

# AI 데이터 공유·활용 서비스

상주 개발 웹 기타(IT 서비스 구축), 기타(AI·머신러닝), 네트워크·서버 운영

배경 성과 핵심 기능 프로젝트 상세

## 프로젝트 배경

### 1) 문제점

- 연구데이터 생산·활용은 급격히 증가하고 있으나, 인공지능 학습 데이터 공개는 여전히 부족함.
- 연구 투명성과 재현성을 위한 데이터 공유·개방은 확대되고 있으나, AI 연구자 관점에서 활용 가능한 데이터 인프라는 미흡함.
- 연구자들이 원하는 방식으로 데이터에 접근하고 분석할 수 있는 환경과, 사전학습 모델·맞춤형 모델을 활용할 수 있는 기반이 필요함.

### 2) 프로젝트 목표

- 인공지능 연구를 위한 데이터 공유·활용 서비스 개발 및 운영
- 대규모 연구데이터를 효율적으로 저장·관리할 수 있는 아키텍처(과학기술 데이터 댐) 설계 및 개발
- 데이터·AI 기반 문제 해결을 지원하는 플랫폼 구현 및 데이터톤 개최

### 3) 주안점

- 데이터 품질 강화 및 기계가 읽을 수 있는 형태로 정제
- DOI 기반 검색, OpenAPI 제공을 통한 데이터 접근성·재사용성 향상
- 머신러닝 코드·모델을 표준화된 방식으로 공유할 수 있는 체계 마련
- 대규모 데이터 저장·관리 아키텍처(과학기술 데이터 댐) 구현
- 효율적인 데이터 처리·분석 환경과 AI 모델 연계 기반 확보
- 사전학습·분석 모델 제공을 통한 연구 생산성 제고
- 데이터톤·리더보드·워크플로우 기반 협업을 통한 실증 및 우수 모델 확산
- 연구자 커뮤니티 활성화 및 문제 해결 중심의 개방형 협력 생태계 조성

## 프로젝트 성과

### 월간 활성 사용자(MAU) 4만 명 달성

서비스 개시 2년 만에 4000 MAU -> 40000 MAU 10배 달성

### 대규모 연구데이터 저장·관리 아키텍처(과학기술 데이터 댐) 구축

약 6,000GB에 달하는 2,033,174건의 데이터 저장 및 제공

### AI 데이터 공유·활용 서비스 개발 및 안정적 운영

- 69 건의 모델/데이터셋 제공
- 연간 서비스 가용성 99.9% 이상 유지
- 중대 장애 0건

### OpenAPI, DOI 기반 검색 기능 제공으로 데이터 접근성 및 재사용성 강화

- 8000 여건 이상 외부 기관에서 OpenAPI 활용
- 69건의 데이터셋에 대해 DOI를 부여하여, 모든 데이터셋이 학술적으로 인용 가능한 형태로 제공됨

### 데이터톤·리더보드 운영을 통한 우수 AI 모델 발굴·확산



paulhana6006

개발 · 개인

#### 참여 기간

2022.06. ~ 2023.11.

#### 참여율 ⓘ

30%

#### 고객사

KISTI

#### 역할

개발

#### 관련 기술

IRODS jpa spring cache

Pebble Spring Boot

PostgreSQL Docker

Vue.js Redis

Spring Scheduler

spring Security soa

WordPress Maven

- 연간 1회 데이터톤 개최 약 140명 이상 참가
- 문제 신청 40건 이상

## 핵심 기능



### 대규모 연구데이터 저장·관리 아키텍처 (iRODS 기반 데이터 댐)

- 대규모 과학기술 데이터를 안정적으로 저장/관리
- SFTP, davrods, Redis, PostgreSQL 등과 연계된 분산형 저장소 운영



### 데이터 검색·활용 서비스 (DOI·Solr·OpenAPI 기반)

- DOI 기반 데이터 검색 및 메타데이터 제공
- OpenAPI를 통한 외부 서비스 연동 및 데이터 재사용성 확대



#### AI 연구자 포털 (사용자·관리자 포털 제공)

- 사용자: 데이터 활용 신청·조회·저장소 제공, 개인화 서비스
- 관리자: 모델/데이터셋, 사용자, 로그·통계 관리



#### AI 경진대회 플랫폼

- 데이터톤, 리더보드, 문제/Q&A 관리
- 연구자 커뮤니티 활성화 및 우수 모델 확산

### 프로젝트 상세

두개의 프로젝트가 진행된 포트폴리오입니다.

KISTI 과학기술기계학습데이터서비스(2022.06~2022.11)

KISTI AI 데이터 공유·활용 서비스 구축 및 운영(2023.06~2023.11)

#### 1) 시스템 구조

[Frontend Layer]

- NGINX front
- Next wordpressClient

#### [Application Layer]

- Tomcat adm
- Tomcat serv
- Tomcat openapi
- Tomcat user
- Tomcat competition
- Tomcat kafe
- Wordpress

#### [Infra / Platform Layer]

- Solr
- davrods
- davrods-ticket
- sftp

#### [Database Layer]

- Postgres repo-db
- Redis session-db

#### [External Storage]

- iRODS Cluster (3 Servers)

#### [Monitoring Layer]

- Prometheus
- Grafana

## 2) 주요 모듈 스펙

### common 모듈

개발 언어: Java 8

프레임워크: Spring Boot 2.5

- Spring Security (SAML 지원)
- Spring Session (Core, Redis)
- Spring Boot Starter Data Redis
- Pebble (템플릿 엔진)
- Springfox Boot Starter (Swagger/OpenAPI)

데이터베이스: Redis 6

기타 주요 라이브러리:

- Apache SolrJ (검색)
- netCDF (데이터 파일 처리)
- AspectJ Weaver (AOP)
- OpenSAML (SAML 처리)

### user모듈

개발 언어: Java 8

프레임워크: Spring Boot 2.5

- Spring Boot Starter Web
- Spring Boot Starter Validation
- Spring Boot Starter Cache
- Spring Boot Starter Data JPA
- Spring Boot Starter Actuator (Prometheus 모니터링)

데이터베이스: PostgreSQL 14

기타 주요 라이브러리:

- kr.re.kisti.idr: common (공통 모듈)
- Apache POI 4.1.0 / POI-OOXML 4.1.0 (엑셀/오피스 문서 처리)
- Apache SolrJ (검색)
- jsoup 1.12.1 (HTML 파싱)
- ModelMapper 2.3.8 (객체 매핑)
- Micrometer Registry Prometheus (Prometheus 모니터링 메트릭)

#### adm모듈

개발 언어:

- Java 8

프레임워크: Spring Boot 2.5

- Spring Boot Starter Web
- Spring Boot Starter Validation
- Spring Boot Starter Cache
- Spring Boot Starter Data JPA
- Spring Boot Starter Actuator (Prometheus 모니터링)

데이터베이스:

- PostgreSQL 14
- H2 Database (테스트용)

기타 주요 라이브러리:

- kr.re.kisti.idr:common 4.0.0 (공통 모듈)
- Apache POI 4.1.0 / POI-OOXML 4.1.0 (엑셀/오피스 문서 처리)
- Liquibase Core 4.7.1 (DB 마이그레이션)
- JTS Core 1.18.2
- Spatial4j 0.8
- iRODS Jargon Core 4.3.2.5
- iRODS Jargon Ticket 4.3.2.5
- OpenCSV 5.5.2
- Springfox Swagger UI 3.0.0 (API 문서화)
- Micrometer Registry Prometheus
- JavaMail 1.4.7
- XMLSchema Core 2.2.5
- DSpace Handle 9.2

#### kafe 모듈

개발 언어: Java 8

프레임워크: Spring Boot 2.5

- Spring Boot Starter Security
- Spring Session (Core, Data Redis)
- Spring Boot Starter Data Redis

보안 및 인증:

- Spring Security SAML2 Core (SAML 기반 인증)

#### serv 모듈

개발 언어: Java 8

프레임워크: Spring Boot 2.5

- Spring Boot Starter Data JPA (spring-boot.version 사용, Tomcat JDBC 제외)

데이터베이스: PostgreSQL 14

주요 라이브러리:

- kr.re.kisti.idr:common (4.0.0)
- JThumbnail (이미지 썸네일)
- Apache Tika (1.20, 파일 형식 식별/파싱)
- iTextPDF (5.5.10, PDF 처리)
- ModelMapper (2.3.3, 객체 매핑)
- Spire.Office.Free (3.9.0, 오피스 문서 처리)
- GeoTools (24.0, gt-shapefile / gt-epsg-wkt)
- Apache Commons Compress (1.20)
- Apache Commons CSV (1.4)

#### sftp모듈

개발 언어: Java 8

주요 라이브러리:

- Apache FtpServer Core (1.1.4)
- iRODS Jargon Core (4.3.1.0-RELEASE)

#### wordpressClient모듈

개발 언어: JavaScript (Node.js >= 14.15.4, npm >= 6.14.10)

실행 환경: Node.js

프레임워크: Next.js (10.2.0)

프론트엔드 라이브러리:

- React (17.0.2)
- React DOM (17.0.2)
- classnames (2.3.1)
- next-seo (4.24.0, SEO 관리)
- nprogress (0.2.0, 로딩 표시)
- sass (1.32.13, 스타일링)
- @wordpress/base-styles (3.5.2)
- @wordpress/block-library (3.2.4)
- react-cookie (4.0.3)
- react-mailchimp-subscribe (2.1.3)

주요 유틸리티 라이브러리:

- @apollo/client (3.3.16, GraphQL 클라이언트)
- graphql (15.5.0)
- axios (0.21.1, HTTP 클라이언트)
- cookie (0.4.1)
- lodash (4.17.21)
- dompurify (2.2.8, XSS 방어)
- uuid (8.3.2, 고유 ID 생성)
- validator (13.6.0, 데이터 유효성 검사)
- TailwindCSS (1.8.10)
- prop-types (15.7.2)

openapi 모듈

개발 언어: Java 8

프레임워크: Spring Boot 2.5

- Spring Security Core
- Spring Data Commons
- SnakeYAML

기타 주요 라이브러리:

- kr.re.kisti.idr:common (SAML, Spring Security, Redis 관련 의존성은 제외)

competition 모듈

개발 언어: Java 8

프레임워크: Spring Boot 2.5

- Spring MVC (spring-boot-starter-web)
- Spring Boot Actuator (Prometheus 모니터링)
- Spring Session (Redis)
- Spring Security (OAuth2, SAML2, egovframe 보안 모듈)

기타 주요 라이브러리:

- Apache Tomcat Embed Jasper (JSP 지원)
- Apache Commons (DBCP2, Codec, Compress, FileUpload 등)
- Apache POI (엑셀 파싱, poi / poi-ooxml)
- PostgreSQL JDBC Driver
- log4jdbc (SQL 로깅)
- Log4j2 (보안 이슈 대응 버전 2.17.2)
- Tiles (뷰 레이아웃 관리)
- Jackson (Databind, Core, Annotations)
- HttpClient / HttpMime (RestTemplate, 파일 전송)
- JSON-Simple
- ICU4J (국제화 지원)
- Smile-core (FScore 관련)
- egovframe RTE 4.0 (표준프레임워크: mvc, dataaccess, idgnr, property, security)

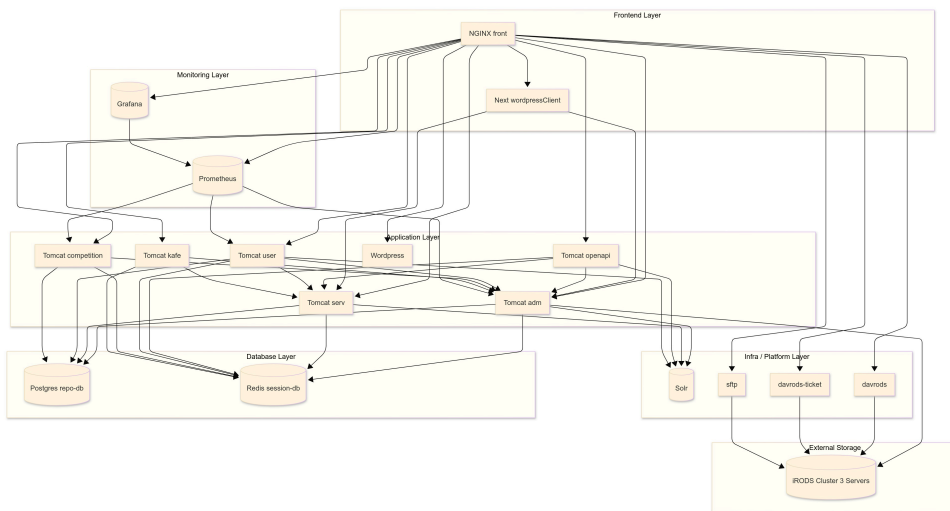
3) 주요 서비스 역할

- common: 시스템 전반에서 재사용되는 공통 모듈(도메인 객체, 변환기, 검색, 세션, 서비스 로직 보조, 유틸리티, 상수/열거형 등)

- adm: 운영자가 시스템을 관리·운영할 수 있는 관리자 포털(통계, 모델/데이터셋, 사용자, IRODS 저장소, 활용신청, 게시판, 로그 등)
- user: 사용자에게 데이터·모델을 제공하는 포털(데이터 활용신청·조회·제출, 개인화 기능(알림, 마이페이지, 저장소))
- kafe: Kafe 기반 인증 구현 모듈(redis 세션연동을 통한 사용자 인증 및 권한 부여)
- serv: 데이터 파일 처리의 중심 모듈(텍스트 추출, 미리보기, 개인정보 식별, 파일 복사 기능 등)
- sftp: SFTP 모듈(IRODS 시스템과의 연동 해 파일 업로드/다운로드 처리)
- openapi: 데이터 조회·다운로드·컬렉션 관리·북마크·활용신청·저장소 연계·통계 등 외부 서비스 연동을 위한 API 제공
- competition: 데이터·AI 경진대회를 운영을 위한 서비스(문제, 참여자, QnA, 공지사항, 심사위원 기능 등)
- wordpressClient: WordPress 서버와 연동한 프론트엔드 클라이언트로 관련 문서 및 가이드 콘텐츠 제공

#### 4) 서비스 아키텍처 구조도

\* 포트폴리오 이미지를 참고해주세요



서비스 아키텍처 구조도(docker 기반 MSA)