

## 과제 기술서: 한국어 어문 규범 기반 생성(RAG)-㉔ 유형\*

\* ㉔ 유형(데이터 제한): 외부 데이터를 사용하거나 데이터를 증강하여 사용 불가

### 1. 과제 개요

‘한국어 어문 규범 기반 생성(RAG)’ 과제는 한국어 어문 규범 관련 질문에 대하여 국어 지식을 참조하여 답변을 생성하는 과제이다. 이 과제는 국립국어원의 ‘2024년 글쓰기 첨삭 지원을 위한 지시문 기반 생성 말뭉치 구축 연구’ 사업의 결과물인 ‘글쓰기 첨삭 지원을 위한 기초 자료’를 활용하여 설계되었다.

이 과제를 해결하기 위해서는 어문 규범에 알맞은 표현을 고르거나 어문 규범에 따라 문장을 교정하고 그 이유를 설명해야 한다. 이를 위하여 모델 개발 시 검색 증강 생성(RAG) 기술을 활용할 수 있도록 국어 지식 자료를 별도 제공한다. 이 과제의 입력은 문제이고, 출력은 주어진 국어 지식 자료를 활용하여 생성된 정답과 정답에 대한 이유이다.

유형	항목	내용
선택형	문제	"가축을 기를 때에는 {먹이량/먹이양}을 조절해 주어야 한다." 가운데 올바른 것을 선택하고, 그 이유를 설명하세요.
	정답	"가축을 기를 때에는 먹이양을 조절해 주어야 한다."가 옳다. 한 음절의 한자어는 앞말이 고유어나 외래어일 때는 독립적인 한 단어로 인식하여 두음 법칙을 적용하고, 앞말이 한자어일 때는 하나의 단어로 인정하지 않아 두음 법칙을 적용하지 않는다. 따라서 한자어 '량'은 앞말이 고유어나 외래어일 때는 '양'이 되고 한자어일 때는 '량'이 된다. '먹이'는 고유어이므로 '먹이양'이 맞는 말이다.
교정형	문제	다음 문장에서 어문 규범에 부합하지 않는 부분을 찾아 고치고, 그렇게 고친 이유를 설명하세요. "어서 쾌차하시길 바라요."
	정답	"어서 쾌차하시길 바라요."가 옳다. 동사 '바라다'에 어미 '-아요'가 결합한 형태이므로 '바라요'로 표기해야 한다. '바람', '바래요', '바랬다'는 모두 비표준형이고 '바람', '바라요', '바라다'가 표준형이다.

<표 1> 한국어 어문 규범 기반 생성(RAG) 과제의 예시

### 2. 과제 정의

과제는 어문 규범에 알맞은 표현을 고르거나 어문 규범에 따라 문장을 교정하고 그 이유를 설명하는 것이다. 이때 한글 맞춤법, 표준어 사정 원칙, 문장 부호 규정, 외래어 표기법 등의 한국어 어문 규범을 참조한다. 이에 따라 선택형 문제 해결 시에는 선택지 중 어문 규범을 따르는 표현을 고르고 해당 표현을 선택한 이유를 국어 지식 자료를 참조하여 설명해야 한다. 그리고 교정형 문제 해결 시에는 국어 지식 자료를 참조하여 문제의 문장을 어문 규범에 따라 수정하고 수정한 이유를 설명해야 한다. 정답은 {선택:교정 문장}이/가 옳다. {이유}의 형식으로 출력해야 한다.

출력 형식 중 {선택:교정 문장}이/가 옳다.'는 완전 일치(Exact Match)로 평가하며, '{이유}'는 루지(ROUGE)와 버트스코어(BERTScore), 블루알트(BLEURT) 세 지표의 평균으로 모델의 성능을 정량 평가한다.

분류	유형	예시	비고
선택형	입력	{ "input": { "question_type": "선택형", "question": "\"가축을 기를 때에는 {먹이량/먹이양}을 조절해 주어야 한다.\" 가운데 올바른 것을 선택하고, 그 이유를 설명하세요." }, }	문자열
	출력	"output": { "answer": "\"가축을 기를 때에는 먹이양을 조절해 주어야 한다.\"가 옳다. 한 음절의 한자어는 앞말이 고유어나 외래어일 때는 독립적인 한 단어로 인식하여 두음 법칙을 적용하고, 앞말이 한자어일 때는 하나의 단어로 인정하지 않아 두음 법칙을 적용하지 않는다. 따라서 한자어 '량'은 앞말이 고유어나 외래어일 때는 '양'이 되고 한자어일 때는 '량'이 된다. '먹이'는 고유어이므로 '먹이양'이 맞는 말이다." }	문자열
평가	완전 일치(Exact Match), 루지(ROUGE), 버트스코어(BERTScore), 블루알트(BLEURT)의 평균		
교정형	입력	{ "input": { "question_type": "교정형", "question": "다음 문장에서 어문 규범에 부합하지 않는 부분을 찾아 고치고, 그렇게 고친 이유를 설명하세요.\n\n\"어서 쾌차하시길 바라요.\"" }, }	문자열
	출력	"output": { "answer": "\"어서 쾌차하시길 바라요.\"가 옳다. 동사 '바라다'에 어미 '-아요'가 결합한 형태이므로 '바라요'로 표기해야 한다. '바람', '바래요', '바랬다'는 모두 비표준형이고 '바람', '바라요', '바라다'가 표준형이다." }	문자열
평가	완전 일치(Exact Match), 루지(ROUGE), 버트스코어(BERTScore), 블루알트(BLEURT)의 평균		

<표 2> 한국어 어문 규범 기반 생성(RAG) 과제 입력과 출력의 예

### 3. 자료 형식

데이터 세트는 제이슨(JSON) 형식으로 제공되며 각 제이슨(JSON)은 질문, 참조 문서, 정답을 포함한다. <표 3>은 데이터 형식의 예시이다. 주어진 훈련 데이터와 시험 데이터는 동일한 제이슨(JSON) 형식으로 제공되며 시험 데이터의 경우에는 각 텍스트에 대한 출력(output) 항목이 없다. 참가 팀은 모델의 출력 결과를 추가하여 제출한다.

※ 훈련 데이터와 제출 데이터의 형태 및 구성 요소는 동일하다.

항목	내용
훈련 데이터 형식의 예	<pre>[   {     "id": "knowledge-qa-rag-train-0001",     "input": {       "question_type": "선택형",       "question": "\"가축을 기를 때에는 {먹이량/먹이양}을 조절해 주어야 한다.\" 가운데 올바른 것을 선택하고, 그 이유를 설명하세요."     },     "output": {       "answer": "\"가축을 기를 때에는 먹이양을 조절해 주어야 한다.\"가 옳다. 한 음절의 한자어는 앞말이 고유어나 외래어일 때는 독립적인 한 단어로 인식하여 두음 법칙을 적용하고, 앞말이 한자어일 때는 하나의 단어로 인정하지 않아 두음 법칙을 적용하지 않는다. 따라서 한자어 '물'은 앞말이 고유어나 외래어일 때는 '양'이 되고 한자어일 때는 '량'이 된다. '먹이'는 고유어이므로 '먹이양'이 맞는 말이다."     }   },   ... ]</pre> <p>- 아이디(id)와 입력(input), 그리고 출력(output)으로 구성</p>
평가 데이터 형식의 예 (제출 전)	<pre>[   {     "id": "knowledge-qa-rag-train-0001",     "input": {       "question_type": "선택형",       "question": "\"가축을 기를 때에는 {먹이량/먹이양}을 조절해 주어야 한다.\" 가운데 올바른 것을 선택하고, 그 이유를 설명하세요."     },     ...   },   ... ]</pre> <p>- 학습용 데이터와 동일한 형태 - "output" 키와 값을 제거한 데이터</p>
제출 데이터 형식의 예	<pre>[   {     "id": "knowledge-qa-rag-train-0001",     "input": {       "question_type": "선택형",       "question": "\"가축을 기를 때에는 {먹이량/먹이양}을 조절해 주어야 한다.\" 가운데 올바른 것을 선택하고, 그 이유를 설명하세요."     },     "output": {       "answer": "\"가축을 기를 때에는 먹이양을 조절해 주어야 한다.\"가 옳다. 한 음절의 한자어는 앞말이 고유어나 외래어일 때는 독립적인 한 단어로 인식하여 두음 법칙을 적용하고, 앞말이 한자어일 때는 하나의 단어로 인정하지 않아 두음 법칙을 적용하지 않는다. 따라서 한자어 '물'은 앞말이 고유어나 외래어일 때는 '양'이 되고 한자어일 때는 '량'이 된다. '먹이'는 고유어이므로 '먹이양'이 맞는 말이다."     }   },   ... ]</pre>

	<pre>{   ... }</pre>
	- 평가용 데이터에 "output"키와 값을 생성.

<표 3> 데이터 형식의 예

<표 4>는 데이터의 규모이다. 데이터 세트는 훈련(train), 검증(validation), 시험(test) 데이터로 무작위 분할하여 제공한다.

	훈련	검증	시험
대화 수	622	127	498

<표 4> 데이터 규모



#### 4. 기준 모델

기준 모델로는 사전 학습된 거대언어모델(LLM)을 사용하였다. 과제 수행 시 기준 모델에 대한 평가를 재현하기 위한 코드는 깃허브(github)에서 확인할 수 있다.

- [기준 모델] Qwen3-8B<sup>1)</sup>, HyperClova Text 1.5B<sup>2)</sup>
- [코드] [https://github.com/teddysum/Korean\\_QA\\_RAG\\_2025](https://github.com/teddysum/Korean_QA_RAG_2025)

#### 5. 대회 참가 규정

- 인공지능의 한국어 능력 평가에 관심 있는 누구나 팀을 구성하여 참가할 수 있다.
- 팀 구성원은 국립국어원 언어정보나눔터 회원이어야 한다.
- 팀장은 팀원의 전자 우편 주소(회원 가입된 전자 우편 주소)를 입력하여 팀을 구성할 수 있으며, 모든 팀원이 참가 신청에 동의하면 팀 구성이 완료된다.
- 참가자는 접수 기간 중 여러 개의 팀에 참여할 수 있으며, 팀 구성 인원 제한은 없다. 다만, 접수 기간 종료 후 정성 평가 대상이 되는 팀이 여러 개라면 참가자는 과제별 하나의 팀을 소속 팀으로 결정하여야 한다.
- 참가 팀은 과제를 해결한 결과를 정해진 양식에 맞추어 제출한다.(1일 5개까지 제출 가능)
- 라이선스에 문제가 없는 모델을 개발하여야 한다.
- 외부 API를 통해 호출하는 모델(OpenAI API 등)은 제출할 수 없다.
- 개발한 모델은 RTX 4090 24GB 1개에서 구동(추론)할 수 있어야 한다.
- 외부에 공개된 데이터는 사용이 불가하며 과제 데이터(말뭉치)를 거대언어모델(LLM) 등을 활용해 변형 및 증강하여 사용할 수 없다(㉠ 유형). 모델 입력을 위한 형식 및 형태 변형만이 허용된다.
- 순위표 기준 정성 평가 신청서를 제출한 상위 10개 팀을 정성 평가 및 발표 평가 대상으로 선정한다. (해당 팀에게는 별도 안내 예정)
- 최종 수상자는 정량 평가, 정성 평가, 발표 평가를 종합하여 선정한다.

#### 6. 시스템 사용 방법

- 인공지능(AI)말평 → 알림 → 사용 안내 메뉴 내 동영상 참고

#### 7. 문의

- '한국어 어문 규범 기반 생성(RAG)-㉠ 유형' 클릭 → 문의(들여가기 후 작성 가능)

1) <https://huggingface.co/Qwen/Qwen3-8B>

2) <https://huggingface.co/naver-hyperclova/HyperCLOVAX-SEED-Text-Instruct-1.5B>