

공공데이터의 이해와 활용 가치 분석 보고서

2025.10.24
강원대학교 컴퓨터공학과 심우현

개요

공공데이터의 개념과 활용 가치를 이해하고, 향후 더존 ICT 인턴십 실무 프로젝트에서의 응용 가능성을 탐색하기 위해 작성하였습니다.

공공데이터의 개념 및 주요 제공 형태

1. 공공데이터는 정부 및 공공기관이 보유한 정보를 국민과 기업이 자유롭게 활용할 수 있도록 개방한 데이터입니다.
2. 공공데이터포털(data.go.kr)에는 약 4만 2천여 건의 데이터셋과 5천여 개의 오픈 API가 등록되어 있습니다.
3. 「공공데이터의 제공 및 이용 활성화에 관한 법률」에 따라 개방되며, 주요 형태는 아래와 같습니다.

구분	데이터 유형	주요 특징	제공 형식	대표 기관(예시)
①	오픈 API형	· 실시간 연동 가능 · 자동화 서비스 구현 용이 · 일일 트래픽 제한 존재, 응답 지연 발생 가능	REST API JSON/XML	기상청, 금융감독원
②	파일 데이터형	· 정적 데이터 · 분석·시각화 중심 활용	CSV, XLSX JSON	통계청, 국토교통부
③	LOD형	· 데이터 간 연결·통합 분석 · 시맨틱 웹 기술 활용 · 국내 적용 제한적	RDF	국립중앙도서관
④	공간정보형	· 위치 기반 시각화 · 지도 서비스 연계	GeoJSON Shapefile	통계청(SGIS), V-World

공공데이터의 활용 방식 및 기대 효과

1. 공공데이터는 다양한 산업·행정 분야에서 업무 효율화 및 자동화를 위한 기반으로 활용되고 있습니다.
2. 주요 기대 효과는 다음과 같습니다.
 - 행정 업무 효율화: ERP, 회계, 전자결재 등 내부 시스템과 연계
 - 정보 서비스 개선: 대시보드, 통계포털, 행정지원형 웹서비스 구현
 - 데이터 기반 의사결정 강화: 실시간·정적 데이터 분석을 통한 정책 지원
 - 산업 혁신 촉진: 민간 서비스와의 융합으로 신규 비즈니스 창출

공공데이터 접근 제약 이슈 및 대응 방안

1. 2025년 9월, 대전 데이터센터 화재로 인해 공공데이터포털(data.go.kr) 서비스가 일시 중단되며 4,000여 개의 API가 불능 상태가 되었습니다.
2. 이로 인해 실무 프로젝트 진행에도 제약이 있습니다.
 - 공공데이터 접근 제약 상황에 대응하기 위한 주요 방안을 아래 표와 같이 정리하였습니다.

구분	방안	주요 내용	비고
①	정적 데이터 기반 분석 전환	CSV 등 오프라인 데이터 활용, 일정 지연 최소화	단기 대응
②	대체 포털 활용	기상자료개방포털, OpenDART 등 사용	병행 가능
③	더미 데이터 기반 개발	복구 시 실시간 API로 전환	장기 대응
④	① + ③ 혼합 전략	단기: CSV 기반 프로토타입 구축 → 장기: API 방식 전환	가장 현실적인 방안

- 단기적으로는 정적 CSV 데이터를 활용하여 프로토타입을 구축하고, 복구 이후에는 실시간 API로 전환 함으로써 서비스 완성도와 안정성을 확보할 수 있습니다.

결론

1. 공공데이터는 행정 효율성과 산업 경쟁력을 동시에 높일 수 있는 핵심 인프라입니다.
2. 조사한 내용을 바탕으로 공공데이터를 활용한 행정지원형 대시보드 및 자동화 서비스 개발 기획을 추진하겠습니다.
3. 대전 데이터센터 화재와 관련해서는, 단기적으로 파일 데이터를 활용한 개발을 우선 수행하고, 복구 이후에는 실시간 API 방식으로 전환하여 안정적인 데이터 서비스 구조를 구축하겠습니다.

첨부 1. 오픈 API형 세부 내용 및 예시. 1부.

첨부 2. 파일 데이터형 세부 내용 및 예시. 1부.

첨부 3. LOD형 세부 내용 및 예시. 1부.

첨부 4. 공간정보형 세부 내용 및 예시. 1부. 끝.

오픈 API형 세부 내용 및 예시

오픈 API 사용 방법

1. 공공데이터포털에서 API 활용신청

- 공공데이터포털(www.data.go.kr) 회원가입 후 로그인
- 원하는 API 검색 및 '활용신청' 버튼 클릭

2. 인증키(API Key) 발급 받기

- 활용목적 작성 후 신청 완료 (일반적으로 즉시 승인)
- 마이페이지에서 일반 인증키(Encoding) 또는 Decoding 키 확인

3. 서비스 URL에 Key를 포함하여 HTTP 요청(GET/POST)

- API 문서에 명시된 요청 변수 확인
- URL에 필수 파라미터와 인증키를 포함하여 호출

4. 응답받은 JSON/XML 데이터를 파싱하여 실시간 서비스에 연동

- 응답 데이터를 파싱하여 필요한 정보 추출
- 웹/앱 서비스에 실시간 표시 또는 DB 저장

데이터 처리 방법

- JSON 파서로 필요한 필드 추출: response.body.items 배열에서 필요한 데이터만 선택적으로 추출
- 실시간 업데이트가 필요한 서비스에 직접 연동: 주기적(5분~1시간)으로 API 호출하여 최신 데이터 유지
- DB에 저장하거나 화면에 즉시 표시: 데이터 특성에 따라 실시간 표시 또는 히스토리 저장

데이터 세부 내용 및 특징

- 실시간성이 높아 최신 정보를 제공할 수 있으나, 응답 속도 및 트래픽 제한에 영향을 받음
- 기관별 API 구조가 달라, 필드명·응답 포맷의 표준화가 미흡한 경우가 있음
- 데이터의 변동성이 커, 서비스 설계 시 캐싱 또는 주기적 백업이 필요함
- API 호출 횟수 제한(일 1,000~10,000회)이 존재하여 트래픽 관리가 중요함

응답 데이터 예시 (기상청 단기예보 API, JSON형식)

```
{
  "response": {
    "header": { "resultCode": "00", "resultMsg": "NORMAL_SERVICE" },
    "body": {
      "items": [ { "baseDate": "20251024", "category": "T1H", "obsrValue": "15" } ],
      ...
    }
  }
}
```

파일 데이터형 세부 내용 및 예시

파일 데이터 사용 방법

1. 공공데이터포털 또는 기관 사이트 접속

- 공공데이터포털(www.data.go.kr) 또는 해당 기관 사이트 접속
- 원하는 데이터셋 검색 (키워드 또는 카테고리별 탐색)

2. 데이터셋 다운로드

- 데이터셋 상세 페이지에서 파일 형식 확인 (CSV, XLSX, JSON 등)
- '다운로드' 버튼 클릭하여 로컬에 저장
- 일부 데이터는 회원가입 없이 즉시 다운로드 가능

3. 로컬 환경에서 파일 열기

- Excel, 메모장, 또는 데이터 분석 도구로 파일 열기
- 파일 인코딩 확인 (UTF-8, EUC-KR 등)

4. 데이터 분석 및 활용

- 데이터 전처리 (결측치 제거, 형식 변환 등)
- 통계 분석, 시각화, 머신러닝 등 목적에 맞게 활용

데이터 처리 방법

- 엑셀 또는 분석 도구로 파일 로드: pandas(Python), readr(R) 등의 라이브러리 활용
- 데이터 정제 및 전처리: 결측치 처리, 이상치 제거, 데이터 타입 변환
- 통계 분석 및 시각화: 그래프 생성, 대시보드 구축, 보고서 작성

데이터 세부 내용 및 특징

- 정적 형태로 제공되어 안정적인 분석·시각화에 적합함
- 데이터 갱신 주기가 있어, 최신성 유지에는 한계 존재
- 대용량 파일은 메모리 부담이 크며, 일부 항목은 결측치 처리 필요
- 데이터 구조가 일정해 비교·추세 분석에 유리함

파일 데이터 예시 (지역별 인구수, CSV형식)

```
지역명,인구수,측정연도,증감률
서울특별시,9400000,2025,0.2
부산광역시,3300000,2025,-0.5
대구광역시,2400000,2025,-0.3
인천광역시,2900000,2025,0.8
광주광역시,1450000,2025,-0.2
대전광역시,1470000,2025,0.1
울산광역시,1140000,2025,-0.4
```

LOD형 세부 내용 및 예시

LOD(Linked Open Data) 사용 방법

1. 데이터 제공기관의 SPARQL 엔드포인트 확인

- 제공기관 웹사이트에서 LOD 서비스 페이지 접속
- SPARQL 엔드포인트 URL 확인 (예: <http://data.example.org/sparql>)
- 제공되는 온톨로지(Ontology) 및 스키마 문서 확인

2. SPARQL 쿼리 작성

- SPARQL 질의 문법에 따라 쿼리 작성
- PREFIX로 네임스페이스 정의
- SELECT, WHERE 절로 원하는 데이터 조건 지정

3. 엔드포인트에 쿼리 전송

- HTTP GET/POST 방식으로 엔드포인트에 쿼리 전송
- 웹 인터페이스 또는 프로그래밍 방식으로 실행

4. RDF 형식의 결과 데이터 수신 및 처리

- XML, JSON, Turtle 등의 형식으로 결과 수신
- RDF 라이브러리를 통해 데이터 파싱 및 활용

데이터 처리 방법

- SPARQL 질의어로 의미적 관계 탐색: 데이터 간의 연결 관계를 따라 정보 탐색
- 여러 데이터셋 간의 연결 관계 추론: 서로 다른 출처의 데이터를 의미적으로 통합
- 온톨로지 기반 지식 그래프 구축: 개념과 관계를 체계적으로 구조화
- 시맨틱 검색 및 추론: 단순 키워드가 아닌 의미 기반 검색 수행

데이터 세부 내용 및 특징

- RDF 구조를 기반으로, 데이터 간 관계·의미를 함께 표현 가능
- SPARQL 쿼리를 통해 지식 그래프형 분석 수행 가능
- 온톨로지 설계 수준에 따라 데이터 품질 및 활용 범위가 달라짐
- 국내 LOD 데이터는 제한적이므로, 주로 연구·실험 목적에 활용됨

LOD 데이터 예시 (도서관 장서 정보, RDF/Turtle 형식)

```
@prefix ex: <http://example.org/library/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

ex:Book001 rdf:type ex:Book ;
  dc:title "한국의 공공데이터 정책" ;
  dc:creator ex:Author001 ;
  ex:publishYear "2024" .

ex:Author001 ex:name "홍길동" ;
  ex:affiliation "서울대학교" .
```

공간정보형 세부 내용 및 예시

공간정보 데이터 사용 방법

1. 공간정보 오픈플랫폼에서 데이터 다운로드

- V-World(www.vworld.kr), SGIS(sgis.kostat.go.kr) 등 접속
- 원하는 지역 및 레이어(행정경계, 도로, 건물 등) 선택
- 파일 형식 선택 (GeoJSON, Shapefile 등) 후 다운로드

2. GIS 소프트웨어 또는 웹 라이브러리로 파일 열기

- QGIS, ArcGIS 등 전문 GIS 소프트웨어 활용
- 또는 Leaflet, Mapbox 등 웹 지도 라이브러리 사용

3. 좌표계(EPSG) 확인 및 필요시 변환

- 데이터의 좌표계 확인 (EPSG:4326, EPSG:5179 등)
- 웹 지도는 주로 WGS84(EPSG:4326) 사용
- 좌표계가 다를 경우 변환 작업 필요

4. 지도 위에 레이어로 시각화

- 포인트, 라인, 폴리곤 형태로 지도에 표시
- 속성 정보(이름, 주소 등)를 팝업으로 연결
- 사용자 위치 기반 서비스 구현

데이터 처리 방법

- 좌표 데이터(위도/경도) 추출: GeoJSON의 coordinates 필드에서 위치 정보 추출
- 웹 지도 API에 GeoJSON 로드: Leaflet, 카카오맵 등에 직접 로드하여 시각화
- 공간 분석 수행: 버퍼링(반경 검색), 교차 분석, 거리 계산 등
- 사용자 위치 기반 서비스 구현: GPS 정보와 결합하여 주변 시설 안내 등

데이터 세부 내용 및 특징

- 위치(위도/경도) 정보를 포함하여 시각화 및 공간 분석에 활용 가능
- 데이터 용량이 커, 필요한 행정구역 또는 지역 단위만 필터링하는 것이 효율적
- 좌표계 변환 과정에서 오차가 발생할 수 있으므로 변환 작업 필수
- 행정지도, 도시계획, 인프라 분석 등 공간기반 정책 수립에 활용됨

공간정보 데이터 예시 (서울시청 좌표, GeoJSON 형식)

```
{
  "type": "Feature",
  "geometry": {
    "type": "Point",
    "coordinates": [126.9780, 37.5665]
  },
  "properties": {
    "name": "서울시청",
    "address": "서울특별시 중구 세종대로 110"
  }
}
```