SK네트웍스 **Family AI** 과정 **14**기

데이터 수집 및 저장 데이터 조회 프로그램

산출물 단계	데이터 수집 및 저장
평가 산출물	데이터 조회 프로그램
제출 일자	2025. 08. 22
깃허브 경로	https://github.com/skn-ai14-250409/SKN14-Final-2Team
작성 팀원	유용환

원본 데이터 (텍스트)	**[Web scraping 기반 향수 정보 수집]** 1. 자료 출처 : 해외직구 플랫폼 “바이슈코” 2. 수집 방식 : 웹 스크래핑(Web Scraping) 3. 수집항목 : - 브랜드, 제품명, 용량, 가격, 제품 상세 url, description (부향률, 메인어코드, 탑/미들/베이스 노트, 향 설명) 4. Output 파일 : perfume.csv
데이터 전처리 과정	**[수집데이터 구조 및 전처리 대상]** 1. 전처리 대상 : perfume.csv 2. 처리 방식 : description 텍스트에서 키워드 기반으로 향수 특성 정보를 추출하고, 별도 컬럼으로 분리 3. 추가 생성된 컬럼 : - 부향률, 메인 어코드, 탑 노트, 미들 노트, 베이스 노트, 향 설명 4. 전처리 된 최종 파일 : after_prepro_del_txt.csv
DB 사용 용도	**[Vector DB 구축 및 활용 목적]** 1. 사용 목적 - 사용자의 쿼리(Query)에 적합한 향수를 의미기반으로 검색하여 추천 - 자연어 검색 질의(Query)에 대응하기 위해 텍스트를 임베딩하고 벡터 DB에 저장 - 향수 상세정보에 대한 설명 또는 비교 분석 질문에 대응 가능 2. 활용 구조 - 모든 컬럼을 “page_content + metadata” 구조로 변환 - LangChain + Pinecone 사용하여 데이터 조회 프로그램 구현

VectorDB 구조

<Pinecone Vector DB>

```python
field_queries = {
    "메인어코드": ("우디", 1.0),
    "향설명": ("남성", 10.0),
    "브랜드명": ("딥디크", 2.0),
}

df = search_with_field_weights(field_queries, top_k=20)
display(df)
```

Python

```
메인어코드: 우디
향설명: 남성
브랜드명: 딥디크
```

	brand	name	향 설명	메인 어코드	탑 노트	미들 노트	베이스 노트
0	메종 마르지엘라	재즈 클럽 오 드 뚜왈렛	프라이빗 재즈 클럽에서 느껴지는 남성적이고 할기찬 향	타바코, 스위트, 럼	핑크페퍼, 네롤리, 레몬	럼, 자바 제티버 오일, 세이지	타바코 리프, 바닐라 빈, 때죽나무
1	입생로랑	쿠로스 오 드 뚜왈렛	승리하는 남성성의 시대를 초월한 향	머스크, 아로마틱, 파우더리	알데하드, 고수, 클라리 세이지, 아르테미시아, 베르가못	패출리, 카네이션, 제라늄, 베티버, 시나몬, 자스민, 오리스 뿌리, 라벤더	시빗, 꿀, 레더, 머스키, 오크모스, 앰버, 통카 빈, 바닐라
2	샤넬	블루 드 샤넬 퍼퓸	성취감과 자신감을 보여주는 강렬한 남성의 향기	우디, 시트러스, 아로마틱	레몬 제스트, 베르가못, 민트, 아르테미지아	라벤더, 파인애플, 제라늄, 그린 노트	샌달우드, 시더, 앰버우드, 통카 빈, 이소 이 수퍼
3	입생로랑	르 옴므 오 드 뚜왈렛	신선한 우디 향의 매혹적이고 관능적인 우아함	월 스파이시, 시트러스, 아로마틱	진저, 베르가못, 레몬	향료, 바이올렛 잎, 화이트 페퍼, 바질	통카 빈, 시더, 타히티 베티버
4	입생로랑	르 옴므 오 드 뚜왈렛	신선한 우디 향의 매혹적이고 관능적인 우아함	월 스파이시, 시트러스, 아로마틱	진저, 베르가못, 레몬	향료, 바이올렛 잎, 화이트 페퍼, 바질	통카 빈, 시더, 타히티 베티버
5	샤넬	블루 드 샤넬 오 드 퍼퓸	현실과 규칙에서 벗어나 진취적으로 삶을 살아가는 남성을 위한 향	시트러스, 앰버, 우디	그레이프 프룻, 레몬, 민트, 핑크 페퍼, 베르가못, 알데하이드, 코리안더	진저, 너트맥, 자스민, 멜론	인센스, 앰버, 시더, 샌달우드, 패출리, 앰버우드, 랍다넘
6	샤넬	알뤼르 옴므 오 드 뚜왈렛	과감한 결단력과 카리스마를 지닌 남성을 표현한 향	시트러스, 바닐라, 월 스파이시	레몬, 피치, 진저, 만다린 오렌지, 라벤더, 베르가못	페퍼, 시더, 패출리, 베티버, 브라질리안 로즈우드, 로즈, 자스민, 가데니아, 아...	바닐라, 통카 빈, 샌달우드, 코코넛, 앰버, 벤조인, 머스크, 레더, 오크모스

<Data_Search_Program>