



데이터 분석



학교 급식실에 잔반 측정기가 설치되었다. 식판에 음식을 담아 잔반 측정기에 한 번 스캔하고, 식사를 마친 후 다시 식판을 스캔하여 잔반량을 측정한다. 잔반량 측정을 통해 얻을 수 있는 데이터에는 무엇이 있을지 생각해 보자.



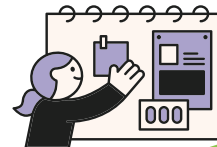
생각 나누기

급식 잔반 측정기를 통해 어떤 데이터가 수집되고, 이것을 어떻게 활용할 수 있을지 이야기해 보자.

01 데이터의 수집과 관리

학습 목표

- 문제 해결 목적에 적합한 데이터를 다양한 방법으로 수집할 수 있다.
- 수집한 데이터를 목적에 맞게 구분할 수 있다.
- 응용 소프트웨어를 이용하여 수집한 데이터를 관리할 수 있다.



1 문제 해결에 적합한 데이터

우리는 일상생활에서 다양한 문제를 마주하게 된다. 문제에 따라 데이터 없이 해결 가능한 것도 있지만, 다양한 데이터를 수집하고 분석하여 해결해야 하는 문제도 있다. 데이터를 이용해 문제를 해결하기 위해서는 문제 해결에 적합한 데이터를 수집하는 것이 중요하다.

예를 들어, ‘스마트폰 사용 습관을 개선하려면 어떻게 해야 할까?’라는 문제를 해결하기 위해서는 자주 사용하는 애플리케이션의 종류와 사용 횟수, 사용 장소, 스마트폰 총사용 시간 등의 데이터가 활용될 수 있다.

문제 스마트폰 사용 습관을 개선하려면 어떻게 해야 할까?



데이터 사용 앱 • 사용 횟수 • 사용 장소 • 총사용 시간 등

문제 앞으로의 날씨는 어떻게 예상할 수 있을까?



데이터 기온 • 습도 • 강수량 • 바람 세기 등

문제 운동 경기의 승부는 어떻게 예측할까?



데이터 신체 조건 • 개인 성적 • 경기 기록 • 팀 성적 등

문제 급식 잔반량을 줄일 수 있는 방법은 무엇일까?



데이터 식단 • 급식 인원 • 잔반량 • 날짜 등

[그림 1-7] 문제 해결에 적합한 데이터의 예시

2 데이터의 수집

❖ 검색 엔진

인터넷에서 원하는 데이터나 정보를 쉽게 찾을 수 있도록 도와주는 프로그램을 말한다.

데이터를 수집하는 방법에는 인터뷰, 관찰, 설문 조사, 센서 이용과 같이 데이터를 직접 수집하는 방법과 ❖ 검색 엔진, 데이터 전문 웹 사이트 등에서 이미 정리해 놓은 데이터를 수집하는 방법 등이 있다. 문제 해결 목적과 상황에 따라 적절한 데이터의 수집 방법을 선택한다.

● 데이터를 직접 수집하는 방법 | 목적에 맞는 정확하고 다양한 종류의 데이터를 수집할 수 있는 장점이 있으나 시간과 비용이 많이 드는 단점이 있다.



인터뷰

문제에 관한 질문과 답변을 통해 정보나 의견 등을 알아낸다.



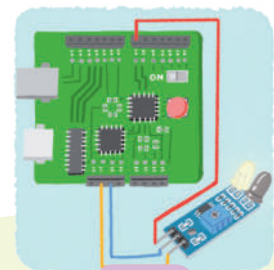
관찰

일정 시간 동안 대상을 오감 또는 관찰 도구를 사용하여 관찰하고 그 결과를 기록한다.



설문 조사

설문지를 이용하여 질문을 하고 답변을 모은다.



센서 이용

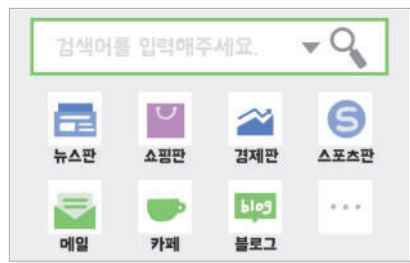
센서를 이용하여 빛, 온도, 습도 등의 다양한 데이터를 수집한다.

❖ 데이터 전문 웹 사이트

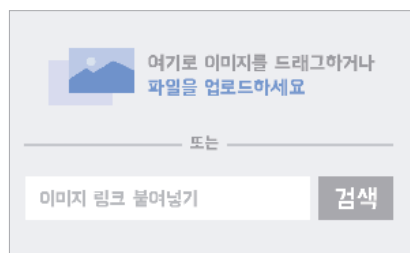
정부나 지방 자치 단체에서 제공하는 공공 데이터 웹 사이트와 데이터 전문 기업에서 제공하는 웹 사이트 등이 있다.

● 검색 엔진과 데이터 전문 웹 사이트를 이용하는 방법 | 원하는 데이터를 빠르고 편리하게 수집할 수 있지만, 검색 엔진은 데이터의 검증이 필요하고, ❖ 데이터 전문 웹 사이트는 문제 해결에 적합한 데이터를 찾기 어려울 수 있다.

검색 엔진



① 검색



② 이미지 검색

데이터 전문 웹 사이트



① 지방행정 인허가 데이터 개방



② 공공데이터포털



클릭!

공공데이터포털

<https://www.data.go.kr>

국가통계포털

<https://kosis.kr>

AI 허브

<https://www.aihub.or.kr>

캐글

<https://www.kaggle.com>

방송통계포털

<https://www.mediastat.or.kr>

3 데이터의 구분

수집한 데이터는 그 의미를 파악하여 공통점을 기준으로 구분하거나 **+** 형식에 따라 구분하여 문제 해결 목적에 맞게 활용될 수 있다. 예를 들어 다음과 같이 검색 엔진으로 수집한 일기 예보 뉴스 데이터를 살펴보자.

- 오늘 서울과 대전은 흐리고, 강릉에는 비가 내리겠으며, 광주와 부산, 제주는 맑겠습
- 니다. 서울의 최고 기온은 23 °C, 대전은 22 °C, 강릉은 20 °C, 광주는 26 °C, 부산은
- 28 °C, 제주는 29 °C로 예상됩니다.

● **의미, 공통점에 따른 구분** | 일기 예보 뉴스 데이터에서 데이터의 의미를 파악하고, 공통점을 찾아 묶을 수 있는 부분을 찾으면 지역, 날씨, 최고 기온이 있다. 이것은 다음과 같이 나타낼 수 있다.

- 지역: 서울, 대전, 강릉, 광주, 부산, 제주
- 날씨: 흐림, 흐림, 비, 맑음, 맑음, 맑음
- 최고 기온: 23 °C, 22 °C, 20 °C, 26 °C, 28 °C, 29 °C

여기서 지역, 날씨, 최고 기온과 같이 데이터의 의미, 공통점을 기준으로 묶을 수 있는 부분에 해당하는 것을 **+** 데이터 속성이라고 한다. 이렇게 데이터를 속성 별로 구분하여 정리하면 목적에 따라 필요한 정보를 빠르게 찾을 수 있고, 문제 발견과 분석, 의사 결정에 도움을 주어 데이터의 활용도를 높일 수 있다.

+ **형식에 따른 데이터 구분**
필요에 따라 문자, 이미지, 소리, 동영상 등 데이터의 유형 별로 데이터를 구분하여 정리할 수도 있다.

예 '고양이' 데이터 수집

- ① 고양이 소개 기사(문자)
- ② 고양이 사진(이미지)
- ③ 고양이 울음소리(소리)

➡ 고양이 모습을 보고 싶을 때는 이미지, 고양이 울음소리가 궁금할 때는 소리 등을 이용하여 목적에 맞게 활용할 수 있다.

+ **데이터 속성**

데이터의 특징을 말하며, 의미를 지니면서 더 이상 분리되지 않는 데이터의 최소 단위이다.

예 '학생'이라는 데이터에는 번호, 이름, 나이, 성별 등 학생을 설명하는 여러 속성이 있다.

지식 충전소

데이터 특성에 따른 종류

데이터의 종류는 특성에 따라 범주형과 수치형으로 구분할 수 있다. 범주형 데이터와 수치형 데이터는 각각 다른 분석 방법이 적용되어 컴퓨터가 데이터를 효과적으로 처리할 수 있다.

범주형 문자나 기호로 표시하여 대상을 구분할 수 있다.

수치형 숫자로 대상이 가진 값을 표시하고 크기나 차이를 비교할 수 있다.



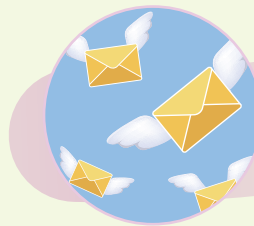
성별

예 남자, 여자



성취 등급

예 A, B, C



전자 우편 수

예 11개, 365개



추의 무게

예 1 kg, 2 kg

4 데이터의 관리

+ 응용 소프트웨어

컴퓨터를 특정한 목적(문서 작성, 이미지 편집 등)으로 사용하기 위해 제작한 소프트웨어이다.

▶ 데이터 관리

데이터를 입력한 후 수정, 추가, 삭제, 저장 등의 작업을 하는 것을 뜻한다.

+ 코답

(<https://codap.concord.org>)

데이터 관리 및 분석을 위한 웹 기반 실험 도구로 다음과 같은 장점이 있다.

- 프로그램 설치, 로그인, 회원 가입 없이 이용할 수 있다.
- 한글을 지원하며 직관적인 메뉴와 조작으로 데이터의 관리, 시각화, 분석을 쉽게 할 수 있다.

▶ 코답 메뉴



▶ '가족 여행 계획하기' 데이터의 종류별 구분

- 수치형: 위도, 경도
- 범주형: 여행지, 여행 기간, 교통수단, 숙박 시설

+ 응용 소프트웨어를 이용하면 수집한 데이터를 문제 해결 목적에 맞게 구분하여 효율적으로 관리할 수 있다. '가족 여행 계획하기'라는 문제를 해결하기 위해 수집한 데이터를 구분하고 관리하는 방법을 알아보자.

가족 여행 계획하기

데이터의 수집과 구분 | '가족 여행 계획하기'라는 문제를 해결하기 위해 필요한 데이터에는 여행지, 여행지 위치, 여행 기간, 교통수단, 숙박 시설 등이 있고, 데이터별로 수집 방법을 결정한다. 예를 들어 여행지는 인터뷰를 통해 추천받을 수 있으며, 여행지의 위치는 검색 엔진을 이용해 확인할 수 있다.

수집한 데이터

- ✓ 여행지: 경주, 부산, 여수, 전주, 제주 등
- ✓ 여행지 위치: 여행지별 위도와 경도 정보
- ✓ 여행 기간: 2박 3일, 3박 4일 등
- ✓ 교통수단: 버스, 기차, 비행기 등
- ✓ 숙박 시설: 펜션, 리조트, 호텔 등



수집한 데이터를 +코답(CODAP)을 이용하여 속성별로 구분하여 입력해 보자. 코답 웹 사이트에서 Launch CODAP - 새 데이터셋 - 테이블(테이블 아이콘)을 선택한다. 이어서 메뉴에서 새 데이터셋을 클릭한 후 여행지, 위도, 경도, 여행 기간, 교통수단, 숙박 시설의 데이터 속성을 기준으로 구분하여 입력한다.

| 가족 여행 계획 | | | | | | |
|-------------|-----|-------|--------|-------|------|-------|
| 데이터 (5 케이스) | | | | | | |
| 인덱스 | 여행지 | 위도 | 경도 | 여행 기간 | 교통수단 | 숙박 시설 |
| 1 | 경주 | 35.86 | 129.22 | 3박 4일 | 버스 | 펜션 |
| 2 | 부산 | 35.18 | 129.08 | 4박 5일 | 기차 | 리조트 |
| 3 | 여수 | 34.76 | 127.66 | 2박 3일 | 기차 | 리조트 |
| 4 | 전주 | 35.82 | 127.15 | 2박 3일 | 버스 | 호텔 |
| 5 | 제주 | 33.5 | 126.53 | 4박 5일 | 비행기 | 호텔 |
| | | | | | | |

[그림 II-8] 데이터 수집과 구분 후 입력 완료된 데이터

1 데이터의 정렬·편집·변환

여행 기간, 교통수단, 숙박 시설의 **+**데이터 속성을 기준으로 데이터를 정렬하여 보자. 예시로 교통수단 데이터를 ‘가나다’ 순서로 정렬하면 교통수단이 기차, 버스, 비행기 순으로 각각 구분되어 데이터를 교통수단별로 살펴볼 수 있다.

행: 가로로 나열된 데이터의 개별 항목

열: 세로로 나열된 데이터의 속성이나 범주

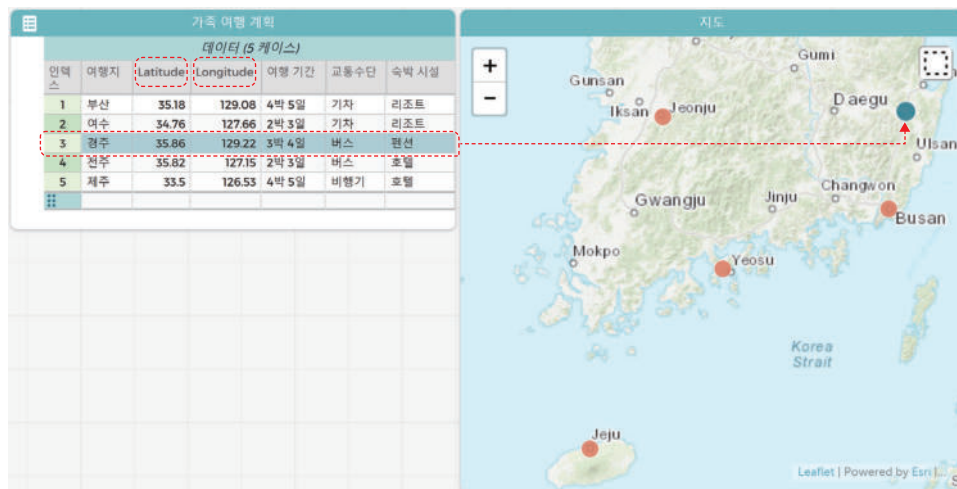
| 가족 여행 계획 | | | | | | |
|-------------|-----|-------|--------|-------|------|-------|
| 데이터 (5 케이스) | | | | | | |
| 인덱스 | 여행지 | 위도 | 경도 | 여행 기간 | 교통수단 | 숙박 시설 |
| 1 | 부산 | 35.18 | 129.08 | 4박 5일 | 기차 | 리조트 |
| 2 | 여수 | 34.76 | 127.66 | 2박 3일 | 기차 | 리조트 |
| 3 | 경주 | 35.86 | 129.22 | 3박 4일 | 버스 | 펜션 |
| 4 | 전주 | 35.82 | 127.15 | 2박 3일 | 버스 | 호텔 |
| 5 | 제주 | 33.5 | 126.53 | 4박 5일 | 비행기 | 호텔 |

칸(셀): 행과 열이 만나는 곳으로 값을 입력하는 단위이다.

[그림 9-] 교통수단 데이터를 ‘가나다’ 순서로 정렬한 결과

필요에 따라 새로운 열(데이터 속성)을 추가하거나 불필요한 열을 삭제할 수 있으며, 새로운 행(데이터 값)을 추가하거나 불필요한 행을 삭제할 수 있다. 그리고 데이터를 ‘더블 클릭’하여 데이터의 내용을 수정하며 편집할 수 있다.

지도(🌐) 기능을 사용하면 위도와 경도로 표시된 여행지 위치를 이미지 데이터로 변환하여 지도에 나타낼 수 있다. 숫자로 된 위치 정보를 지도 형태의 이미지 데이터로 변환하면 이해하기 쉬워서 데이터를 좀 더 효과적으로 관리할 수 있다.



[그림 10-] 여행지 위치를 지도 이미지 데이터로 나타낸 결과

+ 데이터 속성 기준 정렬하기

여행 기간, 숙박 시설도 각각 속성을 기준으로 정렬을 수행하여 데이터를 살펴본다.

▶ 교통수단 기준 정렬 방법

교통수단 속성을 클릭하고 메뉴에서 [오름차순 정렬(A → Z, 0 → 9)]을 선택한다.

- 이름 바꾸기
- 너비 맞추기
- 속성의 특성 편집
- 수식 편집
- 수식 제거(값 유지)
- 랜덤하게 만들기
- 오름차순 정렬(A → Z, 0 → 9)
- 내림차순 정렬(9 → 0, Z → A)
- 속성 숨기기
- 속성 삭제

▶ 코답 메뉴

- 테이블
- 그래프
- 지도
- 슬라이더
- 계산기
- 텍스트
- 플러그인

지도 기능을 이용하기 위해서는 위도는 Latitude, 경도는 Longitude로 각각 변경해야 해요.



2 데이터의 저장과 공유

목적에 맞게 수집한 데이터를 저장하여 정리하면 필요할 때 다시 불러와서 이용할 수 있다.

데이터를 저장할 때에는 프로그램 왼쪽 상단의 메뉴 목록 버튼(☰) - [저장] - [로컬 파일]을 선택한 후, [컴퓨터에 저장하기]를 클릭하면 ‘파일 이름.codap’ 형식으로 저장된다.



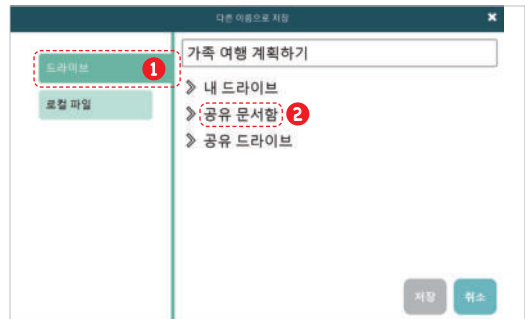
[그림 II-11] 데이터 저장

+ 클라우드 서비스

어디서나 사용자가 필요할 때마다 스마트폰이나 컴퓨터로 인터넷에 접속하여 문서, 음악, 동영상 등을 편리하게 저장, 처리, 이용할 수 있는 서비스이다

사회 관계망 서비스(SNS)나 클라우드 서비스 등을 이용하여 다른 사람과 협업하여 데이터를 새롭게 편집·가공하는 방식으로 데이터를 관리할 수 있다.

데이터를 저장할 때 [드라이브]를 클릭하고 원하는 위치를 지정하면 데이터를 공유할 수 있다.



[그림 II-12] 데이터 공유

지식 충전소

데이터 관리를 위한 응용 소프트웨어

데이터 관리에는 스프레드시트, 데이터베이스 소프트웨어를 이용할 수 있으며, 다음과 같은 특징이 있다.

| 종류 | 특징 | 활용 분야 |
|--------|------------------------|---|
| 스프레드시트 | 표 형태로 데이터를 분석·계산·관리한다. | 용돈(수입/지출) 기록장, 과목 점수별 성취도 자동 계산, 개인 일정 관리, 그래프 작성 등 |
| 데이터베이스 | 대량의 데이터를 효율적으로 관리한다. | 중학교 전체 학생의 성적, 인터넷 쇼핑몰의 고객 및 상품 데이터 관리 등 |

데이터 전문 웹 사이트의 데이터와 응용 소프트웨어를 활용하여 '내가 사는 지역의 무료 와이파이 위치'를 지도에 표시하고 파일로 저장해 보자.

1 데이터 전문 웹 사이트에서 문제 해결에 필요한 데이터를 수집해 보자.

- ① '지방행정 인허가 데이터 개방 웹 사이트(<https://www.localdata.go.kr>)'에 접속하여 상단 메뉴 [생활 편의 정보] - [생활 밀착 데이터] - [무료 와이파이 정보]를 클릭한다.
- ② 자신이 사는 지역(시도)을 선택하고, [검색] 버튼을 클릭한 후 **EXCEL** 을 클릭하여 데이터를 다운로드한다.



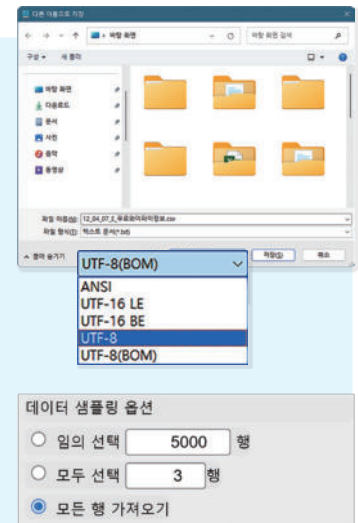
2 수집한 데이터를 코답으로 가져와 보자.

- ① 스프레드시트 프로그램을 이용하여 다운로드한 파일을 열고, [다른 이름으로 저장하기]를 이용하여 파일 형식을 csv 문서로 저장한다.

※ 코답 사용 가능 파일 형식: csv, txt, codap

Tip 한글이 보이지 않을 때는 [마우스 오른쪽 버튼] - [연결 프로그램] - [메모장]을 선택하여 파일을 열고, [파일] - [다른 이름으로 저장]을 클릭한 후 '인코딩'에서 'UTF-8'을 선택하여 저장한다.

- ② 코답을 실행하고, **새 문서** 를 클릭한 후 데이터 파일을 코답으로 끌어 놓는다. ('모든 행 가져오기'를 선택하여 전체 데이터를 가져온다.)



3 데이터를 편집, 변환, 저장, 공유하여 관리해 보자.

- ① 데이터 속성 중에서 설치 장소명, 설치 장소 상세, 와이파이 SSID, 소재지 도로명 주소, WGS84 위도, WGS84 경도만 남기고 나머지 데이터 속성들을 각각 클릭한 후 [속성 삭제]를 선택하여 모두 삭제한다.
- ② [이름 바꾸기]를 이용하여 속성명 WGS84 위도는 'Latitude'로, WGS84 경도는 'Longitude'로 각각 변경한다. (지도에 위치를 표시하기 위해 속성명을 영어로 바꾼다.)
- ③ 메뉴에서 지도(📍)를 클릭하고 오른쪽 화면과 같이 지도에 무료 와이파이 위치가 표시되면 파일을 저장하고 친구들과 공유한다.



예 무료 와이파이 지도 표시

| 설치장소명 | 설치장소상세 | 소재지도로명주소 | WGS84 위도 | WGS84 경도 |
|--------|--------|--------------------|----------|----------|
| 대우주식회사 | 대우주식회사 | 서울특별시 강남구 테헤란로 154 | 37.483 | 127.033 |
| 대우주식회사 | 대우주식회사 | 서울특별시 강남구 테헤란로 154 | 37.483 | 127.033 |
| 대우주식회사 | 대우주식회사 | 서울특별시 강남구 테헤란로 154 | 37.483 | 127.033 |
| 대우주식회사 | 대우주식회사 | 서울특별시 강남구 테헤란로 154 | 37.483 | 127.033 |

※ 같은 방식으로 지역의 '공중 화장실 위치', '자전거 보관대 위치' 등을 지도에 표시할 수 있다.