



Prompt Engineering

김준재

ML Engineer

Use Case #2. 정보 추출

주어진 문맥에서 필요한 정보만을 추출하는 프롬프트 엔지니어링에 대해 알아보니다.

개념 소개

정보 추출이란?

- 비정형 텍스트 데이터 속에서 특정 유형의 정보(예: 이름, 날짜, 장소, 제품명 등)를 식별하고 뽑아내는 기술.
- LLM은 문맥 이해 능력이 뛰어나 이 작업을 매우 잘 수행할 수 있지만, 원하는 정보를 정확히 얻으려면 프롬프트를 잘 작성하는 것이 중요.

활용처

- 수많은 문서에서 필요한 정보만 빠르게 요약/정리. (예: 뉴스 기사에서 핵심 인물/사건 추출)
- 데이터 분석을 위한 정형 데이터 구축. (예: 고객 리뷰에서 제품명, 만족도 추출)
- 자동화된 워크플로우 구축에 기여. (예: 이메일에서 미팅 날짜/시간 추출)

정보 추출 프롬프트의 핵심 요소

명확한 작업 지시(Task)

- 무엇을 추출하고 싶은지 명확하게 정의.

문맥 제공(Context)

- 정보를 추출할 원본 텍스트를 명확히 구분하여 제공. (예: ``` , --- 등의 구분자 사용)

원하는 출력 형식 지정(Format)

- 추출된 정보가 어떤 형태로 나오길 원하는지 지정. (예: 목록, JSON, 표 형식)

예시 패턴 1: 간단한 정보 추출

작업 지시

- 다음 텍스트에서 출력 형식에 해당하는 정보를 추출하세요:

문맥

- ```The Solar Pro model supports up to 32k tokens.
We tested it with the Galactica paper.```

출력 형식

- 출력 형식:
 - 모델 리스트: [모델 이름]
 - 논문 리스트: [논문 이름]

예시 패턴 2: 보다 구조화된 정보 추출 (JSON)

작업 지시

- 다음 텍스트에서 출력 형식에 해당하는 정보를 추출하세요, 오직 결과값만 반환하세요:

문맥

- ```The quantized Solar Pro requires two NVIDIA A10G GPUs, so an AWS ml.g5.12xlarge instance, which includes four A10G GPUs, can support two Solar Pro deployments simultaneously.```

출력 형식

- 추출할 정보: 모델 이름, 파라미터 수, 필요한 GPU 수, 인스턴스명. 누락된 정보는 null로 표시하세요.

JSON 형식:

```
{
  "model_name": "",
  "parameters": "",
  "required_gpu_count": 0,
  "instance_name": ""
}
```

실습

목표

- 주어진 문맥에서 필요한 정보만을 추출하는 프롬프트 엔지니어링 숙달.

예상 결과물

- 단일 정보 추출.
- 정보 목록 추출.
- JSON으로 규격화된 정보 추출.

필요 환경

- 구글 코랩
 - openai 라이브러리
-



Building intelligence for the future of work