

자연어 기반 분류기 구현

1. 기능 개요

사용자의 자연어 질문을 soccer 또는 other로 분류하여, 질문이 축구 관련 질문인지 일반 사내 질문인지 구분하는 경량 라우팅 분류기입니다.

2. 구현 방법

기술/라이브러리	설명
SentenceTransformer	사전 학습된 한국어 SBERT 모델 (jhgan/ko-sroberta-multitask)을 활용해 문장 임베딩 생성
FAISS	고속 최근접 이웃(Nearest Neighbor) 검색 라이브러리. 임베딩 기반 벡터 유사도 비교
Ko-SRoBERTa	한국어 자연어처리를 위한 사전 학습 모델
NumPy	벡터 연산 처리
FAISS IndexFlatL2	L2 거리 기반 최근접 벡터 탐색 (비압축 인덱스)

3. 데이터 구성



총 100+ 문장을 기반으로 두 가지 카테고리로 구분:

- "soccer": 축구 선수 정보, 비교, 추천 질문
- "other": 일반 사내 질문 (복지, 연차, 출근 등)

예시)

```
("손흥민은 어떤 선수야?", "soccer")
("복지포인트는 어디서 확인해?", "other")
```

4. 결과

```
print(classify("퇴사하려면 뭐부터 해야 해?")) #  True (other)
print(classify("득점력 좋은 미드필더는 누구야?")) #  False (soccer)
```