

# 수요자 중심의 빅데이터 기반 공공정보시스템 구축에 관한 연구

조경모\*

\*(주)지오인프라

e-mail:chokm@gioinfra.co.kr

## A Study on the Public Information System Construction of Consumer-Centric Big Data Foundation

Kyung-Mo Cho\*

\*Gioinfra Co., Ltd

### 요약

4차산업혁명 시대의 핵심기술인 빅데이터 기술에 대한 연구는 국내·외에서 활발하게 연구를 수행하고 있으며, 이를 기반으로 서비스를 제공하는 정보시스템이 설계되고 있다. 빅데이터 기술을 기반으로 수요자의 정보서비스의 품질을 높이기 위해 국가기관 및 민간기관에서 빅데이터 센터가 구축되어 정보서비스를 제공하고 있다. 그러므로 본 논문에서는 공공정보시스템에서 빅데이터 기술을 기반으로 수요자 맞춤형 정보시스템 구축에 대한 연구를 수행하고자 한다.

## 1. 서론

거대하고 방대한 정보를 빅데이터라 정의하고 있으며, 이러한 정보는 미래 시대의 핵심 전략기술로 주목받고 있다. 과학기술정보통신부에서는 공공과 민간의 협업을 통해 데이터 생산, 수집, 분석, 유통에 대한 빅데이터 기술을 연구하여 혁신 서비스 발굴, 확산하는 생태계를 조성하려고 하고 있다[1][2][6].

빅데이터 기술은 복잡하지만, 다변화된 현대사회의 수요자 중심의 정보서비스를 제공하기에는 적합하고, 이러한 기술을 기반으로 정확하게 예측된 데이터를 제공함으로써 사용자의 편의성 및 서비스 향상을 높일 수 있다. 본 논문에서는 수요자 중심의 빅데이터 기반의 공공정보시스템 구축에 관한 연구를 하고자 한다.

## 2. 관련연구

### 2.1 빅데이터 정의 및 특징

빅데이터는 대규모 데이터이며, 기존의 전통적인 방식으로 처리할 수 없다. 빅데이터는 “3V”로 불리며 Volume(규모), Variety(다양성), 속도(Velocity)의 특성을 가지고 있다. 이후 Value(가치)와 Veracity(정확성)이 추가되어서 빅데이터의 활용도는 확장되었다[1][5].

빅데이터의 주요 특징은 다음과 같다. 첫번째 Volume(규모)는 데이터 크기는 물리적 한계를 초과할 정도로 거대하며 기하급수적으로 증가하고 있다. 두 번째 Variety(다양성)은 데이터 정형, 비정형으로 다양하게 섞여 있다. 텍스트, 데이터베이스 뿐만 아니라 외부의 SNS 데이터, 로그기록과 같은 데이터 유형이 있다. 세 번째 속도(Velocity)에서는 실시간으로 데이터를 빠르게 처리한다. 네 번째 Value(가치)에서는 데이터의 정확성과 시간성을 관련지어서 데이터의 가치를 찾아야 한다. 마지막으로 Veracity(정확성)에서는 데이터의 노이즈를 제거하고 시그널을 확보하여 데이터의 신뢰성을 확보할 수 있어야 한다[2][3][6].

### 2.2 빅데이터 처리 과정에 대한 절차

빅데이터의 처리 과정은 6단계로 처리된다. 첫 번째 데이터 인식단계가 있으며 각종 데이터에 따른 데이터에 대한 유형을 인식한다. 두 번째 데이터의 수집단계는 외부에서 수신된 소스에 대해서 수집한다. 세 번째 저장단계에서는 데이터의 분석 및 활용을 수행하는 단계이다. 네 번째 데이터 처리단계에서는 일련의 병렬 및 업무처리 단계를 거쳐서 처리업무를 수행한다. 다섯 번째 분석단계에서는 사실, 추세, 관계, 패턴 등의 정보를 찾아가는 과정을 수행한다. 여섯 번째 데이터 표현단계에서는 시각적인 수단으로 데이터를 사용자에게 정보를 전달한다[4-6].

### 3. 시스템 설계

#### 3.1 시스템 구축에 따른 수행 절차

빅데이터 기반의 공공정보시스템 체계개선을 수행하기 위해서는 운영하고 있는 정보시스템의 통합데이터 분석을 수행해야 한다. 공공정보시스템을 방문하는 사용자가 하나의 키워드로 다양한 공공데이터의 정보를 제공하기 위해서는 통합 정보서비스에 대한 제공기능을 수행해야 한다. 시각적인 부가적 기능 수행을 위해서는 시각화모듈(그래프, 맵, 테이블, 지도 등)이 설계되어야 하고, 공공데이터 포털, 각종 외부 정보제공서비스(SNS)의 연계를 통한 통합정보가 제공되어야 한다. 그리고 사용자 편의를 위한 공공정보시스템의 GUI(그래픽 유저 인터페이스) 개선이 필요하고, 사용자의 다양한 단말기(스마트폰, 태블릿PC, 소형 노트북 등)를 통해 손쉽게 정보제공 및 활용을 위한 접근성이 개선되어야 한다.

앞서 언급한 시스템 구축을 위해서는 Agile 개발 방법론을 통해 수요자 중심의 빅데이터 기반 공공정보 시스템이 설계되어야 한다. Agile 개발 방법론은 아래 [그림 1]과 같다. 첫 번째 개발환경 조성을 위해 개발 시스템장비 설계 및 구축, 네트워크 환경 설계 및 구축, 샘플 데이터가 준비되어야 한다. 이후 요구사항 분석, Use Case에 따른 시나리오 설계, 품질관리 기준이 설정되어야 한다. 다음으로 사용자 인터페이스 구현, 기능 단위 모듈 및 시범 서비스가 구현되어야 한다. 마지막으로 품질 테스트를 통해 기능, 성능, 사용자 경험 환경에 따른 테스트가 완료되어야 한다.



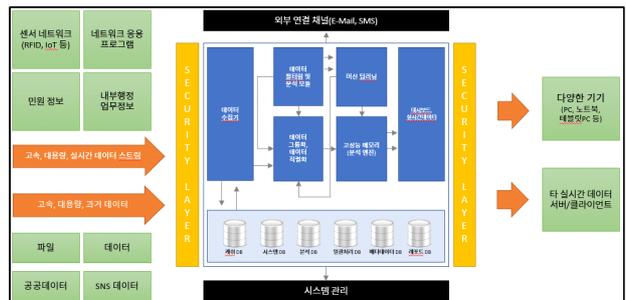
<Figure 1> Agile 방법론을 통한 시스템개발 환경 구축 절차

#### 3.2 시스템 구축에 따른 수행 절차

빅데이터 기반의 공공정보시스템을 구축하여 수요자로부터 다양한 정보를 제공하기 위해 센서 디바이스, 네트워크 응용 프로그램, 각종 민원정보, 내부 업무정보에 따른 데이터가 실시간으로 수집되어야 한다. 파일, 데이터, 공공데이터, SNS 데이터 형식으로 분류되어 수집된 데이터는 정보보호 체계가 갖춰진 시스템으로 전송된다. 운영하고 있는 시스템은 데이터가 안전하게 관리 및 운영되기 위해서 정보보

호의 따른 모니터링이 실시간으로 수행된다. 운영하는 공공정보시스템은 데이터 수집에 따른 데이터 필터링, 분석 모듈, 데이터 그룹화 및 직렬화 등 수집에 필요한 기능이 필수적으로 구축되어야 하고, 빅데이터, AI, 차세대 데이터 현황 관리 모듈과 같은 신기술이 도입되어 수집된 데이터가 안전하게 관리되어야 한다. 또한 운영하고 있는 DB에서는 캐쉬, 시스템, 분석, 일괄처리, 메타데이터, 레포트 등과 같은 기능들이 안전하게 운영되어야 한다. 이후 분석 및 가공된 데이터는 공공정보시스템을 방문하는 수요자로부터 안전하게 통신이 수행되어야 한다. 이 과정에서는 데이터 서버/클라이언트를 통해 실시간으로 데이터가 빠르게 전송되어야 한다. 부가적으로 운영중인 외부연결채널(E-mail, SMS)를 통해 데이터를 개방할 수 있어야 한다.

빅데이터 기반의 신뢰성 있는 정보제공 프로세스에 대한 흐름도는 아래 [그림 2]와 같다. 앞서 언급된 빅데이터 기반의 공공정보시스템의 정보제공 프로세스는 과거 데이터 및 실시간 데이터를 수집하여 분석 템플릿을 적용하여 데이터 분석을 목표하고 있다. 시스템을 활용하는 수요자로부터 제공한 분석결과 리포트 제공 및 데이터 공유가 가능하다. 수요자 단말기의 맞춤형 데이터 제공을 통해 동일한 결과가 조회할 수 있도록 구축하도록 한다. 시스템 구축을 통한 기대효과는 빅데이터 기반 공공정보시스템을 통해 정보를 제공받는 수요자의 정보 제공 활성화를 목표로 한다. 부가적으로 공공데이터 활용을 통해 수요자의 접근성 강화 및 민원소통 활성화를 목표로 한다.



<Figure 2> 빅데이터 기반 정보제공 운영관리 시스템 프로세스 구성도

### 4. 시스템 도출 사례

#### 4.1 빅데이터 기술을 이용한 사용자 맞춤형 정보제공

빅데이터 기반의 공공정보시스템에서 검색포털의 탭에서 검색시스템의 빅데이터라는 키워드를 입력했을 때 사용자 특성 및 관심 분야에 맞는 다양한 키워드를 제공받을 수 있는 사용자 맞춤형 정보제공 시스템을 구축한다. 정보제공 서비스의 도출화면은 아래 [그림 3]과 같다.



<Figure 3> 빅데이터 기술을 이용한 사용자 맞춤형 정보제공 서비스 화면(안)



<Figure 4> 빅데이터 수집을 통한 사업장 사고예측 시스템(안)

## 5. 결론

공공정보시스템을 접속한 사용자 대상으로 빅데이터를 검색 시 정형데이터, 비정형데이터, 데이터 마이닝, 데이터통합, 위치정보, 데이터베이스, 공공데이터 등 빅데이터에 관련한 정보가 나온다. 예를 들어 사용자 1은 데이터 관련으로 검색 시 데이터통합, 마이닝, 데이터베이스, 정형데이터, 비정형데이터에 대한 정보가 나온다. 사용자 2는 데이터의 양을 검색했을 때 내부데이터, 외부데이터, 다양성, 생성속도, 데이터 속성의 크기에 대한 정보가 나온다. 사용자 3은 공공데이터에 대한 값을 입력 시 나오는 데이터는 사고정보, 위치정보, 기상정보, 민원정보, 사고정보 등 사용자별 특징이나 관심 분야별 공공기관 맞춤 정보에 대한 결과값을 제공한다.

### 4.2 빅데이터 수집을 통한 사업장 사고예측

이상징후의 사고 예측에 대한 결과 예측값을 알아보기 위해 공공정보시스템을 접속한 사용자는 작업부주의, 정기교육, 시설관리사항, 위험 사항에 관련한 정보를 제공받을 수 있다. 빅데이터 기반의 사고예측 시스템은 아래 [그림 4]와 같다. 사업장 1, 사업장 2, 사업장3에서 수집된 데이터를 통해서 작업자 부주의, 시설관리, 이전 사고이력, 운송차량 상태, 지역적 이상기후, 정기교육여부, 시설노후화 등과 같은 다양한 정보를 분석한다. 분석된 정보를 기반으로 사고발생 위험이 높은 사업장으로부터 사고예측 위험도와 같은 정량적인 수치와 가공된 사업장 운영정보에 대한 빅데이터 가공 결과값을 제공한다.

본 논문에서는 수요자 중심의 빅데이터 기반의 공공정보시스템 구축에 대하여 연구하였다. Agile 방법론을 통해 빅데이터 기반의 공공정보시스템 개발 절차에 대해서 설계하였다. 그리고 수요자의 다양한 단말기에서 실시간 데이터 수집에 따른 안전한 정보제공 운영관리 시스템 프로세스를 제안하였다.

설계된 시스템을 기반으로 사용자 맞춤형 정보제공 서비스, 사고 예측 관리 서비스에 대한 화면을 구성하여, 사용자, 민원인, 운영자에 대한 정보제공의 효율성을 증대하였다.

향후 세부적인 개선 방향을 연구하여 다양한 단말기를 기반으로 수요자 중심의 안전한 정보서비스를 제공하기 위해 보안성이 향상된 통신시스템 설계 방안에 대해 연구가 수행되어야 한다.

### 참고문헌

- [1] 노현경 외6, 공공 데이터 기반 빅데이터 분석 시스템, 한국인터넷학회논문지, 제 20권 5호, pp. 195-205, 2020. 10.
- [2] 윤상오, 공공데이터 개방정책의 실태분석 및 개선방안에 관한 연구: 공공데이터 포털의 국가중점데이터 개방 사례를 중심으로, 한국공공관리학회, 2019, vol.33, no.1, pp. 219-247
- [3] 김성준, 한국에서 빅데이터를 활용한 범죄예방시스템 구축을 위한 연구, 한국인터넷방송통신학회논문지, 제 17권, 5호, pp. 217-221, 2017. 10
- [4] 오영선, 빅데이터에 대한 금융기관의 기대와 우려, 하나금융경영연구소
- [5] 빅데이터 이해하기, <https://12bme.tistory.com/83>
- [6] 빅데이터 처리 (2)처리과정, <https://ikkison.tistory.com/69>