

# 카드 데이터베이스 구축 및 검색 서비스 설계

정승주 이연걸 이상평 박철준 이혁준 김동주  
광운대학교 컴퓨터정보공학부

wjdtmd987@kw.ac.kr, dusrf87@kw.ac.kr, miaspl105@kw.ac.kr

cheoljun99@kw.ac.kr, hlee@kw.ac.kr, gggg8657@gmail.com

## Design of CAD File Database and Search Service

Seungju Jeong Yeongeol Lee Sangpyeong Lee Cheoljun Park Hyukjoon Lee Dongju Kim

School of Computer and Information Engineering, Kwangwoon University

### 요 약

본 논문은 빠르게 증가하는 데이터와 끊임없이 변화하는 요구사항에 대응할 수 있는 데이터 수집, 저장, 검색에 최적화된 서비스를 설계한다. 이러한 서비스를 제공하기 위해 비정형 데이터인 카드 데이터에 적합한 비관계형 데이터베이스인 MongoDB와 전문 검색 엔진인 엘라스틱서치(Elasticsearch)를 사용한다.

### I. 서 론

컴퓨터 보조 설계(Computer Aided Design: CAD)는 그 효율성과 높은 생산성 때문에 설계 분야에서 널리 활용되고 있다. 그러나, 다수의 엔지니어링 기업들은 빠르게 증가하는 카드 데이터를 관리하지 못해 많은 어려움을 겪고 있고, 또한 카드 데이터는 비정형 데이터이기 때문에 스키마와 구조가 없고 색인 생성 및 읽기, 정보 추출이 어렵다.[1]. 이를 해결하기 위해 엔지니어링 서비스 지원을 위한 클라우드 기반 빅데이터 플랫폼 개발을 연구하고 있고[2], 한국생산기술연구원에서 운영중인 데이터를 보관하고 검색을 지원해주는 빅데이터 플랫폼을 개발한 사례가 있다[3]. 그러나 기업의 입장에서 외부 플랫폼을 이용하는 것은 핵심 데이터를 외부 기업에 노출해야 하는 위험이 있기 때문에 온프레미스 방식을 고수하고 있다. 하지만 자체 서버를 두고 관리하는 온프레미스 방식은 데이터의 확장성과 유연성을 확보할 수 없다. 이에 온디맨드 방식인 클라우드에 저장하며, 클라우드 속 카드 데이터에 대한 접근성을 위해 데이터 종류와 속성 분석을 통하여 데이터베이스를 구축한다. 이를 통해 카드 데이터의 속성과 종류, 제목, 색인을 알 수 있으면 검색했을 때에 해당 데이터에 손쉽게 접근할 수 있다. 따라서 본 논문에서는 카드 파일 내에 포함된 데이터의 종류 및 속성을 분석하고 이를 바탕으로 데이터베이스를 설계한다. 또한 원하는 파일을 쉽게 찾을 수 있는 검색 서비스 시스템을 설계한다.

카드 파일의 속성을 나타내는 외부 데이터를 활용하여 데이터베이스를 구축하고 상세 검색을 위해 내부 데이터를 활용한다.

### 2.2 DBMS 설계 구성도

본 논문에서 사용하는 카드 데이터는 고정된 형태가 없어 스키마를 만들기 어렵고 스키마 변경이 자주 발생한다. 따라서 관계형 데이터베이스보다 비관계형 데이터베이스를 사용하여 유연하고 자유로운 데이터 구조로 구현한다.[4] 비관계형 데이터베이스의 특성상 단일 테이블 형식으로 작성해야 하고 각 카드 파일에서 데이터 추출의 결과가 파일마다 다르기 때문에 하나의 인덱스에 저장해야 한다. 또한 같은 프로젝트 내에 동일한 이름의 카드 파일이 있는 경우도 고려해야 하기에 기존 파일 경로를 유지해야 하는 필요성이 존재한다. 따라서 이에 부합하는 모델 중 칼럼 패밀리 모델을 이용하여 데이터베이스를 구축하였다.

```
{
  "id": "12jdb55hflien3@j",
  "name": "범례 및 일반사항",
  "extension": "dwg",
  "mainCategory": "스마트도시",
  "subCategory": "공공화장실비상벨",
  "index": {
    "mainIndex": "diagram",
    "subIndex": [
      "circuit wiring diagram",
      "switchboard diagram",
      "..."
    ]
  }
}
```

그림 2 - 비관계형 데이터베이스를 이용한 DBMS 구성도

카드 파일 분석을 통해 설계된 데이터 구조에 따라, 칼럼으로 사용된 mainCategory는 카드 파일이 위치한 프로젝트의 제목이고 subCategory는 해당 프로젝트의 기술 분야, 용역 분야에 해당하는 데이터를 의미한다. 카테고리별 메인과 서브로 나눔으로써 프로젝트 분야 별 성격과 용역 분야 별 성격에 따라 분리하여 효율적인 데이터 접근이 가능하다. mainIndex와 subIndex는 각각 카드 데이터의 최빈 텍스트와 제목데이터, 추출된 모든 인덱스를 의미하며 mainIndex의 존재로 인하여 마찬가지로 효율적인 데이터 접근과 검색의 효율성을 높였다.

## II. 카드 데이터 분석 및 데이터베이스 설계

### 2.1 카드 데이터 분석

	Name	Extension	mainCategory	subCategory	index	
					mainIndex	subIndex
1	인전 1호선 역사 무선망 배전망	dwg	인전 1호선	배전망	Diagram	Circuit wiring diagram
2	경춘선 망우 - 춘천간 철도 통합무선망 시스템 구성도	dwg	경춘선	철도 통합 무선망 시스템 구성도	안테나	GPS안테나

그림 1 - 카드 데이터 분석 결과

카드 데이터에는 외부 데이터와 내부 데이터가 있다. 내부 데이터는 카드 파일에 기재된 각종 장비의 정보, 위치 정보, 수치 정보, 통신 선의 종류 등이 있고, 외부 데이터는 카드 파일이 속한 프로젝트의 종류, 카드 파일의 속성, 카드 파일의 제목이 있다. 카드 내부 데이터는 검색을 위해 분류할 수 있는 기준이 모호하고 형식이 다르기 때문에 이를 데이터베이스 설계를 위해 활용하기 어렵다. 따라서

### III. 검색 시스템 설계

#### 3.1 검색 시스템 구성도

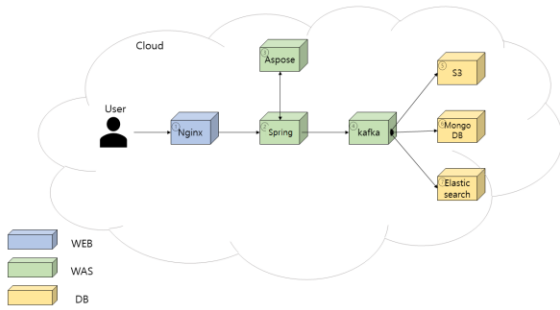


그림 3 - 검색 시스템 구성도

다음은 검색 시스템 구성도이다. 보안의 민감한 데이터를 다루기 위해 WEB, WAS, DB 3 Tier 로 구성되어 있다. WEB 에는 프록시를 통한 보안 향상을 위해 Nginx 를 사용하였고 WAS 에는 추출을 위한 Aspose 라이브러리, 라이브러리 사용과 요청 분기를 위한 spring, 원활한 데이터 통신을 위한 kafka 를 사용하였다. DB 에는 효율적인 로우 데이터 저장을 위한 클라우드 스토리지인 S3, MongoDB, 역색인 검색을 사용해 빠르게 검색 결과를 가져오는 엘라스틱서치로 구성되어 있다. 이러한 모든 프로그램은 클라우드에서 동작한다.

전체적인 흐름은 다음과 같다. 유저가 홈페이지에서 검색 혹은 저장 버튼을 누르면 명령이 Nginx 를 통하여 spring 으로 들어간다. 명령어가 저장일 경우 Aspose 라이브러리를 통하여 키워드 데이터를 추출하여 원본 데이터는 S3 에, MongoDB 에는 키워드 데이터가 저장된다. 명령어가 검색일 경우 엘라스틱서치를 통하여 검색이 이루어진 후 검색 결과를 사용자에게 시각적으로 보여준다.

#### 3.2 WEB 설계 구성도

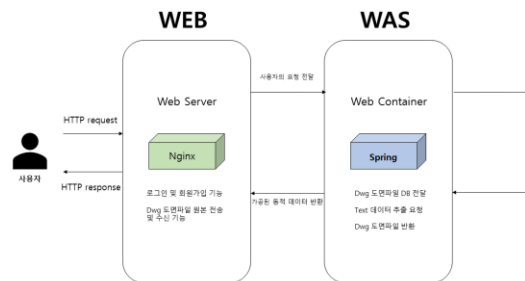


그림 4 - WEB 설계 구성도

웹 서버는 사용자 인터페이스를 제공한다. 모든 사용자에게 파일 저장 권한이 있다면 보안문제가 발생할 수 있다. 때문에 회원 인증을 받은 사용자만 HTTP 요청 권한이 있고 이 사용자가 파일 저장 요청을 할 경우 웹 서버는 캐드 파일을 입력 받아 파일 저장 요청을 WAS 서버로 전달한다. 그 다음 WAS 서버로부터 결과 동적 데이터를 받아 사용자에게 HTTP 응답을 한다. 사용자가 검색 요청을 할 경우 웹 서버는 검색어를 입력 받아 검색 요청을 WAS 서버에게 전달하고 WAS 서버로부터 검색 동적 결과를 받아 사용자에게 HTTP 응답을 한다.

#### 3.3 WAS 설계 구성도

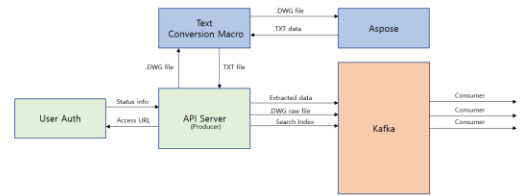


그림 5 - WAS 설계 구성도

WAS(Web Application Server)는 사용자 인증 기능, Aspose 라이브러리를 활용한 캐드 파일에서의 텍스트 데이터 추출 기능, 도면 파일과 추출된 텍스트 데이터를 S3, MongoDB 에 저장하는 기능, 추출된 텍스트 데이터를 엘라스틱서치를 활용해 검색된 데이터를 사용자에게 보여주는 검색 기능이 있다. 그 중 캐드 파일에서 텍스트 데이터를 추출하는 기능은 Aspose 라이브러리가 사용된다. 본 라이브러리는 Aspose사에서 만든 것으로 캐드 파일을 넘겨주면 파일 내에 존재하는 텍스트 데이터를 모두 추출하는 역할이다. 따라서 해당 라이브러리를 통하여 도면 파일의 텍스트 데이터를 추출한다. 다음은 엔지니어링 원본 데이터와 추출된 텍스트 데이터를 S3, MongoDB 에 저장하는 기능이다. 엔지니어링 데이터 중 도면 파일은 보통 MB 단위로, 해당 데이터를 MongoDB 에 저장하면 검색을 하거나 접근하여 가져올 때 상당히 많은 시간이 소요된다. 따라서 클라우드 S3 스토리지를 사용하여 원본 엔지니어링 데이터를 S3 스토리지에 저장한다. 또한 추출된 텍스트 데이터는 상대적으로 크기가 작기 때문에 MongoDB 에 저장하고 S3 에 있는 원본데이터와 링크를 시켜 MongoDB 탐색 한 번으로 S3 에 접근해 원본 엔지니어링 데이터를 가져올 수 있도록 한다. 마지막은 엘라스틱서치를 이용한 검색 기능이다. 방대한 양의 데이터를 실시간으로 검색할 수 있도록 엘라스틱서치를 사용한다. 저장기능에는 탁월하지만 검색기능의 성능은 떨어지는 MongoDB 와 연동하여 사용하면 데이터베이스에서 전체 텍스트 검색을 실시간으로 수행하는 것을 최적화할 수 있다.

### IV. 결론 및 향후계획

본 논문에서는 점점 거대해지는 캐드 데이터를 기업이 직접 효과적으로 관리하는 방법을 제시하고 있다. 이를 통해 기업은 추가되는 데이터에 대하여 편리하게 관리할 수 있고 검색 시스템을 구현하여 원하는 데이터에 손쉽게 접근할 수 있다. 더불어 보안에 민감한 데이터를 클라우드에서 직접 관리함으로써 데이터 안전성에 대한 불안감도 해소하는 효과를 기대할 수 있다. 향후에는 구축한 데이터베이스와 추가되는 데이터를 참고자료로 활용하여 더 정확하고 연관성 있는 용역 설계가 가능할 것으로 기대된다. 또한 데이터베이스에 머신러닝 알고리즘을 적용하여 연관된 도면 파일을 데이터베이스에서 찾아 초안을 자동으로 만들어 주는 연구를 진행할 계획이다.

### 참 고 문 헌

- [1] 최명석, 이상환, “데이터 관리 계획의 국내 현황 및 과제”, 한국콘텐츠학회 논문지, vol.20, no.6, pp. 220-229, 2020
- [2] 서동우, 김명일, 박상진, 김재성, 정석찬, “엔지니어링 서비스 지원을 위한 클라우드 기반 빅데이터 플랫폼 개발 연구”, 한국빅데이터학회지, vol 4, no.1, pp. 119-127, 2019
- [3] “ENG BIG DATA”, 엔지니어링 빅데이터 플랫폼, 2023년 1월 18일 접속, <https://bigdata-eng.com/>
- [4] 이미영, 최완, “빅데이터 처리 및 저장관리 기술동향 및 전망”, 한국정보통신학회지 vol 13 no.1, pp. 33-39, 2012