

도배를

하자하자!팀

김유민 이아영 박은혜

	id	질문_1	질문_2	category	답변_1	답변_2	답변_3	답변_4	답변_5
0	TRAIN_000	면진장치가 뭐야?	면진장치에 사용되는 주요 기술은 무엇인가요?	건축구조	면진장치란 지반에서 오는 진동 에너지를 흡수하여 건물에 주는 진동을 줄여주는 진동 ...	면진장치란 건물의 지반에서 발생하는 진동 에너지를 흡수 하여 건물을 보호하고, 진동을...	면진장치란 지반으로부터 발생 하는 진동 에너지를 흡수하여 건물에 전달되는 진동을 줄여...	면진장치는 건물의 지반으로부터 오는 진동 에너지를 흡수하여 건물에 전달되는 진동을 ...	면진장치는 건물에 오는 지반 진동의 영향을 최대한으로 흡수하여 건물에 전달되는 진동...
1	TRAIN_001	내진설계의 종류 중 알려줘	내진설계에는 어떤 종류가 있는지 자세히 알려주실 수 있나요?	건축구조	내진 설계의 종류로 내진구조, 제진구조, 면진구조가 있습니다.	내진설계에는 내진구조, 제진구조, 면진구조가 있습니다. 내진 구조는 건물 구조물이 지...	내진설계에는 주로 내진구조, 제진구조, 면진구조의 세 가지 종류가 있습니다. 이들은...	내진설계에는 주로 내진구조, 제진구조, 면진구조가 사용됩니다. 내진구조는 건물 구조...	내진 설계에는 다양한 종류가 있지만, 대표적으로 내진구조, 제진구조, 면진구조가 있...
2	TRAIN_002	철골구조의 장점이 뭐야?	철골구조의 장점을 알려줘?	건축구조	철골구조는 건물의 외벽에는 그다지 하중이 걸리지 않기 때문에 고층 건물의 건축이 가...	철골구조의 장점은 건물의 외벽에는 그다지 하중이 걸리지 않기 때문에 고층 건물의 건...	철골구조의 장점은 건물의 외벽에 하중이 적게 걸리기 때문에 고층 건물의 건축이 용이...	철골구조의 장점은 건물의 외벽이 하중이 걸리지 않아 공간 활용이 용이하고, 고층 건...	철골구조의 장점은 건물의 외벽에 하중이 크게 걸리지 않아 고층 건물을 건축할 수 있...
3	TRAIN_003	철골철근 콘크리트 구조가 뭐야?	철골철근 콘크리트 구조의 장점과 단점에는 무엇이 있을까요?	건축구조	철근철골콘크리트는 철골과 철근, 그리고 콘크리트를 함께 사용하는 건축 구조입니다. ...	철골철근콘크리트 구조는 건축물을 지탱하는 주요 구조물인 철골과 철근, 그리고 콘크리...	철골철근 콘크리트 구조는 건축물을 지탱하기 위한 구조물에서 일반적으로 사용되는 방식...	철골철근콘크리트 구조는 철골과 철근, 그리고 콘크리트를 함께 사용하여 만들어지는 건...	철골철근 콘크리트 구조는 강철 골조와 강철 철근, 그리고 콘크리트를 함께 사용하여 ...
4	TRAIN_004	철골구조는 어떤 방식이 있어?	철골구조의 다양한 방식이 무엇인가요?	건축구조	철골구조는 일반철골구조와 경량철골구조가 있습니다.	철골구조는 일반철골구조와 경량철골구조가 있습니다. 일반철골구조는 주로 대형 건물이나...	철골구조는 주로 일반철골구조와 경량철골구조로 나뉘어집니다. 이들은 건축 시스템에 따...	철골구조는 주로 일반철골구조와 경량철골구조로 구분됩니다. 이외에도 최근에는 고층 건...	철골구조는 일반철골구조와 경량철골구조 두 가지 방식이 주로 사용됩니다. 일반철골구조...

Train 데이터 구조 : 질문 2개, 답변 5개, category 표시

```
train['category'].value_counts()
```

```
마감재      272
인테리어   123
시공       111
마감하자    60
건축구조    31
기타        27
타 마감하자  20
Name: category, dtype: int64
```

1. 모델 관련 논의

- 이전 회의에서 Bert 모델을 사용하기로 결정

#Model

```
model = BertForQuestionAnswering.from_pretrained('bert-large-uncased-whole-word-masking-finetuned-squad')
```

#Tokenizer

```
tokenizer = BertTokenizer.from_pretrained('bert-large-uncased-whole-word-masking-finetuned-squad')
```

- Bert 모델의 데이터 입력 구조인 (질문, 대답) 형태로 데이터 변환 후 학습 시도

```
-----  
ValueError                                Traceback (most recent call last)  
<ipython-input-32-9080f658325a> in <cell line: 9>()  
      7  
      8 # BERT 입력을 PyTorch Tensor로 변환  
----> 9 input_ids = torch.tensor(train['bert_inputs'].sum(), dtype=torch.long)  
     10 attention_mask = torch.ones_like(input_ids) # 각 토큰에 대한 attention mask  
     11  
ValueError: expected sequence of length 155 at dim 1 (got 222)
```

Kogpt 모델

gpt-3 기반 한국어 언어 생성모델

- 한국어를 사전적, 문맥적으로 이해하고 이용자가 원하는 결과값을 보여줌
- 60억개의 매개변수와 2000억개 토큰(token)의 한국어 데이터를 바탕으로 구축

△ 주어진 문장의 긍정과 부정 판단

△ 긴 문장 한줄 요약

△ 문장을 추론해 결론 예측

△ 질문을 하면 문맥 이해해 답변하는 등 언어를 가지고 활용

⇒ **kogpt 모델**을 사용하기로 결정!

2. 데이터 증강 방법 관련 논의

- (1) SR (Synonym Replacement) : 특정 단어를 유의어로 교체
- (2) RI (Random Insertion) : 임의의 단어를 삽입
- (3) RS (Random Swap) : 문장 내 임의의 두 단어의 위치를 바꿈
- (4) RD (Random Deletion) : 임의의 단어를 삭제

- **대체** ⇒ 한국 국어원 유의어 사전을 이용
- 위키피디아에서 카테고리에 해당하는 사전 정보 가져오기

3. 앞으로의 계획

- kogpt 모델과 관련한 파인튜닝 공부 후 적용하기
- 유의어를 대체하는 방법으로 데이터 증강하기
- 검색 증강 생성 (RAG) 에 대해 공부하기