건설공사 사고 예방 및 대응책 생성

: 한솔데코 시즌3 AI 경진대회

팀명: YG

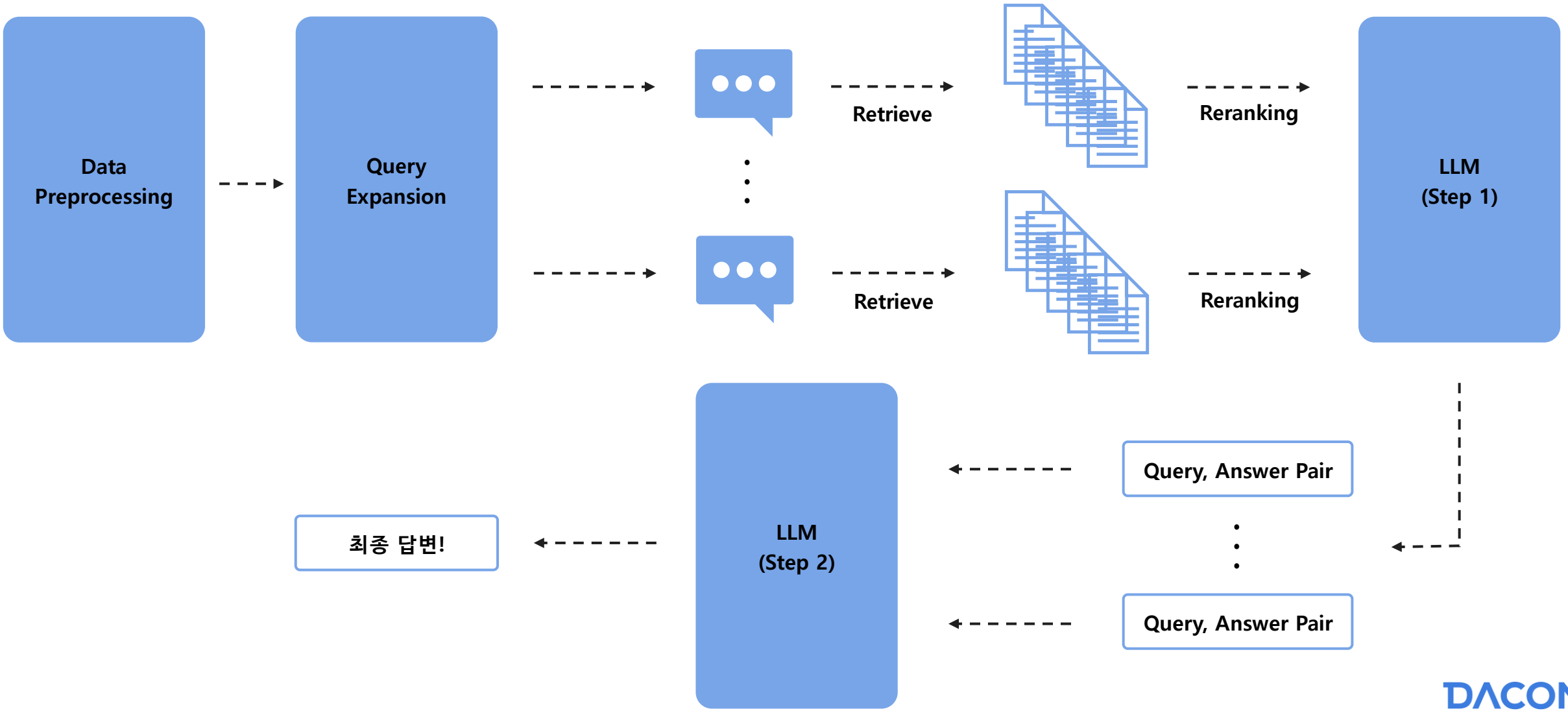
팀원: 최용빈 

홍재민 

DACON

Table of Contents

Overview



DACON

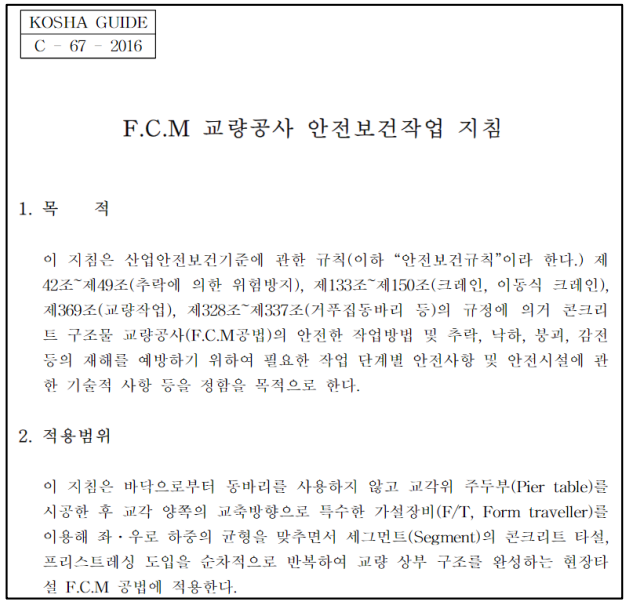
CSV Preprocessing

- 특정 컬럼 값의 경우, Split이 필요하다고 판단
- "/", ">", "(" 의 기호를 기준으로 수행

```
- 공사종류: 건축 / 건축물 / 근린생활시설      → 공사종류1, 공사종류2, 공사종류3
- 인적사고: 떨어짐(5미터 이상 ~ 10미터 미만)  → 인적사고1
- 공종    : 건축 > 철근콘크리트공사           → 공종1, 공종2
- 사고객체: 건설자재 > 철근                   → 사고객체1, 사고객체2
- 장소    : 근린생활시설 / 내부               → 장소1, 장소2
- 부위    : 철근 / 고소                       → 부위1, 부위2
```

PDF Preprocessing 양식을 고려한 전처리

- 표지, 목차 등 필요 없는 페이지 제거
- ***olmOCR*** 을 활용하여 텍스트 추출
- 숫자 패턴 (e.g., “1.”, “1.1”, “1.1.1”), 괄호 숫자 패턴 (e.g., “(1), “(2)”), 괄호 한글 패턴 (e.g., “(가)”, “(나)”)을 기준으로 Chunking 수행



KOSHA GUIDE
C - 67 - 2016

F.C.M 교량공사 안전보건작업 지침

1. 목 적

이 지침은 산업안전보건기준에 관한 규칙(이하 “안전보건규칙”이라 한다.) 제42조~제49조(추락에 의한 위험방지), 제133조~제150조(크레인, 이동식 크레인), 제369조(교량작업), 제328조~제337조(거푸집동바리 등)의 규정에 의거 콘크리트 구조물 교량공사(F.C.M공법)의 안전한 작업방법 및 추락, 낙하, 붕괴, 감전 등의 재해를 예방하기 위하여 필요한 작업 단계별 안전사항 및 안전시설에 관한 기술적 사항 등을 정함을 목적으로 한다.

2. 적용범위

이 지침은 바닥으로부터 동바리를 사용하지 않고 교각위 주두부(Pier table)를 시공한 후 교각 양쪽의 교축방향으로 특수한 가설장비(F/T, Form traveller)를 이용해 좌·우로 하중의 균형을 맞추면서 세그먼트(Segment)의 콘크리트 타설, 프리스트레싱 도입을 순차적으로 반복하여 교량 상부 구조를 완성하는 현장타설 F.C.M 공법에 적용한다.

```
1. 목 적\n\n이 지침은 산업안전보건기준에 관한 규칙(이하 “안전보건규칙”이라 한다.) 제42조~제49조(추락에 의한 위험방지), 제133조~제150조(크레인, 이동식 크레인), 제369조(교량작업), 제328조~제337조(거푸집등바리 등)의 규정에 의거 콘크리트 구조물 교량공사(F.C.M공법)의 안전한 작업방법 및 추락, 낙하, 붕괴, 감전 등의 재해를 예방하기 위하여 필요한 작업 단계별 안전사항 및 안전시설에 관한 기술적 사항 등을 정함을 목적으로 한다.

---

2. 적용범위\n\n이 지침은 바닥으로부터 동바리를 사용하지 않고 교각위 주두부(Pier table)를 시공한 후 교각 양쪽의 교축방향으로 특수한 가설장비(F/T, Form traveller)를 이용해 좌·우로 하중의 균형을 맞추면서 세그먼트(Segment)의 콘크리트 타설, 프리스트레싱 도입을 순차적으로 반복하여 교량 상부 구조를 완성하는 현장타설 F.C.M 공법에 적용한다.

...
```

DACON

Query Expansion 배경

1. 초기 접근 방식의 한계
    - 검색 및 답변 생성에 효과적으로 활용할 만한 Query가 없었음
2. 초기 방법론
    - train/test.csv 내 특정 컬럼 값들의 조합을 Query로 사용
    - train.csv의 "재발방지대책 및 향후조치계획" 컬럼 값을 문서(Document)로 사용
    - Retrieve를 통해 Top k의 관련 문서를 Context로 추출하고, 이를 종합하여 최종 답변을 생성
3. 답변의 실용성 문제
    - 위 방식으로 생성된 답변은 실제 사고 상황 분석 및 대응에 직접적인 도움을 주기 어렵다고 판단
    - PDF를 문서로 활용하기로 결정 → *"PDF의 내용을 최대한 답변에 담아보자!"*
4. Query Expansion의 필요성 (대안)
    - 기존 Query와 PDF 내용의 관련성이 낮음
    - 따라서, PDF 내용과 관련성이 높도록 새로운 Query를 생성하는 것이 필요

Query Expansion

- 사용 모델: HumanF-MarkrAI/Gukbap-Gemma2-9B
- 활용한 컬럼: “공종2”, “작업프로세스”, “사고객체1”, “사고객체2”, “인적사고1”, “사고원인”
- 3개의 Query로 확장

```
You are a construction safety expert. \
Your task is to generate focused questions that will help identify preventive measures from safety guideline documents based on given accident scenarios. \
Each question should be concise, clear, and focused on a single topic or perspective. \
These questions will be used to search through safety documentation to find relevant preventive measures and protocols. \
Avoid compound questions or questions that address multiple scenarios simultaneously. \
Please generate 3 questions in Korean.

While performing '{job_process}' at '{gongjong}', an incident involving '{human_accident}' occurred at '{accident_object}'. \
The cause of the accident is as follows: \
'{accident_cause}'

We are trying to find preventive measures in the safety guideline documents. Please generate 3 questions in Korean.

<Output Format>
{{
    "questions": [
        "Question 1",
        "Question 2",
        "Question 3"
    ]
}}
```

<Prompt>

DACON

Query Expansion

- 사용 모델: HumanF-MarkrAI/Gukbap-Gemma2-9B
- 활용한 컬럼: “공종2”, “작업프로세스”, “사고객체1”, “사고객체2”, “인적사고1”, “사고원인”
- 3개의 Query로 확장

```
You are a construction safety expert. \
Your task is to generate focused questions that will help identify preventive measures from safety guideline documents based on given accident scenarios. \
Each question should be concise, clear, and focused on a single topic or perspective. \
These questions will be used to search through safety documentation to find relevant preventive measures and protocols. \
Avoid compound questions or questions that address multiple scenarios simultaneously. \
Please generate 3 questions in Korean.

While performing '타설작업' at '철근콘크리트공사', an incident involving '부딪힘' occurred at '건설기계, 콘크리트펌프'. \
The cause of the accident is as follows: \
'펌프카 아웃트리거 바닥 고임목을 3단으로 보강 했음에도, 지반 침하(아웃트리거 우측 상부 1개소)가 발생하였고, 좌, 우측 아웃트리거의 펼친 길이가 상이하고 타설
위치가 건물 끝부분 모서리에 위치하여 붐대호스를 최대로 펼치다보니 장비에 대한 무게중심이 한쪽으로 쏠려 일부 전도되는 사고가 발생된 것으로 판단됨'

We are trying to find preventive measures in the safety guideline documents. Please generate 3 questions in Korean.

<Output Format>
{{
    "questions": [
        "Question 1",
        "Question 2",
        "Question 3"
    ]
}}
```

<Prompt 적용 예시>

DACON

Inference Step 1

- 사용 모델: HumanF-MarkrAI/Gukbap-Gemma2-9B
- 각 3개의 Query에 대해 Retrieve 후, 검색된 Top k의 문서에 대해 Reranking 수행
- Query별로 답변을 생성하여, 총 3개의 쌍 (질문, 답변) 이 도출

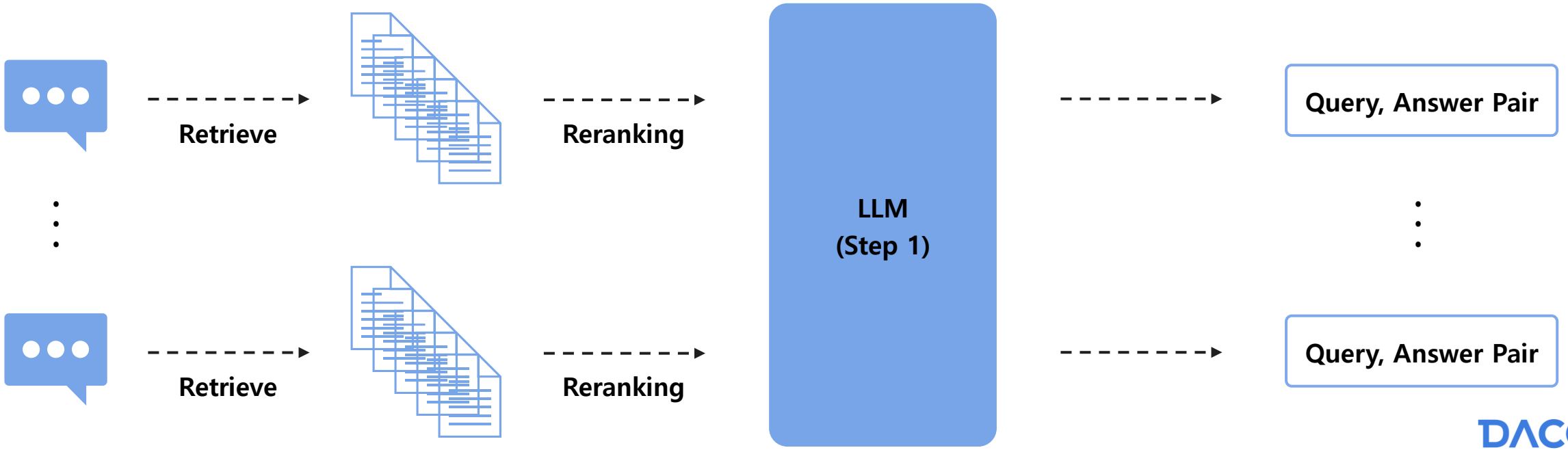
```
You are a construction safety expert. \
Your task is  to answer preventive measures for the cause of an incident given as a query based on context.

The following responses were generated for different aspects of the same safety concern. Each response addresses a specific question related to the main safety issue.

Follow these guidelines when responding:
- Identify the core safety concern across all questions and answers
- Extract and combine the most critical safety measures from all responses
- Create a unified, comprehensive safety guideline that addresses the central issue
- Ensure the final answer is clear, actionable, and directly applicable to construction sites
- Keep the response within ONLY ONE sentence in KOREAN.

Query: {query}

Context:
{context}

Answer:
```

<Prompt>

DACON

Inference Step 1

- 사용 모델: HumanF-MarkrAI/Gukbap-Gemma2-9B
- 각 3개의 Query에 대해 Retrieve 후, 검색된 Top k의 문서에 대해 Reranking 수행
- Query별로 답변을 생성하여, 총 3개의 쌍 (질문, 답변) 이 도출



Inference Step 2

- 사용 모델: HumanF-MarkrAI/Gukbap-Gemma2-9B
- 3개의 쌍을 Context로 주어서, 종합 답변을 최종적으로 생성

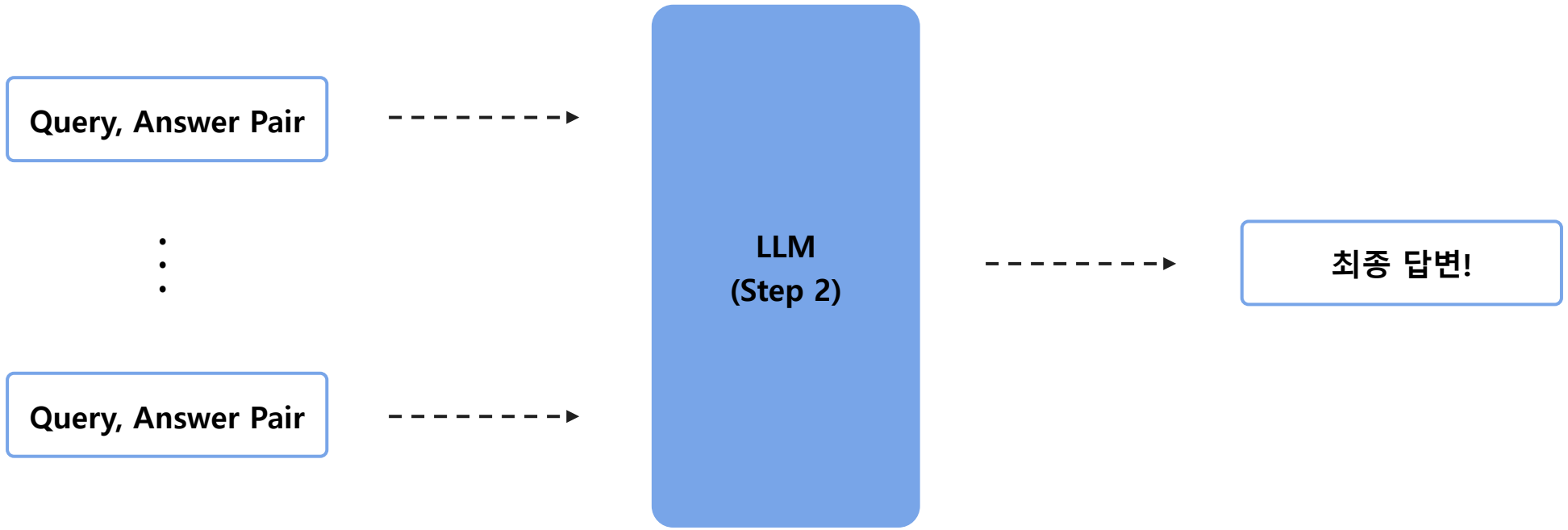
```
You are a construction safety expert. \
Based on the provided question and retrieved safety guidelines, generate a single, comprehensive preventive measure that incorporates key points from the guidelines. \
Follow these guidelines when responding:
- Do not provide multiple alternative measures—combine key ideas into one well-structured sentence.
- Ensure the measure is clear, actionable, and directly applicable to construction sites.
- Keep the response within ONLY ONE sentence in KOREAN.

Question: {question}
Context:
{context}

Answer:
```

<Prompt>

DACON

Inference Step 2

- 사용 모델: HumanF-MarkrAI/Gukbap-Gemma2-9B
- 3개의 쌍을 Context로 주어서, 종합 답변을 최종적으로 생성



감사합니다 🤗

팀명: YG

팀원: 최용빈 

홍재민 

DACON

Appendix

- GitHub Link : https://github.com/whybe-choi/dacon-hansoldeco-season3
- Requirements

```
torch==2.4.0
peft==0.14.0
trl==0.15.2
transformers==4.50.0
datasets==3.3.2
accelerate==1.4.0
evaluate==0.4.3
bitsandbytes==0.45.3
wandb==0.17.4
deepspeed==0.15.4
sentence-transformers==3.4.1
langchain==0.3.19
langchain-community==0.3.18
langchain-huggingface==0.1.2
langchain-openai==0.3.7
langchain-text-splitters==0.3.6
langchain-qdrant==0.2.0
pymupdf==1.25.3
pymupdf4llm==0.0.17
qdrant-client==1.13.3
```