

데이터 수집 및 저장 데이터 수집 보고서

□ 개요

- 산출물 단계 : 데이터 수집 및 저장
- 평가 산출물 : 데이터 수집 보고서
- 제출 일자 : 2025.02.14
- 깃허브 경로 : [깃허브 주소 기재](#)
- 작성 팀원 : 조해원

데이터 수집 목적	<ul style="list-style-type: none">- 사용자가 입력한 식재료를 기반으로 적절한 요리 레시피를 추천- 다양한 출처에서 요리 레시피 데이터를 수집하여 정보의 다양성을 확보하고 사용자 맞춤형 서비스 제공- 레시피와 연계된 영상을 함께 제공하여 직관적인 조리 방법 안내 가능
데이터 수집 방법	<ul style="list-style-type: none">- 방법<ul style="list-style-type: none">- 웹 크롤링(Web Crawling) 기법을 활용하여 데이터 수집 (BeautifulSoup, Selenium 사용)- 소요 시간<ul style="list-style-type: none">- 코드 수정 및 최적화 과정 약 2주- 최종 데이터 크롤링 작업 1일 소요

수집 데이터(요약)

- 출처

- 냉장고를 부탁해 [15분 레시피 | 냉장고를 부탁해 | JTBC](#)

[레시피 영상 | 냉장고를 부탁해 | JTBC](#)

(레시피 제목과 영상 제목을 연결하여 데이터 구성)

- 만개의 레시피 [요리를 즐겁게~ 만개의레시피](#)

- 편스토랑 [신상출시 편스토랑 > 편스토랑 레시피 | 디지털 KBS](#)

- Case 수

- 만개: 226,995개

- 냉부: 970개 (레시피) / 869 (영상)

- 편스: 1,816개

- Column 정보

- 만개: name(음식명), intro(소개), info(정보), photo(사진), ingredient(재료), recipe(조리방법), views(조회수), video(영상), date(등록일자), category(카테고리)

- 냉부: name(음식명), image(사진), ingredient(재료), video(영상), recipe(조리방법)

- 편스: name(음식명), photo(사진), ingredient(재료)

- 객체 탐지(object detection)

데이터셋: roboflow [project2 Dataset > Overview](#), [Fridgify Dataset > Overview](#)