지, 마이 KT 어플에서 SMS 또는 신용카드 인증 후 확인이 가능합니다.

⋮





## 3. Proposed Model

### ■ Proposed Model

- 인코더는 2단계 훈련을 거쳐 훈련 되는 방식
- 1단계: Train 셋의 (Q, A) 쌍들을 이용하여 In-Batch Negative 세팅에서  $L(c^+, q)$ 를 최소화하는 훈련

$$P(c^+|q) = \sum_{c' \in B} \frac{\exp(\text{score}(c^+, q))}{\exp(\text{score}(c', q))}$$

- 2단계: 컬렉션 인코더를 프리징 하고, 훈련 컬렉션과 함께  $L(c^+, q)$ 를 최소화하는 훈련

$$P(c^+|q) = \sum_{c' \in C_{train}} \frac{\exp(\text{score}(c^+, q))}{\exp(\text{score}(c', q))}$$

$$L(c^+, q) = -\log \sum P(c^+|q)$$

## 4. Experiment

### ■ Dataset

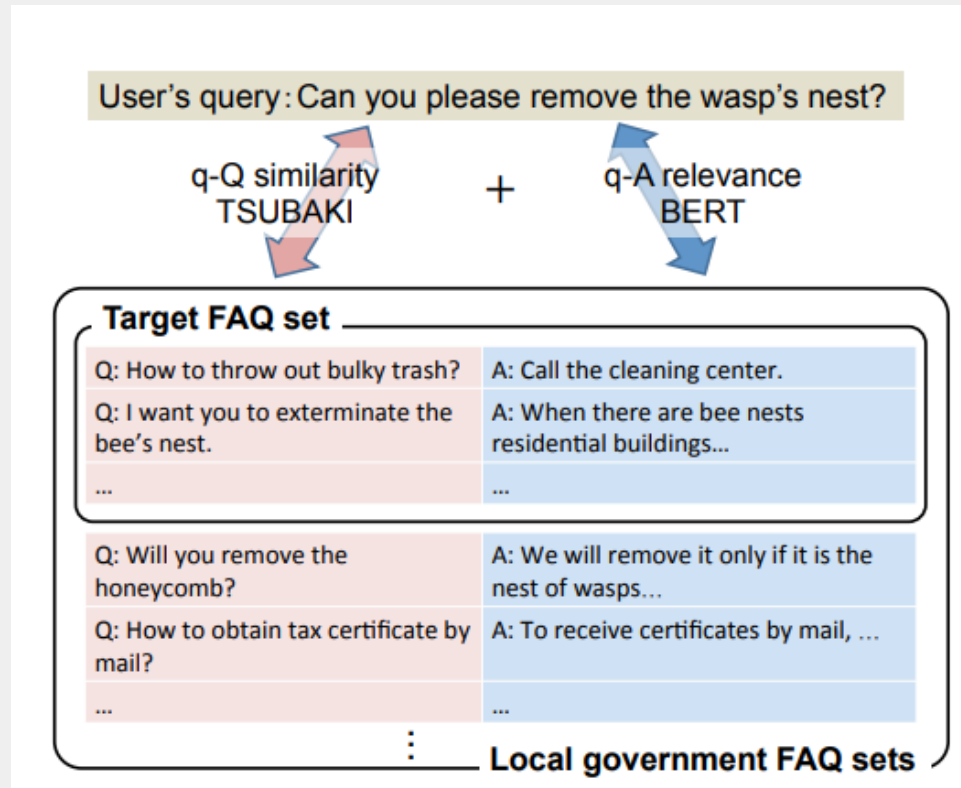
- KT 융합기술원에서 제공한 FAQ 컬렉션과 쿼리 셋
- Train 셋과 Dev 셋은 컬렉션을 나누어서 구축
- Test 셋은 FAQ 컬렉션의 질문에서 패러프레이징이 진행된 질문들로 이루어져 있음.

	Data split	#q, (Q, A)
Collection	Train	15224
	Dev	1000
Query	Test	1000

## 4. Experiment

### ■ Baseline: BM25, Reranker

- BM25와, q-Q BM25를 통해 소규모 후보군을 가져오고, q-A BERT-Reranker를 이용해 재정렬하는 모델.



## 4. Experiment

### ■ Experiment result

Model	Match	Hits@1	Hits@5	MRR@10
BM25	q-Q	0.6329	0.8473	0.7225
	q-A	0.2852	0.3965	0.3294
	q-QA	0.5492	0.7829	0.6483
Reranker	q-Q	0.7608	0.8924	0.8168
	q-A	0.7075	0.8887	0.7786
Dense	q-Q	0.8114	0.9163	0.8545
	q-A	0.7847	0.8740	0.8147
	q-QA	0.8040	0.8859	0.8362
Dense+	q-Q	<b>0.8169</b>	<b>0.9190</b>	<b>0.8570</b>
	q-A	0.7875	0.8767	0.8180
	q-QA	0.8077	0.8850	0.8368

- 모든 모델에서 q-Q 필드 매칭 방식이 높은 검색 성능을 보임.
- Dense retrieval 모델이 가장 좋은 성능을 보이는 가운데, 2단계 학습까지 마친 모델이 가장 높은 성능을 보였다.

## 4. Experiment

### ■ Conclusion

- 한국어 FAQ retrieval을 위한 Dense Retrieval 모델을 제시하고, 기존의 방식인 BM25, BM25 + Reranker 두가지 검색 모델과 함께 훈련, 평가
- 실험결과, Dense retrieval 모델은 테스트 셋에서 MRR@10 0.8570의 성능으로 기존 방법들에 비해 (BM25: 0.7225, Reranker: 0.8168) 더 높은 성능을 보였다.
- 추후에는 FAQ 컬렉션 내에 정답이 없을 때를 판별하여 FAQ Retrieval 알고리즘을 고도화 할 계획이다.