www.osia.or.kr

OSIA S&TR Journal

ISSN 1738-9887 Vol. 30, No. 2, June 2017

Editorial

02 데이터 사이언스
조완섭 교수/충북대

Article

04 자율주행과 5G, Mobile Edge Computing: 빅데이터 요구 사항과 주요 기술 동향
김진철 매니저/SKT

15 보건의료 빅데이터 연계 플랫폼
이경희 교수/충북대

21 공공빅데이터 개방 표준화
박경현 선임/ETRI

26 빅데이터 거버넌스와 표준화
조완섭 교수/충북대

30th OSIA
(사)개방형컴퓨터통신연구회
Open Standards & Internet Association

데이터 사이언스



충북대학교 교수

조 완 섭

빅데이터는 4차 산업혁명의 원동력이라고 할 만큼 공공과 민간의 전체 산업을 혁신적으로 발전시키는데 핵심 역할을 담당할 것으로 인식되고 있다. 빅데이터의 활용으로 미국 혹은 유럽에서만 분야별로 연간 수천억 달러의 가치를 생성하고, 이에 따라 수많은 전문 인력이 필요할 것이라는 전망으로 인해 빅데이터를 선점하기 위한 각국의 경쟁도 심화되고 있다. 미국이나 영국 등 데이터 선진국이라고 부르는 나라에서는 우리보다 먼저 빅데이터의 가치를 인식하고, 정부가 먼저 공공 빅데이터의 개방 활성화를 추진하고, 빅데이터 활용에 걸림돌을 제거하면서, 이를 통해 자국 산업의 혁명적인 발전을 추구하고 있다. 우리나라도 지난 정부에서 정부 3.0의 일환으로 공공데이터 개방을 적극적으로 추진하고 있지만 아직은 후발주자에 그치고 있어 아쉽다. 전자정부 등에서 1위를 해 왔는데 데이터 부분에서의 경쟁력은 10위권을 넘어 20위 이하로 뒤떨어진 평가 결과가 나오고 있다. 이는 "데이터를 IT (HW, SW)의 일부로 간주하고, IT만 지원하면 데이터는 자동으로 해결된다"라는 잘못된 인식과 데이터 기반의 과학적 의사결정에 익숙하지 못한 우리의 문화적 특성에 기인한 것으로 보인다. 실제 정부 주도로 추진해오고 있는 공공데이터 개방도 수요자의 요구보다는 데이터 보유기관 (공급자) 편의성 위주로 데이터가 공개되고 있어 활용도가 10% 정도에 머물고 있다는 조사 결과도 있다. 이러한 상황에서 각 분야별로 빅데이터 관련 현황과 이슈를 다룸으로써 이 분야의 향후 발전방향을 제시하고자 한다.

첫 번째 논문은 자율주행차 분야에서 빅데이터의 중요성을 강조한 논문이다. 자율주행차는 개별 자동차에서 생성되는 데이터뿐 아니라 주변의 차량과 첨단 도로환경 등에서 끊임없이 생성되는 대규모 데이터를 실시간으로 수집하고 처리해야만 경쟁력을 가질 수 있음을 논의한다.

두 번째 논문은 보건의료 분야의 빅데이터 수집과 연계 및 활용체계 구축 방안을 소개한 논문이다. 보건의료 분야는 가장 복잡하고, 초대규모 데이터를 갖고 있으며, 분석 활용 시 가장 가치가 높은 분야 중 하나로 알려져 있다. 특히 여러 기관에 산재한 개인의 건강 관련 데이터를 환자 중심으로 연계, 통합하여 분석함으로써 높은 가치를 창출할 수 있으므로 관심이 매우 높은 분야이다.

세 번째 논문은 정부 및 공공기관이 보유한 빅데이터의 개방을 위한 플랫폼 및 표준화 방안을 소개한 논문이다. 현재 정부에서 구축하여 운영 중인 공공데이터 포털과 각 지자체 및 공공기관에서 운영 중인 수많은 오픈 데이터 포털들이 표준화 없이 개별적으로 구축되어 운영되고 있어 상호운영성이 없으며, 사용자의 혼란이 가중되고 있는 상황에서, 이에 대한 해결책을 제시하고 있다.

네 번째 논문은 지속 가능한 빅데이터의 활용 가치 증진 방안으로 빅데이터 거버넌스와 표준화 현황을 소개한 논문이다. 빅데이터 거버넌스는 데이터에 대한 품질관리, 표준화, 프라이버시 관리, 메타 데이터 및 마스터 데이터 관리, 데이터 수명관리, 데이터 관련 조직과 인력 등 제도적인 측면에서의 관리체계를 갖추는 것으로, 이를 소홀히 하면 빅데이터가 1회성으로 그치거나 위험을 초래하게 되어 지속적인 성과 창출이 어렵게 됨을 지적하고 있다.

훌륭한 원고를 제출하여 주신 저자분들께 감사를 드리며 빅데이터에 관심을 갖고 계시는 독자분들에게 관련 자료들이 도움이 되기를 기대합니다. 감사합니다.

자율주행과 5G, Mobile Edge Computing

– 빅데이터 요구 사항과 주요 기술 동향 –

김진철, 김동현
SK Telecom 네트워크 기술원

Abstract

자율주행 기술의 신뢰성과 안전성을 높이기 위해 C-V2X (Cellular V2X) 통신을 통한 부가 서비스와 자율주행 인지 범위 확장이 중요하며, 이를 위한 자율주행 빅데이터 처리와 V2X 통신을 이용한 차량 간, 차량과 교통 인프라 및 네트워크 서비스간 공유가 필요하다. 자율주행 자동차의 사례를 바탕으로 V2X 기반의 자율주행 및 고급 운전 지원 시스템(ADAS) 서비스에서 논의되고 있는 빅데이터 요구 사항과 이를 해결하기 위한 5G의 기술 요소에 대해서 논의한다. V2X 통신 인프라로서 5G 모바일 네트워크 기술이 필요한 이유를 저지연 처리 및 빅데이터 대역폭의 요구 사항을 중심으로 살펴본다. 커넥티드 카와 자율주행차를 위한 V2X 통신에서의 저지연 서비스를 위해 5G 기술로서 연구되고 있는 모바일 에지 컴퓨팅(Mobile Edge Computing; 이하 MEC)이 활용될 수 있다. MEC 기술의 표준 제정 및 기술 현황에 대해서 살펴보고, 이러한 MEC 기술이 자율주행 빅데이터 처리에 어떠한 이점을 제공하는지 같이 살펴본다. MEC와 포그 컴퓨팅(Fog Computing) 등으로 여러 수준의 계층으로 조직될 5G 인프라에서 계층 자원 스케줄링(hierarchical scheduling) 기술이 주는 이점과 자율주행 빅데이터와의 연관성에 대해서 논의한다. MEC를 이용한 자율주행 및 커넥티드 카의 ADAS 서비스의 모빌리티 제공을 위해 필수적인 MEC에서의 서비스 핸드오버 기술에 대해서 같이 살펴본다.

I. 서론

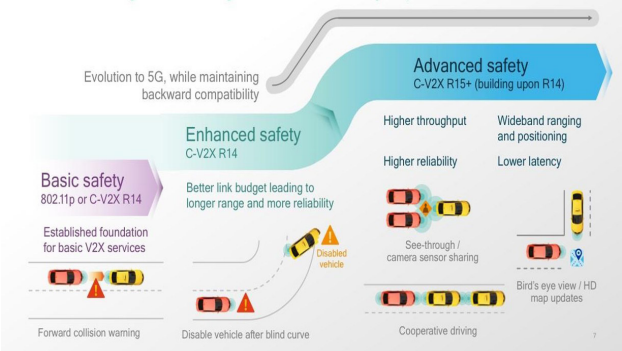
최근 사회 이슈로 급부상하고 있는 자율주행은 자동차 기술의 혁신뿐만 아니라 5G 네트워크 기술과 관련된 다양한 기술적 문제를 해결할 것을 요구하고 있다. 뿐만 아니라, 네트워크를 통해 자율주행차에 제공될 수 있는 HD Map, 화상 전화, 미디어 스트리밍 등의 인포테인먼트 서비스, 차량 내 결제와 같은 부가 서비스들과, 자율주행을 통해 제공될 수 있는 로봇 택시, 카쉐어링 서비스와 같이 다양한 비즈니스 모델의 혁신이 여러 분야에 걸쳐 광범위하게 일어날 것으로 기대되고 있다. 이렇게 자율주행차를 통해 일어나는 모빌리티 혁신은 사회 인프라 및 관련 산업의 에코시스템에도 많은 변화를 끼칠 것으로 전망하고 있다.

이렇게 사회에 큰 영향을 끼칠 자율주행차를 실현하기 위해 아직도 극복해야 할 기술적인 난제들이 많다. 자율주행차의 부상으로 가장 많이 주목을 받게 된 영상 인식 및 기계 학습 기술은 최근 딥러닝 기술이 급격하게 발전하면서 자율주행차의 실현을 앞당기고 있기는 하지만, 자율주행차가 사람과 같은 수준으로 운전하기 위해서는 아직도 신뢰성과 정확도가 많이 부족하다. 이런 인공 지능 기술을 이용한 자율주행 기술은 단순히 인공 지능 기술의 한계로 인한 자율주행의 한계뿐만 아니라, 주행 환경 정보를 수집하는 센서 기술이 사람과 같은 수준으로 정보를 수집, 처리하지 못하기 때문에 현재 기술로써 자율주행차를 사람이 운전하는 것과 같이 만들기는 어렵다.

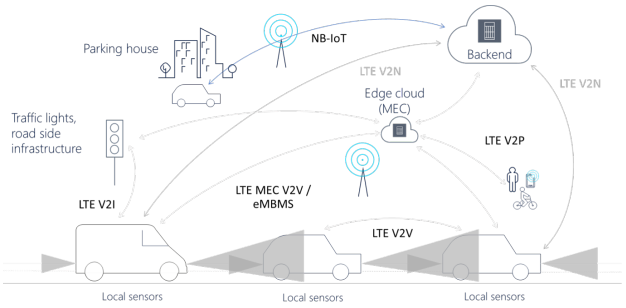
차량 내 자율 주행 제어(in-vehicle autonomous driving control) 기술을 이용한 자율주행 기술만으로 높일 수 없는 자율주행 신뢰성 한계를 V2X 통신과 C-ITS를 이용해 자율주행차의 인지 범위를 확장하여 해결하는 방법이 최근 제안되고 있다. V2X 통신은 커넥티드 카를 위한 서비스 인프라일 뿐만 아니라, 자율주행차가 자체 센서 정보와 차량 내 자율 주행 제어 기술만으로 얻기 어려운 복잡한 주행 정보의 해석과 판단을 중앙의 C-ITS 서비스가 저지연으로 제공할 수 있게 하는 자율주행 서비스 인프라로서 새롭게 조명 받고 있다.

앞으로의 커넥티드 카와 자율주행차 시대에는 자동차의 주행 정보와 센서 정보 등이 모두 V2X 통신을 통해 교통관제센터로 수집되어 분석되고, 분석된 주행 정보들이 고급 주행 보조(ADAS) 및 자율주행 서비스, 인포테인먼트 및 기타 부가가치 서비스를 제공하는 데에 활용될 것이다. 커넥티드 카와 자율주행차 시스템이 소프트웨어 플랫폼을 중심으로 컴퓨팅 시스템화되어 가면서, 개인용 컴퓨터의 운영체제에서 제공하는 원격 소프트웨어 업그레이드와 원격 차량 진단과 같은 서비스도 제공될 것으로 보인다.

자율주행 기술과 차량 내 인포테인먼트 서비스, 전기자동차의 발전과 확산으로 인해 자동차에 탑재되는 컴퓨팅 시스템의 복잡도와 용량은 앞으로 크게 증가될 것으로 보인다. 자동차에서 처리되는 정보가 많아지면서 주행 및 센서 정보, 그리고



〈그림 1. V2X기반 자율주행의 안전성 향상을 위한 5G로의 진화 (그림 출처:[1])〉



〈그림 2. 자율주행에서의 데이터 흐름과 LTE V2X 통신의 활용 개요 (그림 출처:[2])〉

차량 내 컴퓨팅 및 제어 시스템이 생성하는 데이터가 크게 증가할 것으로 보인다. 스마트폰이 확산되면서 모바일 네트워크의 대역폭 요구량이 급증하게 된 것처럼, 플랫폼으로서의 미래 자동차를 통해 생산되는 데이터도 크게 증가할 것으로 보인다. 빅데이터 기술과 서비스에 있어서 스마트폰 이후의 또 하나의 큰 변화가 커넥티드 카와 자율주행차를 중심으로 다가오고 있다.

본 글에서는 자율주행 자동차의 사례를 바탕으로 V2X 통신을 이용한 자율주행 및 고급 운전 지원 시스템(ADAS) 서비스에서 논의되고 있는 빅데이터 요구 사항과 이를 해결하기 위한 5G 네트워크의 기술 요소에 대해서 논한다. V2X 통신 프로토콜로서 5G 모바일 네트워크 기술이 필요한 이유를 저지연 데이터 처리 및 빅데이터 대역폭의 요구 사항을 중심으로 살펴본다. 커넥티드 카와 자율주행차를 위한 저지연 서비스를 위해 5G 기술로서 연구되고 있는 모바일 에지 컴퓨팅(Mobile Edge Computing; 이하 MEC)의 표준 제정 및 기술 현황에 대해서 살펴보고, 이러한 MEC 기술이 자율주행 빅데이터 처리에 어떠한 이점을 제공하는지 같이 살펴본다. MEC와 포그 컴퓨팅(Fog Computing) 등으로 여러 수준의 계층으로 조직될 5G 인프라에서 계층 자원 스케줄링(hierarchical resource scheduling) 기술이 주는 이점과 자율주행 빅데이터와의 연관성에 대해서 같이 논의한다. MEC를 이용한 5G 서비스의 이동성(mobility) 제공을 위해 필수적인 MEC의 서비스 핸드오버 기술의 현황에 대해서 같이 살펴본다.

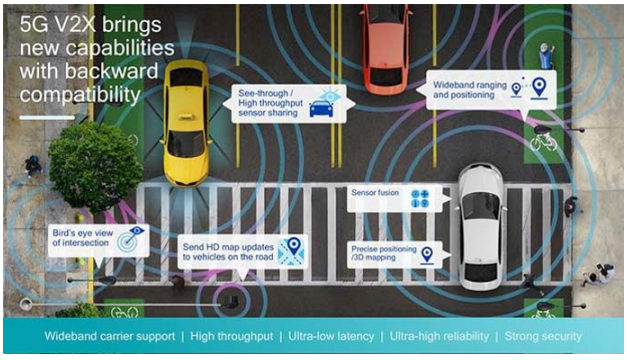
II. 자율주행 관련 서비스를 위한 빅데이터 요구 사항

자율주행차는 우선 높은 신뢰성과 안전성을 보장해야 하지만, 주요 자동차 제조사들이 내놓고 있는 자율주행 관련 기술 및 기능들은 아직 NHTSA 자율주행 수준의 Level 2 정도에 불과하다. 이는 사람의 인지, 판단, 운전 조작 능력을 자율주행 인공지능 기술이 아직 구현하지 못하기 때문이다. 사람 수준의 자율주행이 가능한 자율주행차를 구현할 수 있는 센서계 구성이 아직 업계에서 일정 수준 이상으로 합의되지 않았다. 이러한 센서계에서 수집되는 데이터를 자동차 제어를 위해 10밀리 초에서 100밀리 초의 지연안에 적절하게 연산할 수 있는 차량용 고성능 컴퓨팅 시스템 기술이 아직 성숙되지 않았다. 이런 문제점들로 인해서 앞으로 교통 인프라와 V2X 네트워크를 통한 자율주행 지원 기능이 자동차의 자율주행에 필요할 것으로 보인다.

자율주행과 관련된 빅데이터 이슈는 최근 이런 상황과 관련해서 크게 세 가지 관점에서 살펴볼 수 있다.

① 차량 내부 센서 데이터 및 주행 데이터 처리와 정보 압축

자율주행을 위한 인지, 판단, 제어 연산을 위해서 먼저 센서에서 수집되는 주행 상황에 대한 데이터와 정보들이 제한된 시간 내에 제어 피드백으로 연산되어야 한다. 현재의 32, 또는 64채널 레이더와 같은 센서는 데이터가 크지 않지만, 실제 주행 상황에서는 도로에 많은 숫자의 장애물이나 보행자, 차량 등의 객체들이 등장하여 자율주행에 영향을 줄 수 있기 때문에 이런 주행 상황의 정확한 인지를 위해서는 데이터가 가장 적은 레이더 센서만 해도 앞으로 상용 자율주행차에서 쓰



〈그림 3. 5G V2X의 대표적인 사용 사례. (그림 출처: [1])〉

일 때에는 데이터가 지금보다 훨씬 더 증가한다.

최근 자율주행에서 가장 중요하게 활용되고 있는 센서는 카메라와 같은 영상 센서와 LIDAR이다. 특히 LIDAR 센서는 포인트 클라우드 형태로 저장되는 데이터의 크기가 초당 수백 MB에서 수 GB까지 생성되며, 원시 데이터(raw data)를 그대로 수집하면 시간당 최대 수 TB 정도까지 생성될 수 있다.

영상 센서의 경우에도 32bit의 색상 표현을 가지는 풀 HD급 영상의 경우 영상 원시 데이터의 크기가 카메라 하나의 프레임당 약 수 MB 정도가 된다. 보통 초당 10 ~ 30프레임이기 때문에 초당 카메라 하나가 생성하는 데이터만도 수십 MB가 된다. 보통 자율주행에는 카메라가 여러 대가 쓰이기 때문에 영상 카메라에서만 나오는 데이터도 초당 수백 MB, 시간당 수십에서 수백 GB의 대용량의 데이터가 생성될 수 있다.

이런 센서 빅데이터가 짧게는 10밀리 초에서 100밀리 초, 길게는 수 분 안에 처리되어 제어계에 줄 피드백 연산이 수행되어야 하므로, 자율주행차 내에서도 고성능 컴퓨팅이 수행되어야 한다.

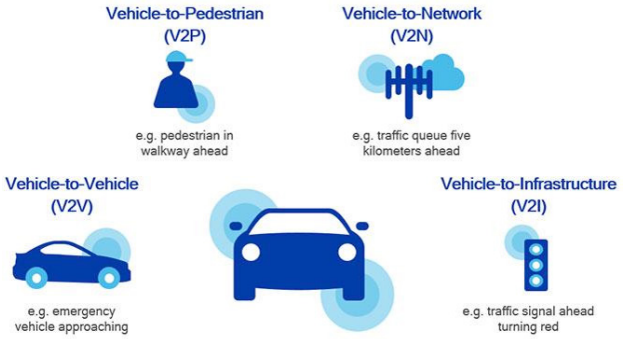
자율주행차 내부에서의 빅데이터 처리는 크게 두 가지 목적이 있다. 첫 번째는 앞에서 언급한 바와 같이 자율주행 제어를 적절한 시간 안에 안전하게 처리하기 위한 제어 연산용이고, 두 번째는 자율주행차의 주행 정보를 V2X 통신을 통해 차량 간에 주고받거나 ITS(Intelligent Transportation System) 또는 C-ITS(Cooperative Intelligent Transportation System)와 같은 교통 인프라로의 주행 데이터 전송을 위한 데이터 압축이다. 센서 데이터를 어떤 방식으로 처리하여 주행 데이터로 변환, 압축할지에 대해서는 센서계 구성 및 데이터 형식 표준 논의와 함께 아직도 많은 연구가 더 진행되어야 한다.

② 교통 인프라에서 수집되는 교통 정보 빅데이터의 수집과 처리

자율주행차에 장착된 센서뿐만 아니라, 교통 인프라에 설치되어 운영되는 센서 데이터 정보들도 V2X나 ITS 네트워크를 통해 전송되어 수집, 처리되어야 한다. 교통 인프라 데이터는 단위 센서당 데이터의 양도 문제가 되지만, 다양한 인프라에 쓰이는 다양한 종류의 센서 데이터를 처리하는 원시 데이터의 다양성(variety) 문제로 해결해야 한다.

이동 중인 많은 차량의 데이터를 연속적으로 수집하여 중앙의 ITS 서비스 인프라에 전송하거나, 에지 컴퓨팅(edge computing) 디바이스 등을 통해 짧은 시간 안에 자율주행차의 요청에 회신을 줄 수 있는 연산을 하여야 하기 때문에 빅데이터의 세 가지 요건 중 데이터 생성의 속도(velocity)와 다양성(variety)에 의한 요구 사항이 강하다.

교통 인프라에 쓰이는 센서들 중에서 단위 센서가 많은 데이터를 생성하는 센서는 아직까지는 CCTV 등의 영상 카메라



〈그림 4. 자율주행 V2X 통신의 네 가지 종류. (그림 출처: [3])〉

센서뿐이다. 교통 신호등 데이터나 교통 인프라 센서들의 데이터는 단위 데이터의 양은 수 바이트에서 수 킬로바이트 정도로 그렇게 크지 않다. 하지만, 이런 작은 데이터라도 많은 수의 센서가 연속적으로 하루 24시간 동안 계속해서 데이터를 수집하기 때문에 하루에 수집되는 데이터의 양은 수 MB에서 수 TB에 이를 수 있다.

자율주행과 커넥티드 카가 점차 대중화되면서 안전 문제, 보험 데이터 수집 등의 사회 이슈 해결을 위해 점점 더 많은 양과 종류의 센서가 교통 인프라에 설치되어 데이터를 수집할 것으로 전망된다. 이런 많은 양의 센서가 연속적으로 수집하는 데이터 스트림이 끊김 없이 수집되도록 하는 네트워크 인프라 아키텍처와 지능형 네트워크 운영 시스템 문제를 해결하는 것도 앞으로의 자율주행을 위한 교통 인프라의 중요한 요구 사항 중 하나이다.

③ 차량 안전 및 신뢰성 향상을 위한 차량 센서 데이터 및 주행 데이터 전송 및 공유

자율주행차의 안전성을 높이기 위해서 현재 논의되고 있는 V2X 통신의 사례(use case)는 크게 세 가지 관점에서 논의되고 있다. 첫 번째로, 자율주행차 간의 통신을 통해 하나의 자율주행차가 인지하지 못하는 주행 상황을 차량 간에 서로 공유, 전달하여 자율주행차의 인지 범위를 넓히고 안전성을 높이는 V2V(Vehicle-to-Vehicle) 통신의 사례(use case)가 있다. 두 번째로 자율주행차와 교통 인프라 간의 통신을 통해 중거리, 원거리에서 일어난 교통사고나 홍수로 인한 도로 유실, 끊김 등의 재난 상황을 사전에 감지하고 자율주행차의 경로를 수정하거나 주행 제어를 변경하는 등의 V2I(Vehicle-to-Infrastructure) 통신의 사용 사례가 있다. 경우에 따라서 데이터 전송 및 네트워크 지연 제약을 해결하기 위한 네트워크 인프라와의 V2N(Vehicle-to-Network) 통신의 사용 사례를 별도로 나누기도 하지만, 본 글에서는 V2N은 V2I의 한 종류로 보고 특별하게 구분하지는 않을 것이다. 세 번째로, 보행자나 사람들이 휴대하고 있는 스마트폰이나 웨어러블 디바이스와의

통신을 통해 자율주행차 호출, 자율주행차를 이용한 주문형 운송서비스, 자율주행차 내 탑승자와 보행자 간 통신에 활용되는 V2P(Vehicle-to-Pedestrian) 사용 사례가 있다.

위 세 경우 중에서 빅데이터 요구 사항이 충족되어야 하는 경우는 V2V, V2I 통신의 경우이다. 우선 V2I 통신의 경우를 생각해보면, 자율주행차 내부에서 센서 데이터 및 주행 데이터를 실시간으로 처리, 압축하여 ITS 및 교통 인프라로 송신하기 때문에 단위 차량에서 생성, 송신하는 데이터는 많지 않을 수 있다. 하지만, 고속도로와 같이 교통량이 많은 곳에서 5G 망의 각 기지국이 빠르게 지나가는 차량별로 수신을 받는 데이터의 양은 교통량에 따라 역시 하루에 수십 GB에서 수 TB에 이를 수 있다. V2V 통신의 경우에는 저 지연성 문제가 더 중요한 문제가 된다. 역시 자율주행차 내부에서 원시 데이터를 한번 가공, 압축하여 전송하면 데이터의 전체 양은 크지 않을 수 있지만, 차량이 빠르게 서로 지나가거나 주행하는 경우 차량 간에 데이터를 주고받을 수 있는 시간은 길지 않기 때문에, 단위 시간당 주고받아야 하는 데이터의 양은 저 지연성의 제약 조건과 함께 훨씬 더 엄격한 조건을 만족해야 한다.

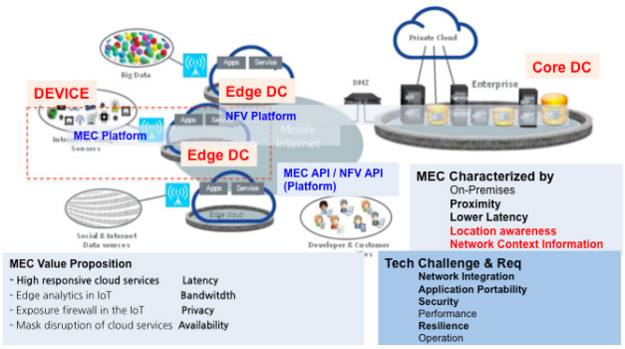
위의 세 경우에서 살펴본 것과 같이, 자율주행차가 사회의 안전한 교통 인프라로서 자리 잡기 위해서는 자율주행차 기술만 발전되어야 하는 것이 아니라, 자율주행차의 안전성을 높일 수 있는 V2X 통신 기술과 교통 인프라 기술이 같이 발전되어야 한다. 자율주행차의 이동성(mobility)에 다른 저 지연성 제약이 안전에 큰 영향을 미치기 때문에, 10배 이상 크게 늘어나게 될 5G의 대역폭뿐만 아니라 5G의 저 지연성이 자율주행을 위한 V2X 통신에 더 적합할 수 있다.

자율주행의 저 지연성 제약 문제를 해결하기 위해서 5G의 모바일 에지 컴퓨팅(Mobile Edge Computing; 이하 MEC)의 중요성이 점점 더 높아지고 있다. MEC는 차량의 V2X 단말기가 5G 망에 접근하는 기지국에 설치되는 컴퓨팅 장비에서의 데이터 처리(에지 컴퓨팅; edge computing)를 위한 5G 네트워크 인프라 기술이다. 5G 망의 코어 및 데이터 센터에서 제공되는 서비스를 거칠 경우에 만족시킬 수 없는 저지연 서비스를 위해, 컴퓨팅 자원의 일부를 엑세스 네트워크 영역, 또는 에지 네트워크 영역에 두고, 코어 네트워크 및 데이터 센터까지 데이터가 전송되지 않고도 MEC에서 데이터를 처리하여 사용자에게 필요한 지연시간 내에 서비스를 제공하는 것이 MEC이다. 자율주행차 및 커넥티드 카에서의 저 지연성 제공을 위한 MEC 기술의 요구 사항과 필요 기술을 앞으로 살펴보고자 한다.

III. 5G 및 Mobile Edge Computing 개요와 주요 기술 동향

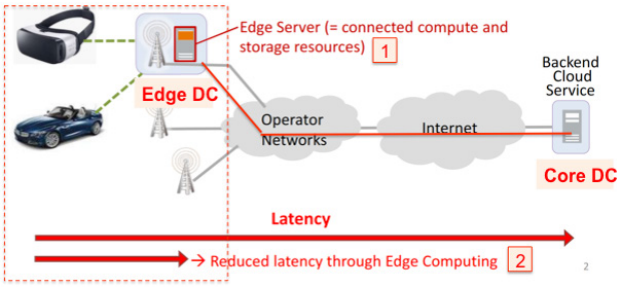
앞에서 논한 것과 같이 V2X 기반 ADAS와 자율주행 서비스에서 짧은 지연(Latency) 특성과 안정적인 연결성(Connectivity) 특성이 필요하게 된다. 이런 특성을 위해 5G는 저 지연성(Low Latency), 높은 대역폭(High Bandwidth), 대규모 연결성(Massive Connectivity)의 특징을 만족하는 네트워크를 제공하며, 개인별, 서비스별로 이 네트워크의 특성이 제공될 수 있게, 서비스 지향 아키텍처(service-oriented architecture)를 가지고 있다.

서비스 테넌트 별로 할당될 수 있는 네트워크 슬라이스(network slice)는 5G 코어 네트워크의 대표적 특성이기도 하다. 또 서비스 지향 모델로써 에지 네트워크(edge network)에 다중 접근(multi-access)의 리소스와 서비스 API를 필요로 하며, 이를 위한 기술이 모바일 에지 컴퓨팅(Mobile Edge Computing / Multi Access Edge Computing; 이하 MEC)이다. 앞서 말한 5G의 세 가지 네트워크 기술적 특성과 함께, 서비스 제공을 위한 프레임워크를 제공한다.

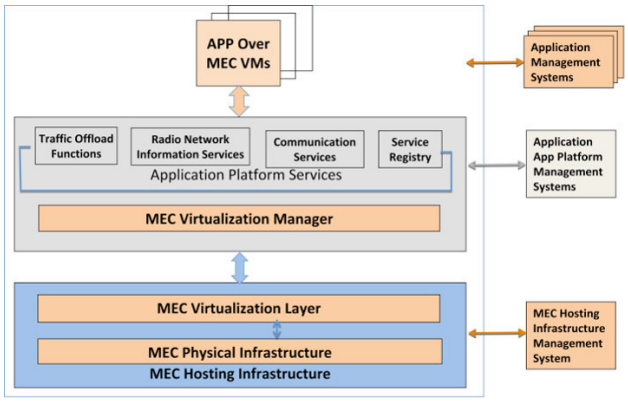


〈그림 5. Edge DC 구조상의 MEC 특징 및 가치 (그림 출처: [16])〉

MEC는 모바일 네트워크 사용자들과 가까운 거리에 있는 RAN (Radio Access Network)과 엑세스 포인트(Access Point)를 통한 접속 등의 다중 접속(multi-access) 상황에서



〈그림 6. Edge Computing 의 제공 리소스와 Latency 특성〉



〈그림 7. MEC Platform의 구조 (그림 출처: [17])〉

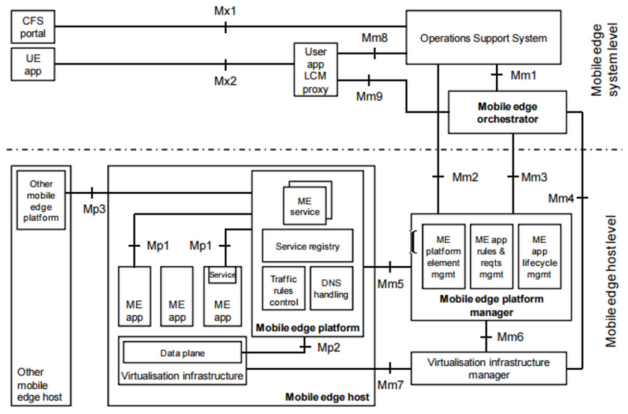
IT 및 클라우드 컴퓨팅을 위한 기능을 제공한다.

기지국 또는 무선 네트워크 컨트롤러에 위치하여 응용 프로그램 및 서비스가 이용하여 컨텍스트 관련 서비스를 제공할 수 있는 가입자 위치 또는 셀로드와 같은 실시간 라디오 및 네트워크 정보에 대한 액세스를 제공한다.

5G 네트워크에서 MEC는 애플리케이션 개발자 및 컨텐츠 제공 업체에게 무선 엑세스 네트워크(RAN) 에지는 근접, 초저 대기 시간, 고 대역폭뿐만 아니라 무선 네트워크 정보 및 위치 인식에 대한 실시간 액세스를 위한 서비스 환경을 제공한다. MEC를 사용하면 컨텐츠, 서비스 및 애플리케이션을 가속하여 다양한 모바일 네트워크 조건에서 고객의 경험을 유지할 수 있다.

MEC는 ETSI 표준의 NFV 환경에 모바일 에지 플랫폼(Mobile Edge Platform)과 에지(Edge)의 컴퓨팅 환경으로 확장한 것이다. 에지 컴퓨팅(Edge Computing)에 가상화 기술을 더해 멀티 테넌시를 지원하는 표준화된 개방형 환경이다. 타사 응용 프로그램을 여러 공급 업체 플랫폼에 효율적이고 원활하게 통합함으로써 이동 통신 사업자가 더 폭넓은 고객을 지원할 수 있다. MEC 산업 표준 그룹(Mobile Edge Computing Industrial Standard Group; MEC ISG)은 업계 전반의 응용 프로그램 개발을 활성화하고 가속화하여 시장 규모를 키우는데 영향력을 끼칠 것으로 예상된다.

모바일 에지 컴퓨팅 산업 규격 그룹 (MEC ISG)의 회의에서 1단계(Phase1)의 5가지 사양에 대한 상세한 표준 합의 작업이 2015년 중순 완료되었다. 전문 용어, 서비스 시나리오, 기술 요구 사항 (사용 사례 및 이점 포함), 프레임 워크 및 참조 아키텍처로 이어지는 개념 프레임워크의 검증을 수행하였다. 이후 플랫폼 서비스, API 및 인터페이스에 대한 내용이 추가되었다. 현재 다중 접근(multi-access) 특징을 포함하여 2단계(Phase 2) 개정 작업이 진행 중이다.



〈그림 8.ETSI Mobile Edge System Reference Architecture (그림 출처: [17])〉

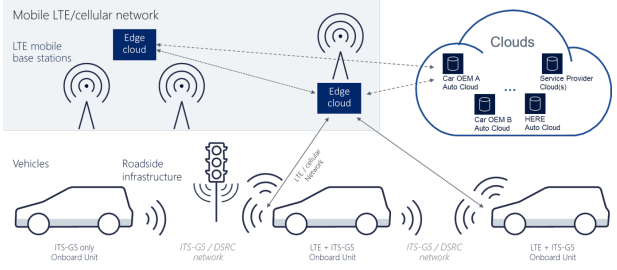
IV. 자율주행 빅데이터 처리를 위한 5G와 Mobile Edge Computing

A. 자율주행 빅데이터 처리를 위한 5G 요구 사항

자율주행 빅데이터 처리 관점에서 5G 인프라는 대량의 데이터를 짧은 시간에 전송할 수 있는 LTE 대비 10배 이상의 대역폭을 제공하고, 서비스 요청부터 응답까지 제약된 시간 안에 서비스를 보장하는 실시간 저 지연성을 보장할 수 있는 요구사항을 만족하여야 한다. V2X를 이용한 자율주행을 지원하기 위해 5G 모바일 네트워크가 만족해야 할 요구 사항에 대해서 대역폭 관점과 저 지연성 관점에서 같이 살펴보도록 한다.

① 자율주행 지원을 위한 5G V2X통신의 대역폭 요구 사항

3GPP에서 V2X 표준을 논의하고 있는 SA1 워킹그룹의 기술 문서 TR 22.886과 TS 22. 186에 따르면, 동시 검출된 10개의 객체에 대한 정보를 전송하기 위해 전 처리된 데이터에 대해 50Mbps, 원시 데이터에 대해서는 1Gbps 이상의 대역폭이 1600바이트의 페이로드 데이터와 함께 전송되어야 할 것으로 예상된다[4-6]. 영상 정보 공유(VaD; Video data sharing for assisted and improved automated driving) 사용 사례의 경우, 군집 주행(platooning)의 성능 요구 사항에



〈그림 9. LTE를 이용한 V2X 통신의 개요 (그림 출처: [1])〉

대해 HD 급, 720p, 초당 30 프레임의 영상 데이터 공유를 위해 최소한 10Mbps의 대역폭이 필요한 것으로 예상된다. 자율주행과 고급 운전자 주행 보조(ADAS) 성능 요구 사항을 기준으로 컴퓨터 비전을 이용한 원시 영상 데이터 공유를 위해서 최소 100Mbps ~ 700Mbps의 대역폭이 필요할 것으로 예상된다[5].

앞의 2장에서도 논의한 바 있지만, 레이더의 데이터가 영상 센서나 LIDAR보다 적은 데이터를 생성한다고 해도, 동시 검출하는 객체수가 많아지면 이에 비례해서 데이터가 많아진다. 위의 동시 검출 객체 수가 늘어나면 당연히 필요한 대역폭도 늘어나게 된다. 동시 검출이 필요한 객체 수가 상대적으로 적은 시외 도로인 경우와, 동시 검출이 필요한 객체 수가 상대적으로 많은 도심 도로 환경에서의 대역폭 요구 사항은 많이 다를 수밖에 없다.

5G의 자율주행 V2X를 위한 대역폭 요구 사항은 자율주행차의 컴퓨팅 성능에 의해서도 많은 제약을 받는다. 만약 자율주행차가 동시에 많은 객체를 검출, 처리하여 데이터를 크게 압축할 수 있다면 대역폭 요구 사항은 줄어들 것이다. 자율주행 데이터 전송 및 처리를 위한 요구사항에는 자율주행차에서 최소한으로 처리하여야 하는 센서와 주행 상황 정보 데이터의 종류, 처리와 표현 방법에 대한 요구 사항도 같이 고려가 되어야 한다. 이러한 주행 상황, 주행 환경 인지 정보전 처리(pre-processing) 수준과 표현 방법에 따라 대역폭 요구 사항이 달라지기 때문이다.

5G V2X에서 대역폭 요구 사항에 영향을 줄 수 있는 또 하나의 요소는 교통량이다. 교통량은 자율주행차 센서계가 동시 검출하는 객체 수와 밀접한 연관이 있다. 교통량이 많은 도로에서는 당연히 자율주행차가 동시 검출해야 하는 차량의 수도 많아질 수밖에 없다. 교통량은 자율주행차의 동시 검출 객체 수와도 연관이 있지만, V2X 기지국에 동시 접속하여 데이터를 전송하는 차량의 수와 접속 빈도, 서비스 핸드오버 빈도가 높아지는 것 때문에 통신망 대역폭 점유율이 높아지는 문제에 더 큰 연관이 있다. 교통량이 많아 동시 접속하는 차량의 수와 빈도가 높아지면 V2X 대역폭 자원이 상대적으로 희소해져 네트워크 품질과 자율주행의 안전성에 영향을 미치게 된다. 교통량이 많아지면 대량의 차가 이동하면서 발생하는 기지국 MEC 자원에서의 서비스(또는 어플리케이션) 핸드오버도 자주 일어나기 때문에 이런 핸드오버에 필요한 대역폭 점유 문제까지 더해져 대역폭 확보와 네트워크 품질에 많은 영향을 주게 된다.

위에서 살펴 본 세 요소, 자율주행 차량에서 동시 검출, 인지하는 객체 수, 자율주행차 컴퓨팅 시스템의 주행 인지 정보 처리 성능과 표현 방법, 그리고 교통량이 자율주행을 위한 V2X 대역폭 요구 사항에 많은 영향을 준다. 이들 세 요소가 서로 상호 연관되어 있어 5G 대역폭을 V2X 요구 사항을 수용할 수 있도록 예측하고 정하는 것이 쉽지 않아 자율주행 V2X 표준을 만드는데 어려움이 예상된다.

② 자율주행 지원을 위한 5G V2X통신의 저 지연성 요구 사항

저 지연성을 위해서는 5G 네트워크 및 컴퓨팅 인프라의 계층성과 이를 이용한 저 지연성 지원이 해결되어야 한다. 5G 네트워크 인프라의 계층성이 저 지연성과 어떻게 연관이 있는지 살펴보자.

V2X를 이용한 자율주행에서 저 지연성 요구 사항은 대개 주행 안전과 관련된다. 예를 들면, V2V 통신을 이용해 앞 차가 본 사고 현장 정보를 뒤 차가 공유 받아 속도를 감속하거나(see-through), 협력 차선 변경 등의 요구 사항을 만족시키기 위해 필요한 지연은 20ms 미만이다. 자율주행을 위한 차량 간 정보 공유(information sharing for automated driving)를 위해서는 10ms 이하의 지연이 요구된다. 비상 상황에서 차량 경로 변경 및 정렬(emergency trajectory alignment)에 대한 요구 사항에서는 3ms 이하의 지연이 요구된다[5]. 현 LTE 네트워크의 실측 종단 간 지연(end-to-end latency)이 50ms ~ 150ms 정도인 것으로 생각하면 자율주행과 관련된 저 지연 요구 사항은 LTE만으로는 만족이 어렵다.

5G 망에서 어떻게 1ms 이하의 저 지연성을 지원할 것인지에 대해서는 아직도 많은 기술적인 논의가 이루어지고 있다. 저 지연성 지원의 가장 유력한 방법 중의 하나는 바로 MEC를 이용하는 것이다. 저 지연성이 요구되는 일부 서비스를 기지국과 같은 무선망 전방, 즉 에지 컴퓨팅 인프라에 위치시켜 코어 및 데이터 센터까지 데이터가 전송되지 않고도 에지 컴퓨팅 인프라에서 응답이 될 수 있도록 하여 지연을 줄이려는 것이다.

MEC를 사용해서 저 지연성을 보장하기 위해 해결해야 할 기술적인 난제들이 있다. 첫 번째로, 5G 인프라 전 지역에 걸쳐 배치된 컴퓨팅 인프라에 저 지연성 보장을 위한 서비스 어플리케이션 인스턴스를 스케줄링하는 문제이다. 기존의 3G/4G 네트워크 인프라와는 달리, 엑세스 네트워크와 에지 네트워크를 위한 기지국이 이제는 단순한 네트워크 인프라가 아닌 컴퓨팅 인프라가 같이 구축되어 서비스되어야 하는 작은 데이터 센터가 된다. 이 경우 네트워크 인프라뿐만 아니라 컴퓨팅 인프라도 네트워크가 구축된 전 지역에 걸쳐 분산되는 구조를 가지게 된다. 네트워크 서비스의 종단 간 지연 요구 사항이 만족되도록 서비스 인스턴스를 분산된 데이터 센터 형태로 조직된 컴퓨팅 인프라 사이에 어떻게 스케줄링하고 배치할 것인지가 어려운 기술적인 문제가 된다.

두 번째로, 자율주행차의 빠른 이동성으로 인해 수시로 바뀌는 자율주행차의 엑세스 네트워크와 기지국 사이에 어떻게

일관성 있게 서비스 핸드오버를 신속하게 할 수 있을지에 대한 문제이다. 과거 3G/4G에서는 엑세스 네트워크의 주파수 대역 자원을 사용자의 디바이스에 할당하고 모바일 네트워크에 대한 연결성만 제공하면 되었지만, 5G에서 MEC를 통해 제공되는 서비스는 일종의 작은 서비스 서버들이기 때문에 할당해야 하는 자원도 컴퓨팅 자원을 포함하여 복잡해지고, 핸드오버 대상도 복잡해진다. 이러한 MEC에서의 서비스 핸드오버 문제는 최근 MEC를 에지 클라우드로서 클라우드화하고, 각 자율주행 차별로 에지 클라우드에서의 멀티테넌시를 지원하며, 서비스 서버를 에지 클라우드 서비스 인스턴스로 보고 라이브 마이그레이션의 형식으로 서비스 핸드오버를 보는 관점으로 바뀌어 가고 있다.

위의 자율주행 V2X의 저 지연성 지원을 위해 5G 망에서 해결해야 할 문제로서 컴퓨팅 자원의 계층 스케줄링 문제와 서비스 인스턴스의 핸드오버 문제를 다음 절에서 좀 더 구체적으로 살펴본다.

B. 자율주행 지원 서비스의 워크로드 분산 및 저 지연 처리를 위한 계층적자원 스케줄링 (hierarchical scheduling)

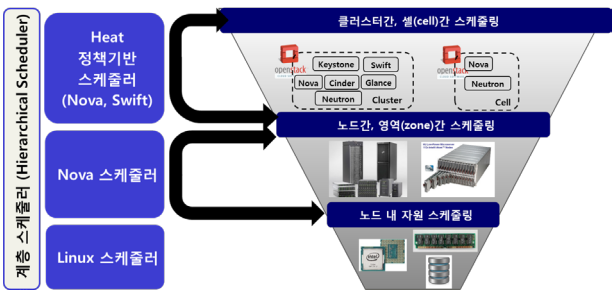
MEC의 핵심은 기존에 코어 네트워크가 위치한 데이터 센터에서 제공되던 서비스를 엑세스 네트워크, 즉 네트워크의 에지 부분에 위치시켜 서비스 요청의 종단 간 지연을 엑세스 네트워크에서의 접속 및 서비스 응답시간만으로 단순화하여 저 지연성을 확보하는 것이다. 이렇게 엑세스 네트워크에서 서비스가 이루어지기 위해서는 기지국에서 서비스를 위한 컴퓨팅 자원이 같이 운영되고 스케줄링 되어야 한다. 5G 이전의 기지국에는 없던 서비스를 위한 IT 장비가 같이 운영되어야 하기 때문에 기지국이 작은 데이터 센터와 같은 설비로 복잡해지게 되며, 데이터 센터에 중앙 집중화되어 관리되던 서비스 IT 인프라들이 네트워크 인프라가 구축된 전 지역에 걸쳐 분산되어 운영, 관리되면서 네트워크 자원과 컴퓨팅 장비 위주의 IT 장비가 계층(hierarchy)을 이루어 조직, 통합된다.

이렇게 저 지연성을 포함한 서비스의 제약 조건과 응답 시간별로 서비스를 제공하는 서버 어플리케이션이 MEC를 제공하는 에지 클라우드에 위치해야 하는지, 또는 코어 네트워크 서비스가 제공되는 중앙 데이터 센터에 위치해야 하는지를 판단해서 서버 어플리케이션의 위치와 필요한 자원을 할당, 관리해주어야 하는데 이를 계층 스케줄링(hierarchical scheduling)이라고 부른다.

원래 계층 스케줄링은 클라우드 컴퓨팅에서 시작되었다. 데이터 센터의 자원들, 즉 서버, 스토리지, 네트워크 스위치 등의 IT 자원들은 물리적인 제약으로 인해 계층을 이룰 수밖에 없다. 데이터 센터 내 하드웨어 자원들의 경우 CPU, 메모리, 스토리지, 네트워크 대역폭 등의 자원이 우선 서버 노드(node) 별로 묶여지고, 다시 노드 단위로 묶여진 자원들은 네트워크 스위치를 통해 연결되어 랙(rack) 단위로 묶여지게 된다. 이렇게 랙 단위로 묶여진 자원들은 용도에 따라 여러 대의 랙 자원들이 하나의 목적을 위해 통합, 조율(coordination) 되어 동작하는 클러스터, 또는 분산 컴퓨팅 시스템으로서 조직되며, 이런 다양한 분산 컴퓨팅 시스템이 모여 데이터 센터 스케일의 자원을 구성하게 된다. 최근의 서비스들은 빅데이터 기반의 분산 컴퓨팅 어플리케이션들이 많아 서비스 운영을 위해 필요한 자원이 쉽게 데이터 센터 스케일의 자원으로 확대된다. 5G 네트워크에서 다루게 될 빅데이터의 수준은 이런 데이터 센터 스케일의 자원이 지리적으로 넓은 지역에 분포하여 통합되어야 하는 분산 데이터 센터 스케일의 자원을 사용하여야 하는 수준으로 발전하고 있다.

클라우드 컴퓨팅 기술의 대표적인 사례로 OpenStack을 들 수 있다. 사용자의 요구 사항에 맞는 컴퓨팅 노드에 가상 머신을 할당하고 운영하기 위해 OpenStack에서 계층 스케줄러 개념을 도입한 자원 스케줄러를 사용하고 있다. OpenStack 자원도 클라우드 컴퓨팅의 특성상 데이터 센터 스케일로 확장이 되며, OpenStack 소프트웨어와 서비스가 설치되고 운영되는 IT 장비들은 데이터 센터 스케일의 자원 계층을 그대로 이루게 된다. 이런 계층 구조의 자원에서 사용자의 CPU, 메모리, 저장 장치와 네트워크 대역 및 지연 요구 사항에 맞는 노드에 가상 머신과 컨테이너 인스턴스를 할당하기 위해 OpenStack의 가상 머신 관리 소프트웨어인 Nova의 컴퓨트 노드에 위치한 Nova 스케줄러가 우선 사용자의 요구 사항에 맞는 노드를 찾아 가상 머신을 할당한다. 할당된 가상 머신은 다시 노드의 Linux 커널 스케줄러와 KVM 등의 가상화 기술을 통해 CPU, 메모리, 가상 디스크를 위한 저장 장치 등의 자원을 할당받아 인스턴스화되고 구동되게 된다.

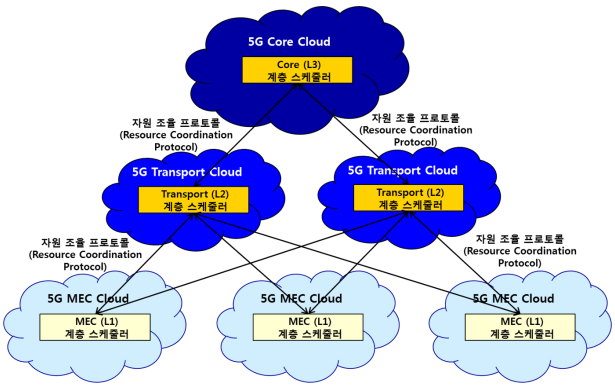
MEC에서의 계층 스케줄링은 OpenStack의 2계층 스케줄링보다 한층 더 복잡할 수 있다. 우선 하나의 데이터 센터 내에서의 자원 스케줄링이 아닌, 분산된 데이터 센터 간의 자원



〈그림 10. OpenStack 클라우드에서의 가상 머신과 컨테이너 인스턴스 자원 할당을 위한 계층 스케줄링 (그림 출처: [15])〉

스케줄링 문제를 고려해야 할 필요가 있다. 작은 데이터 센터로서의 기지국은 그 자체로 보안을 위한 독립된 네트워크 망 구조를 가지게 될 것이며, 기지국 내의 IT 자원 또한 이런 독립된 네트워크 내에서 운영되기 때문에 자원 스케줄링을 위한 정보를 주고받으면서 자원을 할당할 서버 어플리케이션 인스턴스가 프로비저닝되어 실행되기 위해 필요한 자원 통합이 어렵다. 서버 어플리케이션 인스턴스가 중앙의 데이터 센서에서 관리되고 배포되는 것이 아니라, 이런 보안 문제와 저 지연성 문제 때문에 각 기지국별로 보관되고 관리된다면 일관된 서버 어플리케이션 이미지 관리가 매우 어려워져 5G V2X 서비스 품질에 영향을 미치게 된다.

엑세스 망에서 네트워크 접속 및 대역폭 할당 관리만 하면 되었던 것에 비해, MEC를 이용하는 5G 망에서는 엑세스 네트워크 수준부터 자원 할당 문제가 더 복잡해진다. 엑세스 네트워크에 접속한 사용자의 자율주행차가 요구하는 서비스를 제공하기 위한 서버 어플리케이션 인스턴스가 저지연 제약 시간 내에 할당되어 서비스 응답까지 제공해야 하기 때문에 자원 스케줄링의 복잡성과 신뢰성이 보통 클라우드 컴퓨팅에서의 자원 스케줄링보다 매우 높고 어렵다. 자율주행차의 이동성(mobility)으로 서버 어플리케이션에서 가지고 있는 각 자율주행차의 테넌트 정보가 다음 기지국의 서비스 어플리케이션 인스턴스에 신속하고 빠르게 전달되어 끊김 없는 서비스 핸드오버가 이루어져야 하기 때문에 서버 어플리케이션 인스턴스가 매우 가볍거나 핸드오버만을 위한 별도의 메시지 통신 규격이 표준화되고 구현되어야 하는 문제도 있다. 신속한 서버 어플리케이션 인스턴스의 프로비저닝과 핸드오버를 위해서는 컨테이너 기술 같은 경량 가상화 기술이나 마이크로 서비스 아키텍처와 같은 경량 서비스 지향 아키텍처를 사용할 수 있으나, 이런 최신 기술과 아키텍처가 5G V2X 표준화를 위한 MEC 아키텍처로 어떻게 구현되어야 할지에 대해서는 아직 더 연구되어야 한다.



〈그림 11. 5G 인프라에서의 자원 스케줄링을 위한 계층 스케줄러의 개념. 각 계층 스케줄러는 해당 레이어의 자원만을 스케줄링하며, 레이어간, 또는 같은 레이어의 다른 클라우드간 자원 조율과 스케줄링은 자원 조율 프로토콜(resource coordination protocol)을 통해서 자원의 할당과 사용을 계층 스케줄러간 협상한 다음 자원이 할당, 운영되게 된다.〉

MEC 때문에 엑세스 네트워크에 운영이 되는 IT 장비들로 인해 보안 문제도 어려워진다. 기지국별로 보안이 관리되어야 하기 때문에, 분산된 기지국과 중앙 데이터 센터 간의 일관성 있는 보안 정책 적용과 관리도 한층 더 복잡해진다. 기지국에 있는 에지 클라우드들은 클라우드 컴퓨팅의 보안 문제들을 같이 가지고 있기 때문에, 클라우드 컴퓨팅의 보안 문제를 해결할 수 있는 보안 아키텍처 디자인이 같이 적용되어야 한다. 자율주행 V2X의 높은 저지연 성능 요구 사항과 보안 요구 사항을 만족해야 하기 때문에, 5G MEC에서는 기존의 클라우드 컴퓨팅 기술보다 신뢰성과 보안 측면에서 한층 더 강화된 기술이 쓰여야 할 것으로 보인다. 이런 보안 요구 사항을 만족하면서 분산된 에지 클라우드 간 자원 할당과 관리를 할 수 있는 계층 스케줄러가 어떤 식으로 디자인되어 구현되어야 할지 아직 연구되지 않았다. 5G에서 저지연 요구 사항을 위한 MEC 표준을 위해 더 연구되어야 할 문제이다.

C. 자율주행 서비스의 저 지연성 확보를 위한 서비스 핸드오버(application handover) 기술

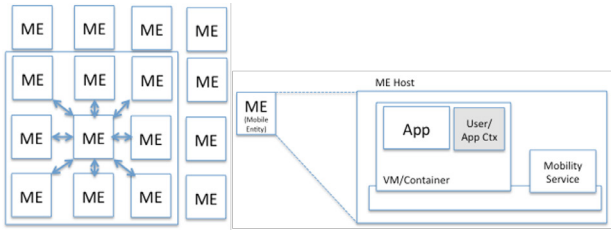
자율주행 서비스의 저 지연성 확보를 위해 MEC 기술을 사용하는 경우, 서비스의 상태를 빠르게 복구하는 서비스 복구(application synthesis)와 이를 기반으로 하는 서비스 핸드오버(application handover) 기술은 핵심기술이 된다.

서비스 핸드오버(application handover)는 기존의 주파수 대역 핸드오버(radio handover)를 기초로 각 에지(Edge) 노드에 배치된 서비스 이동성(application mobility)에 대한 기술이다. 자율주행 서비스에서 저 지연성 확보를 위해 MEC를 이용해 차량과 근접 한 위치에 서비스를 둘 경우, 움직이게 되는 주행 서비스의 특성상 접속하는 기지국의 이동이 일어나지만, 모든 셀 이동마다 서비스 핸드오버(application handover)가 일어 날 필요는 없다. 따라서 서비스의 상태를 보고 이동을 관리하는 기술이 필요하다. MEC에서의 서비스 이동성(application mobility)에 대한 부분은 아직도 합의가 이루어지지 않아 연구가 진행 중이다.

MEC에서의 서비스 핸드오버와 이동성을 위해 MEC의 모바일 엔티티 호스트(mobile entity host; ME host) 내에 모빌리티(Mobility) 서비스와 모빌리티 그룹(Mobility Group)이 필요하며, 이를 서비스 핸드오버에 적용한다. 노드(node) 내부의 저지연을 위해 서비스 기능의 일부가 전용 하드웨어에 오프로드 되어 가속되면 서비스의 지연을 더 개선할 수 있다.

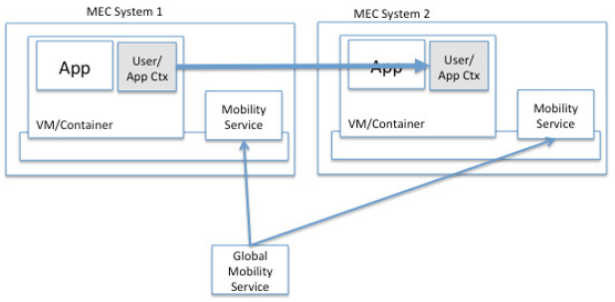
① 모바일 엔티티(Mobile Entity; ME), 모바일 엔티티 호스트(ME Host)와 모빌리티 서비스(Mobility Services)의 확장

각각의 모바일 엔티티는 모바일 엔티티 호스트 내의 가상화된 환경 위에 가상 머신(virtual machine) 혹은 컨테이너(container)로 구성되어 있으며, 빠른 이동성 확보를 위해 사용자 컨텍스트(user context)를 분리하고, 마이그레이션(migration) 속도가 빠른 컨테이너(container)를 사용한다.



〈그림 12. 모바일 엔티티(ME), 모바일 엔티티 호스트(ME Host), 모바일 엔티티 모빌리티 서비스(ME Mobility Services) (그림 출처 [18])〉

모바일 에지(mobile edge)의 이동을 위한 주변 셀(Cell) 정보를 측정 보고(measurement report)를 통해 수집하며, 이 주변 셀 정보 통해 핸드오버 실행하며, 서비스 어플리케이션의 실행 지연 등을 비교하여 애플리케이션 핸드오버를 실행한다.

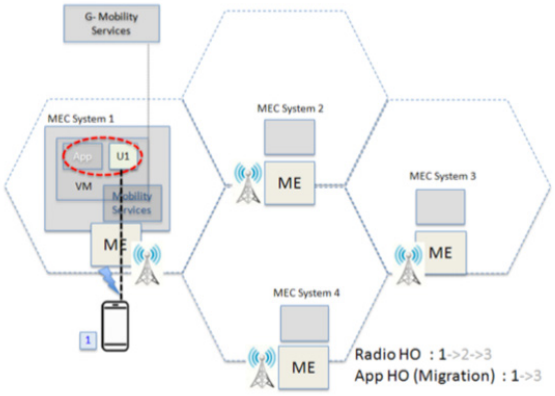


〈그림 13. 서비스 핸드오버시 사용자/서비스 컨텍스트 (user/application context)의 이동〉

② 전역 모바일 엔티티 이동성 서비스 흐름 (ME mobility service flow) 구현 단계

MEC 서비스에 접근한 자율주행 차량의 단말로부터 서비스 탐색 및 발견(Discovery)을 하여 서비스 접속 위치를 찾아낸 후, 인증을 수행한다. 인증이 성공적으로 수행되면 사용자 컨텍스트를 생성하고 API의 호출이 가능하게 된다.

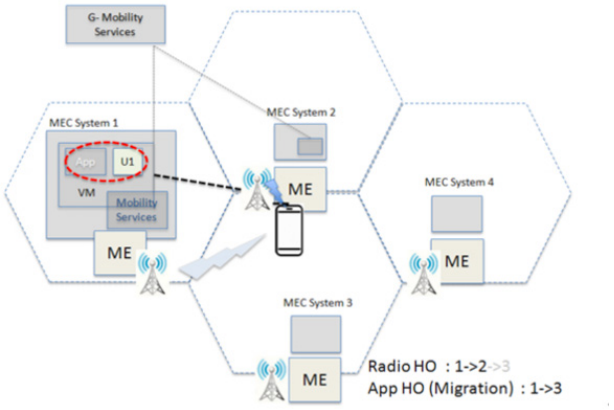
위 인증 후, 차량의 5G 단말에서 MEC 서비스에 대산 세션이 만들어지게 되고, 서비스 사용 내용과 사용자 서비스 컨텍스트의 변경 사항을 저장하게 된다. 모빌리티 그룹(mobility group)의 연결(connectivity) 상태를 확인하여 서비스 인스턴스의 이동에 대비한다. 5G에서 사용하는 주파수 대역의 특성으로 인해 5G RAN의 셀(Cell)이 작기 때문에 주파수 대역 핸드오버 또한 빈번하게 일어나므로 모바일 엔티티의 이동 시 주파수 대역 핸드오버도 같이 대비하여야 한다. 주변 셀(Cell)



〈그림 14. 서비스 세션이 만들어진 후 앵커링(anchoring)된 상태〉

은 모빌리티 그룹(mobility group의 정보를 이용하여 연결 상태를 유지하게 된다. 서비스가 핸드오버 될 경우에는 서비스가 되지 않는 상황의 파라미터 설정을 이용해서 서비스 핸드오버를 유보시킬 수 있다.

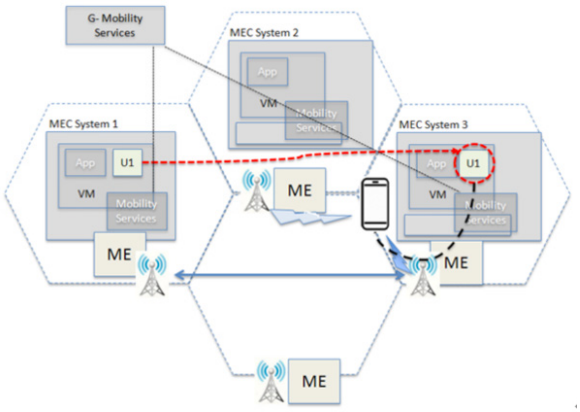
위와 같이 서비스의 세션이 형성되어 핸드오버 요청을 감시하고 있다가, 핸드오버가 일어나야 하는 상황이 되면 먼저 주파수 핸드오버가 일어나야 한다. 주파수 대역 핸드오버가 일어난 후에는 서비스 컨텍스트를 스트리밍 시켜 같이 이동시킨 후, 서비스의 세션을 일단 유지시킨다. 서비스를 구동하는 MEC의 어플리케이션이 사용자의 컨텍스트가 새로운 모바일 엔티티로 이동하면서 컨텍스트 스트리밍이 일어나게 된다.



〈그림 15. 서비스 세션이 서비스 핸드오버 없이 스트리밍되는 상태〉

주파수 대역 핸드오버가 일어나고 사용자 컨텍스트와 서비스 컨텍스트가 새로운 MEC 노드로 스트리밍 된 후에는 서비스 어플리케이션의 세션을 유지한다. MEC 내 지역 세션 정보를 이용해 MEC 노드의 연결(connectivity)을 유지하면서 서비스 핸드오버를 완료하게 된다.

핸드오버 되는 사용자 컨텍스트는 서비스를 핸드오버 받는 MEC 서버의 스토리지에 저장되어야 한다. 서비스 핸드오버 상황에서 서버 간 사용자 컨텍스트 전송 시간을 줄이기 위해



〈그림 16. 서비스 컨텍스트 스트리밍 후 서비스 핸드오버가 완료된 상태〉

컨테이너나 가상 머신의 가상 디스크 이미지 형태로 전송되는 서비스 인스턴스의 정보를 이미지가 변경(delta) 된 부분만을 전송하는 방법을 취해야 한다. 이렇게 이미지의 변경된 부분만 전송하기 위해서는 각 서비스 인스턴스의 이미지 별로 변경 사항만 관리하는 기술이 필요하며, 이를 위해서 가상 머신이나 컨테이너의 라이브 마이그레이션 기술을 사용한다. 특히 컨테이너 이미지를 사용할 경우, 핸드오버 과정이 컨테이너 이미지 마이그레이션으로 가벼워져 지연을 줄이는 효과와 이기종 MEC 노드 간의 이미지 형식을 통합하는 효과를 기대할 수 있다. 컨테이너에서는 이미지 변경 사항만을 관리하여 이미지 전송 시간을 줄이는 기능을 이미 가지고 있으므로, 서비스 컨텍스트 전송에 따른 지연을 줄이는 데에도 효과적이다.

서비스 핸드오버가 수행되고, 사용자 컨텍스트가 새로운 모바일 엔티티로 복사, 전송되고 나면, 자율주행차의 5G 단말은 새로 연결된 셀의 MEC 노드에 마이그레이션 된 서비스로 바로 접속이 가능하게 된다.

V. 결론

완전 자율주행기술이 개발되기까지 자율주행 기술의 신뢰성과 안전성을 높이기 위해 C-V2X(Cellular V2X) 통신을 통한 부가 서비스와 자율주행 인지 범위 확장이 중요하다. 이런 자율주행을 위한 V2X 서비스를 위해서는 자율주행차의 센서에서 생산되는 막대한 양의 빅데이터를 제한된 지연 시간 내에 처리해야 하며, 자율주행차에서 생성되는 빅데이터를 차량간, 차량과 교통 인프라 및 네트워크 서비스 간 공유해 처리, 분석하여 자율주행 제어의 신뢰성과 안전성을 높이게 된다. 자율주행차에서 생성되는 빅데이터뿐만 아니라, 교통 인프라의 다양한 센서 및 인프라에서 생성되는 빅데이터들도 자율주행의 안전성을 향상시키는데 쓰일 수 있다. 자율주행의 실현을 앞당기는데 필요한 요구 사항들을 만족하고, 자율주행차와 교통 인프라 빅데이터를 활용하기 위해 5G 네트워크 기술이 해결해야 할 요구 사항에 대해서 같이 살펴보았다.

자율주행차와 교통 인프라에서 생성되는 빅데이터를 전송, 처리하고 저 지연성을 만족하는 자율주행 V2X 서비스를 제공하기 위한 5G 네트워크 인프라의 조건을 5G 네트워크의 대역폭 요구 사항, 모바일 에지 컴퓨팅(MEC) 인프라를 이용한 분산 처리 및 저 지연성 지원 요구 사항, MEC에서의 서비스 핸드오버를 위한 요구 사항 등으로 나누어 살펴보았다. 특히, 자율주행 빅데이터 처리의 실시간 저지연 처리를 위해서는 MEC를 이용한 분산 처리와 서비스 인스턴스 실행이 필수적이다. 5G 인프라에 MEC를 이용해 빅데이터 처리와 서비스 분산을 구현하기 위해서는 넓은 지역에 걸친 5G 네트워크 인프라의 분산 클라우드 데이터 센터에서의 자원 스케줄링 문제와 분산된 에지 클라우드 간의 서버 어플리케이션의 신속한 핸드오버 기술 문제를 해결해야 함을 살펴보았다. 이들 요구 사항들을 모두 만족하는 5G 기술이 표준화되기는 쉽지 않을 것으로 보이나, 5G V2X 네트워크가 아니더라도 언젠가는 자율주행 서비스의 이동성(mobility)을 위해 모바일 네트워크 기술에서 해결해야 할 문제들이다.

5G 네트워크 기술이 가진 현재의 LTE-A 및 LTE Pro 네트워크 대역폭 대비 10 ~ 100배 대역폭과 1ms 이하의 저 지연성 보장 특성은 자율주행 시대의 빅데이터 처리를 위해 필수적인 특성들이다. 이런 5G 네트워크 기술의 특성이 자율주행의 실현을 앞당길 뿐만 아니라, 자율주행 영상 센서 및 교통 인프라 빅데이터 처리, 분석 문제 해결을 위한 다양한 빅데이터 기술의 발전에도 긍정적인 영향을 끼칠 것으로 기대된다. 빅데이터 기술을 해결하고자 하는 연구자들이 5G 기술과 자율주행에 더 깊은 관심을 가지고 접근해야 하는 이유이다.

References

[1] Maged Zaki, “Accelerating C-V2X toward 5G for autonomous driving,” Qualcomm OnQ Blog, February 25, 2017. (https://goo.gl/GHAuhj)

[2] Pekka Kuure, “Automotive use cases powered by Edge Computing,” Document No. A-170093, a presentation at the 5GAA Working Group 2 Face-to-Face meeting, Amsterdam, 8-10 May, 2017.

[3] Maged Zaki, “The path to 5G: Paving the road to tomorrow’s autonomous vehicles,” Qualcomm OnQ Blog, June 8, 2016. (https://goo.gl/9V1QdU)

[4] LG Electronics, “Discussion on scope of sensor data sharing,” Document No. A-170091, contribution

to 5GAA Working Group 2 technical documents, a presentation at the 5GAA Working Group 2 Face-to-Face meeting, Amsterdam, 8-10 May, 2017.
[5] 3GPP TR 22.886 v.15.0.0, "Study on enhancement of 3GPP support for 5G V2X services."
[6] 3GPP TS 22.186 v.15.0.0, "Enhancement of 3GPP support for V2X scenarios; Stage 1."
[7] ETSI TR 102 863 (V1.1.1), "Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Local Dynamic Map (LDM); Rationale for and guidance on standardization".
[8] Draft ETSI EN 302 895 (V1.0.0), "Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Local Dynamic Map (LDM) Concept for Local Dynamic Maps".
[9] ISO/TS 18750:2015,"Intelligent transport systems -- Cooperative systems -- Definition of a global concept for Local Dynamic Maps."
[10] Delphi Electronically Scanning Radar, http://www.delphi.com
[11] Gary Kotton, GiladZlotkin, "Enterprise-Grade Scheduling - Enterprise-Grade OpenStack from a Scheduler Perspective," a presentation of OpenStack Summit Atlanta, 2014. (https://goo.gl/Mp7T5i)
[12] The OpenStack Documentation Project, "Compute Schedulers," in Configuration Reference, 2017. (https://goo.gl/izALCp)
[13] Yoonsung Nam, Minkyu Kang, Hanul Sung, Jincheol Kim, HyeonsangEom, "Workload-aware resource management for software-defined compute," Cluster Comp., Vol. 19, Iss.3, pp. 1555-1570, 2016.
[14] Yoonsung Nam, Minkyu Kang, Hanul Sung, Jincheol Kim, HyeonsangEom, "How can we allocate "right" resources to virtual machines in virtualized data centers? - Workload-aware hierarchical scheduling with OpenStack," a presentation of the AMGCC Workshop 2015, collocated in the IEEE International Conference on Cloud and Autonomic Computing (ICCAC) 2015, 2015.
[15] Jincheol Kim, Taeil Choi, Jaeseok Ahn, "Can We Really Make an Operating System for Data Center as a Warehouse-scale Computer? The T-RON Operating System with OpenStack," a presentation of the OpenStack Summit Tokyo 2015, October 28, 2015. (https://goo.gl/uv29HA)
[16] Mobile-Edge Computing - Introductory Technical White Paper (http://bit.ly/MECWhitePaper)
[17] ETSI GS MEC 003 V1.1.1 (2016-03)Mobile Edge Computing (MEC); (http://goo.gl/EgsXEx)
[18] Live Service Migration in Mobile Edge Clouds Andrew Machen, Shiqiang Wang, Kin K. Leung, Bong Jun Ko and Theodoros Salonidis (http://goo.gl/JLkL43)

Biographies



김진철

1997년: 한국과학기술원 물리학과 학사
1999년: 포항공과대학교 물리학과 석사 (전산물리학)
2005년: 포항공과대학교 전기 및 컴퓨터공학부 (물리학 전공) 박사
2006년 ~ 2007년: 유럽입자물리학연구소(CERN), 포스트닥
2007년 ~ 2009년: 한국과학기술정보연구원(KISTI) 선임연구원
2009년 ~ 2011년: 포항공과대학교 물리학과 연구교수
2011년 ~ 2013년: 삼성SDS CSP연구소 수석연구원
2013년 ~ 현재: SK Telecom 종합기술원 네트워크기술원 매니저
주요 관심분야: 분산 컴퓨팅, 클라우드 및 빅데이터, 기계 학습(인공신경망 및 학습 알고리즘)
e-mail: Jincheol.B.Kim@sk.com



김동현

2005년: 포항공과대학교 컴퓨터공학부 (분산처리 전공) 석사 졸
2017년 ~ 현재: SK Telecom 종합기술원, 네트워크 기술원 매니저
주요 관심분야: 이동통신, SDN/NFV, MEC 및 분산 컴퓨팅, 클라우드
e-mail: impro.kim@sk.com

보건의료 빅데이터 연계 플랫폼

이경희, 조완섭
충북대학교 비즈니스데이터융합학과, 충북대학교 경영정보학과

Abstract

보건의료 빅데이터는 고도화된 IT 기술과 융합을 통해 수요자 관점에서는 개인맞춤형 진료, 공급자 관점에서는 환자의 안전 증진과 의료서비스 질 향상, 정부 관점에서는 의료비 증가와 의료 형평성 이슈에 대한 대응전략 수립에 널리 활용할 수 있을 것으로 기대되고 있다. 이렇게 빅데이터의 잠재적 가치를 현실화하기 위해서는 단일 기관의 데이터 활용 뿐 아니라 기관간 데이터 연계를 통해서 더 큰 가치 창출이 가능하다. 본 논문에서는 보건의료 데이터 연계에 대한 국내외 국가적 접근방법을 살펴보고, 우리나라 현황에 알맞은 보건의료 빅데이터 연계플랫폼을 설계하고 구축하는 방안을 제안한다.

I. 서론

4차 산업혁명 시대의 빅데이터는 사회 전반에서 현재의 문제를 근원적으로 해결하고 새로운 가치를 창출할 수 있는 혁신의 동력으로 기대를 모으고 있다[5]. 또한 산업 전 분야에 걸쳐 빅데이터와 융합된 기술이 도입되면서 우리나라를 비롯한 많은 국가들은 신산업을 선점하기 위한 정책적 수단 및 전략을 잇달아 발표하고, 포괄적이며 지속가능한 성장을 이어가려는 움직임을 활발히 하고 있다.

특히 보건의료 분야는 빅데이터와 융합을 통해 의료서비스 개선과 비용절감에 크게 기여할 것으로 예상되며, 빅데이터의 잠재적 가치와 활용 가능성이 높은 영역으로 평가되고 있다. 최근의 의료서비스 환경은 고도화된 IT기술과 접목으로 4P로 일컫는 개인맞춤형(Personalized), 예방중심(preventive), 참여형(Participatory), 예측가능(Predictive)한 방향이 강조되고 있다[3]. 이를 반영하여 보건의료서비스는 기존 공급자 중심에서 수요자 중심으로, 질병의 치료 중심에서 지속 건강관리와 예방 중심으로, 경험기반에서 근거기반의 의사결정으로, 일반형에서 맞춤형으로의 전환이 이루어 지고 있다[3].

현 시점에서 정밀의료 및 예측의료 패러다임의 기반이 되는 탄탄하고 효과적인 보건의료 빅데이터 인프라 확보가 매우 중요하며[5], 그것의 기본은 보건의료 빅데이터 연계 플랫폼 구축이라고 할 수 있다.

우리나라는 공적 건강보험제도에 기초한 보건의료 시스템에서 보건의료 빅데이터의 국가적 활용가치를 높이기 위해서는 공익적 활용을 기반을 마련하고, 여기에 산업적 가치 창출을 이끌어낼 수 있는 통합적 전략수립이 필요하다[1]. 또한, 보건의료 빅데이터의 활용은 일부의 표본으로 전체를 추정하는 소극적 방식에서 전 국민의 개인 간 차이를 일시에 고려하여 현안을 통찰하고 일상에서 데이터를 기반으로 의사결정이 가능한 적극적인 방식의 보건의료 시스템 개혁을 재촉할 것이다[1]. 우리나라와 같이 전국민이 공적 건강보험제도를 활용하는 경우 빅데이터를 활용한 현안 통찰은 더욱 유용한 방법이라 할 수 있다.

본 논문은 다음과 같이 구성된다. 2장에서는 관련 연구로 보건의료 데이터 연계의 국가적 접근방법 사례와 보건의료 빅데이터 연계의 우리나라 현황을 살펴본다. 3장에서는 국내 보건의료 빅데이터의 공익적 활용 기반 마련의 한 부분인 보건의료 연계 플랫폼을 구축하기 위한 플랫폼 아키텍처와 소프트웨어 구성을 설명한다. 4장에서는 결론을 맺는다.

II. 관련 연구

본 장에서는 보건의료 데이터 연계에 대한 해외에서의 국가적 접근방법 사례와 우리나라의 현황을 살펴본다.

1. 보건의료 데이터 연계의 국가적 접근방법 사례

민간 또는 공공분야의 현안 통찰과 새로운 서비스, 정책, 비즈니스 발굴을 위해서 서로 다른 소스의 데이터를 통합하고 연결해야 하는 필요성은 점점 높아지고 있다. 개별 기관의 데이터베이스는 그 기관의 고유한 업무에 기반하여 데이터를 수집·관리하므로 새로운 서비스나 정책 발굴을 위해서는 다른 기관의 데이터 연계가 필요하며, 추가정보를 결합함으로써 더 풍부한 데이터 분석 효과를 얻을 수 있기 때문이다.

영국과 미국 등 주요 선진국을 비롯한 세계 각국의 정부는 빅데이터를 통해 공공정보에 대한 접근성을 보장하고 활용도

를 높이기 위한 정책을 활발하게 추진해왔다. 또한, 분산된 보건의료 분야 데이터를 수집, 저장, 연계, 분석하는 독립 조직을 정부 산하에 설립하여 지속적으로 필요한 데이터를 서로 연계하여 활용할 수 있는 전체 과정을 지원하고 관리하도록 하고 있다.

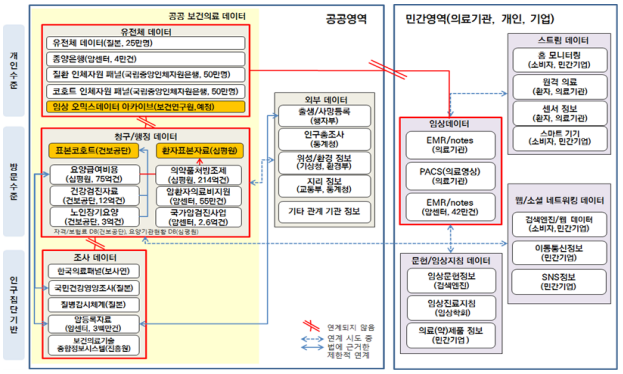
영국의 경우 2013년에 Personalized Health and Care 2020 정책을 발표하고, HSCIC(Health & Social Care Information Center)라는 데이터 연계를 담당하는 조직을 설립하여 국민의 의료와 복지 데이터에 대한 통제권을 강화하고, 분산된 사회보장 데이터를 수집, 저장, 연계, 분석하는 업무를 전담하여 수행하도록 하고 있다 (최근 HSCIC는 NHS Digital 로 명칭변경).

미국의 경우 민간 보험사들을 중심으로 다양한 정보 통합 플랫폼이 활성화되도록 간접적인 조정 역할로서 보건부 산하에 ONC(Office of the National Coordinator)를 설치하여 데이터 상호운용성(interoperability)을 높이기 위한 정보 제공과 분류 표준화 등을 담당하게 하고 있다[1]. 그리고, EHR 도입 및 유의미한 사용(Meaningful Use)에 따른 인센티브 프로그램을 도입하여 의료기관 및 국민들이 EHR 시스템을 도입하도록 유도함으로써 의료기관 및 개인들의 건강정보 교류가 크게 증가하였다[5]. 또한, 미국 국립보건원(NIH)은 2012년부터 빅데이터 R&D 이니셔티브의 일환으로 보건의료 빅데이터 R&D 지원사업인 BD2K(Big Data to Knowledge)를 추진 하고 있다. BD2K는 빅데이터 구축, 분석도구 개발, 인력교육 등에 R&D 지원을 집중하고 있다. 공적자금을 투입한 연구개발의 결과물인 데이터가 지속적으로 활용 및 발전시켜 나갈 수 있도록 연구데이터 공유를 장려하는 정책을 추진중이다. 연구자의 연구실적평가시 논문의 피인용지수를 고려하는 것과 같이 연구데이터의 인용을 수치화하여 연구데이터 공유가 활발한 연구자에게 인센티브를 주어 국가연구개발과제 선정 등에 활용하고 있다. 이와 관련하여 연구데이터의 이해에 필수적인 상세한 수준의 메타데이터를 개발하고 있다.

지속적인 데이터 연계 과정은 개별 기관의 데이터가 다른 기관 데이터와 매칭하는 과정을 반복하게 됨으로써 데이터 상호운용성을 향상시키는데 필요한 프로세스가 계속해서 동작하는 효과를 갖기 때문에 데이터의 완결성과 품질이 높아지게 된다. 예를 들면, 다른 기관 데이터와 연계가능성이 높은 코드성 데이터의 분류체계 관리 및 관련 표준 준수율 향상, 코드성 데이터 변경이력관리 강화, 메타데이터 관리 강화 등 데이터 품질향상 효과를 기대할 수 있다.

2. 보건의료 빅데이터 연계에 대한 우리나라 현황

우리나라는 공적 국민건강보험 시스템 기반 위에 높은 수준



〈그림 1. 우리나라 보건의료 데이터와 연계 현황〉

의 IT 기술을 활용한 데이터 축적이 장기간 진행되어 왔다. 보건의료 빅데이터 활용을 위한 기본계획 수립 연구 보고서[2]에서 밝힌 것과 같이 공공기관 내부의 데이터 통합은 추진되고 있으나 기관 간 데이터 연계는 법적으로 허용된 업무 수행을 위해서만 제한적으로 이루어지고 있는 실정이며, 민간영역에서 보유하고 있는 의미있는 임상데이터가 국가적으로 활용되지 못하고 있다(그림 1).

공적인 보건의료 행정데이터, 지역사회기반 코호트 데이터, 암을 비롯한 국가적 레지스트리 데이터, 인체자원데이터, 각종 조사데이터 등 대규모 공적자금이 투입된 데이터부터 민간병원 중심으로 구축된 환자임상 데이터의 규모가 보건의료 빅데이터라고 하기에 충분하다.

또한, [1]에서는 우리나라 보건의료 빅데이터가 높은 수준의 IT기술 활용으로 공공과 민간 영역 모두 상당한 수준의 규모와 다양성으로 데이터가 구축되고 있으나 국가적으로 보건의료 빅데이터를 구축하고 활용하는 데는 다음과 같은 한계가 있다고 지적하고 있다.

- 기관별로 분산된 보건의료 데이터가 연계·통합되어 국가적으로 의미있는 활용할 수 있는 법적, 기술적, 정책적 기전이 부족함
- 국가 단위 빅데이터 구축의 한계는 오픈데이터를 통한 새로운 가치창출도 제한하고 있음
- 유전체 데이터의 활용을 통한 질병발생 기전 등의 임상지식 창출도 제한적임

미국과 영국의 보건의료 데이터 연계 목적을 달성하기 위한 국가적 컨트롤 타워를 수립하고, 관련 기관이 충실히 참여함으로써 보건의료 빅데이터를 활용할 수 있는 탄탄한 인프라가 구축되고 있음을 알 수 있다.

이와 비교하여 우리나라는 개별 기관 중심의 빅데이터 인프라 설치에 집중하고 있고, 보건의료정보의 교류와 활용을 위한 법령은 아직 마련되지 않았다. 현행 개인정보보호 관련법

령에서는 개인의 신상정보보호를 위한 규정들을 담고 있으나, 건강정보 보호 및 이용(저장 및 이용방식)에 대한 규정은 미비하다[3]. 또한, 보건의료관련 법령에서 의료정보의 비밀을 보호하기 위한 여러 규정을 두고 있으나, 정보교류의 확장성을 고려한 표준 및 인증에 관련된 법령은 아직 마련되지 않았다[3].

이러한 국내 상황에서 최근 보건복지부에서 "보건의료 빅데이터 추진단"을 구성하여 보건의료분야의 4차 산업혁명을 주도하겠다는 목표로 빅데이터 활용과 연계를 위한 법과 제도정비, 새로운 서비스 창출, 데이터 연계의 걸림돌 제거, 안전한 활용을 위한 정보보호 방안, 대국민 소통 증진방안 등을 종합적으로 연구하고 있어 그 결과가 기대되고 있다[6].

III. 보건의료 빅데이터 연계플랫폼 구축

본 장에서는 보건의료 연계 플랫폼을 구축하기 위한 플랫폼 아키텍처와 소프트웨어 구성을 설명한다.

1. 보건의료 빅데이터 연계 플랫폼 아키텍처[7]

보건의료 빅데이터 연계 플랫폼은 현재 국내 보건의료 공공정보, 민간정보,해외기관의 LOD를 다양한 방식으로 수집, 저장 또는 연계하여 일반국민, 연구자, 정책입안자가 하나의 플랫폼에서 데이터 신청부터 수집, 연계, 다운로드를 수행하는 것을 목표로 한다. 플랫폼에서 다루는 데이터는 건강보험공단의 표본코호트DB, 심사평가원의 환자표본데이터, 질병관리본부의 지역사회 기반 코호트, 통계청 사망원인마이크로데이터, 조사데이터(국민건강영양조사, 지역사회건강조사, 의료패널)와 환경데이터(대기질, 수질) 그리고 해외 기관의 LOD 이다.

사용자가 연계데이터포털에서 데이터셋 생성시나리오를 설계하면 데이터를 연계하여 연계 데이터셋을 생성하고 사용자가 다운로드 할 수 있도록 제공한다.

• 보건의료 빅데이터 연계 플랫폼 구조

연계 플랫폼은 데이터의 수집·통합·연계·저장을 담당하는 빅데이터 플랫폼과 관리자의 관리 부담을 줄이면서 사용자 편의성을 제공할 수 있도록 지원하는 사용자 포털(연계 플랫폼 포털)로 구성하였다.

일반적인 빅데이터 플랫폼은 빅데이터 기반 인프라와 분석 플랫폼으로 구성되지만, 본 과제의 경우 보건의료 빅데이터 연계 플랫폼의 개발이 최종 목적이므로 보건의료 원천데이터를 수집·통합·연계·저장한 데이터웨어하우스와 이를 바탕으로 외부 데이터와 연계하여 다양한 연계 데이터셋을 생성하는 기능을 구현하였다. 또한 연계 플랫폼의 사용자 편의성을 제고하기 위해 연계 플랫폼 포털을 구현하며, 이를 통하여 사용자 관리(사용자 인증과 사용자별 데이터 접근통제), 연계 데이터셋 생성 및 수명관리 기능을 제공할 수 있다.

보건의료 빅데이터 연계플랫폼의 구조는 그림 3과 같이 총 6개의 레이어로 구성하였다.

사용자는 서비스 레이어를 통해 보건의료 연계플랫폼을 사용하고, 각 레이어는 사용자에게 제공하는 서비스를 생성, 관리, 제공하기 위한 역할을 담당한다. 레이어별 역할은 다음과 같다.

• 서비스 레이어 : 연계플랫폼 포털의 사용자 서비스 영역
• 거버넌스 레이어 : 보건의료 데이터 관리, 표준화, 품질관리, 사용자 권한관리, 데이터 라이프사이클관리, 정보보안 및 보호를 담당하는 영역
• 응용 레이어 : 보건의료 데이터 연계 플랫폼의 서비스를



〈그림 2. 보건의료 빅데이터 연계 플랫폼 개념도[7]〉



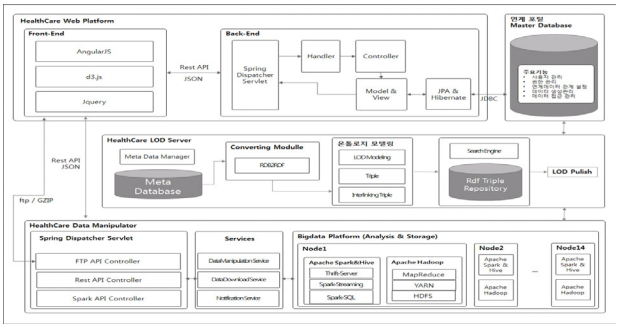
〈그림 3. 보건의료 빅데이터 연계 플랫폼 구조〉

생성하는 영역으로 LOD기반 보건의료 플랫폼 기술연구, 정보연계 및 지식화 연구 그리고 시나리오 기반 분석연구 컴포넌트로 구성

- 주요논리 레이어 : 연계대상이 되는 자원을 관리하는 영역과 시나리오 기반 지식 추출하는 영역
- 저장 레이어 : 보건의료 기관으로부터 수집한 원천소스 데이터 저장영역
- 인프라 스트럭쳐 레이어 : 플랫폼의 하드웨어 영역

2. 보건의료 연계 플랫폼의 소프트웨어 구성

보건의료 연계플랫폼의 소프트웨어 구성은 그림 3과 같이 HealthCare Web Platform, HealthCare Data Manipulator, Master Database의 구성요소를 포함한다.



〈그림 4. 보건의료 연계플랫폼의 소프트웨어 구성도〉

각 구성요소별 기능은 다음과 같다.

HealthCare Web Platform : 보건의료 빅데이터 연계플랫폼의 Master Database는 RDB (SQL-SERVER)로 구성되어 있으며, 거버넌스 체계를 반영한 연계플랫폼의 사용자 계정관리, 데이터접속 권한관리, 사용자 데이터셋 관리 등의 기능을 수행한다.

- HealthCare Web Platform Front-End
  (역할) 사용자와 플랫폼의 인터페이스를 담당하며, 직관적이고 빠르며 유연한 서비스 제공담당
  (구현기술) Spring Web MVC Framework 기술로 구현되어 있으며, 내부적으로 Front-End 는 AngularJS, Javascript, Jquery 언어를 이용하여 구현함

- HealthCare Web Platform Back-End
  (역할) 보건의료 연계플랫폼의 Front-End의 서버측 비즈니스 로직을 수행하고, Database와 통신함
  (구현기술) Spring Web MVC 프레임워크, Spring JPA와 Hibernate 등 ORM(Object Relation Mapping)기술의 유연하고 고도화된 기술을 활용함

HealthCare LOD Server : 보건의료 연계 플랫폼의 LOD(Linked Open Data)를 위한 시스템으로 데이터공개를 위한 데이터 수집/변환/공개 역할을 수행한다.

- Meta Data Manager
  (역할) 연계플랫폼 포털에서 정의된 데이터 거버넌스 체계를 활용하여 LOD 생성을 위한 체계적인 데이터 수집, 변환, 공개 역할을 수행을 담당
  (구현기술) 메타데이터 수집 및 관리기술

- Converting Module
  (역할) 연계플랫폼 포털에서 정의된 데이터 거버넌스 체계를 활용하여 LOD 생성을 위한 체계적인 데이터 수집, 변환, 공개 역할을 수행을 담당
  (구현기술) LOD 모델링 기술, LOD 변환기술, LOD 출판서버

- Ontology Modeling
  (역할) 연계플랫폼 포털에서 정의된 데이터 거버넌스 체계를 활용하여 LOD 생성을 위한 체계적인 데이터 수집, 변환, 공개 역할을 수행을 담당
  (구현기술) Protege

- LOD Search Engine
  (역할) 연계플랫폼 포털에서 정의된 데이터 거버넌스 체계를 활용하여 LOD 생성을 위한 체계적인 데이터 수집, 변환, 공개 역할을 수행을 담당
  (구현기술) 위탁기관 자체 개발 모듈 이용

HealthCare Data Manipulator : 보건의료 연계플랫폼에서 빅데이터 플랫폼의 데이터의 생성, 삭제, 갱신 등의 데이터 조작을 담당하는 소프트웨어로, 실제 빅데이터 플랫폼의 서버측 소프트웨어와 통신하여 제어하는 역할을 수행한다.

- HealthCare Data Manipulator
  (역할) 보건의료 연계플랫폼 Front-End와 빅데이터 플랫폼과의 중간 매개역할을 담당
  (구현기술) Spring, Spring RestAPI, Srping JPA(Java Persistance API; 퍼시스턴스 API), Apache Thrift, Apache Hadoop, Apache Hive, Apache Spark를 활용하여 빅데이터 플랫폼과

고성능으로 효과적으로 데이터를 저장 및 추출, 분석처리 할 수 있음

• Manipulator Service Layer

(역할) Manipulator Service Layer로 연계데이터 생성 및 삭제, 수정 서비스, 초고속 분석서비스, 데이터 다운로드 서비스, 알림 서비스를 제공한다.

(구현기술) Spring, Spring RestAPI, Srping JPA(Java Persistance API; 퍼시스턴스 API), Apache Thrift, Apache Hadoop, Apache Hive, Apache Spark를 활용하여 사용자의 서비스를 구현하기 위한 비즈니스 로직을 빅데이터 플랫폼에 전달하고 태스크의 모니터링 및 스케쥴링 관리를 수행한다.

• BigData Platform

(역할) 대용량 데이터의 안전한 분산 저장 및 초고속 데이터 분석을 위한 서비스를 제공한다.

(구현기술) Spring, Spring RestAPI, Srping JPA(Java Persistance API; 퍼시스턴스 API), Apache Thrift, Apache Hadoop, Apache Hive, Apache Spark를 활용하여 빅데이터 플랫폼과 고성능으로 효과적으로 데이터를 저장 및 추출, 분석처리할 수 있다. 특히, 빅데이터의 안정적 저장을 위해 Hadoop Distributed File System(HDFS)를 사용하여, 하나의 컴퓨팅 노드 고장이 전체 시스템에 영향을 주지 않는 견고한 시스템을 구성하였음. 또한 Spark 데이터 엔진을 탑재하여 대용량의 데이터를 인메모리 기술을 활용하여 초고속으로 처리 및 분석이 가능하도록 구성하였다.

IV. 결론

보건의료 빅데이터는 고도화된 IT 기술과 융합을 통해 수요자 관점에서는 개인맞춤형 진료, 공급자 관점에서는 환자의 안전 증진과 의료서비스 질 향상, 정부 관점에서는 의료비 증가와 의료 형평성 이슈에 대한 대응전략 수립에 활용할 수 있을 것으로 기대하고 있다[5]. 빅데이터의 잠재적 가치를 현실화하기 위해서는 단일 기관의 데이터가 아니라 기관간 데이터 연계를 통해서 가능하다. 이와 같이 데이터의 연계의 중요성이 점점 더 높아지고 있으며, 유럽, 미국, 호주 등에서는 2000년대 초반부터 정부가 주도하는 데이터 연계 정책과 전략을 대규모 장기간 프로젝트로 추진중이다.

본 논문에서 설명하는 보건의료 빅데이터 연계 플랫폼은 현재 국내 보건의료 공공정보, 민간정보, 해외 기관의 LOD를 다양한 방식으로 수집, 저장 또는 연계하여 일반국민, 연구자, 정책입안자가 하나의 플랫폼에서 데이터 신청부터 수집, 연계, 다운로드를 수행하는 것을 목표로 한다. 보건의료 데이터 연계에 대한 국내외 국가적 접근방법을 살펴보고, 우리나라 현황에 알맞은 보건의료 빅데이터 연계플랫폼을 설계하고 구축하는 과정을 설명하였다.

플랫폼은 제도적 전략적 배경 위에서 그 프로세스가 구현되는 것으로써 다양한 서비스가 살아 움직이는 운동장과 같다고 할 수 있다. 그런데, 데이터의 단절이 존재한다면 운동장에 넘어갈 수 없는 블록이 존재하는 것과 같으며 결국 어떤 서비스를 구현하더라도 보건의료 빅데이터 연계 플랫폼은 시뮬레이션 도구로만 존재하게 될 것이다. 보건의료관련 법령에서 의료정보의 비밀을 보호와 정보교류의 확장성을 균형적으로 조율할 수 있는 전략과 법률 마련이 필요하다.

Acknowledgement

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술연구개발사업 지원에 의하여 이루어진 것임(과제고유번호: HI16C1094)

References

[1] 강희정. 보건의료 빅데이터의 정책 현황과 과제. 보건복지포럼, 2016, 55-71.

[2] 강희정 외, 보건의료 빅데이터 활용을 위한 기본계획 수립 연구 보고서, 보건복지부, 2015.

[3] 신지은, 조대곤, 김병인, 고영명, 기영민, 정인수, 강성홍, 박정선, 곽미숙, 황희. (2016). 국가 보건의료정보화의 정보전략계획(ISP) 수립 방안. 대한산업공학회지, 42(3), 198-208.

[4] GROVES, Peter, et al. The big data revolution in healthcare: Accelerating value and innovation, 2016.

[5] The Office of the National Coordinator for Health Information Technology, Federal Health IT Strategic Plan Progress Report, The Office of the National Coordinator for Health Information Technology, 2013.

[6] "보건의료 빅데이터 전략 콘퍼런스, 혁신막은 의료정보 유통규제 걷어내야," 디지털타임스, 2017년 5월 12일

[7] 조완섭 외, 보건의료빅데이터 연계플랫폼, 중간연구보고서, 한국보건산업진흥원 (충북대학교), 2016.

Biographies



이 경 희

1995년: 충북대학교 컴퓨터과학과 학사
1997년: 충북대학교 전자계산학과 석사
2004년: 충북대학교 전자계산학과 박사
주요 관심분야: 빅데이터 컴퓨팅, 데이터 연계, LOD
e-mail : lee.kyunghee@gmail.com



조완섭

1985년: 경북대학교 통계학과 학사
1987년: KAIST 전산학과 석사
1996년: KAIST 전산학과 박사
주요 관심분야: 빅데이터, 비즈니스 인텔리전스,
데이터베이스 보건의료 빅데이터
e-mail : wscho@chungbuk.ac.kr

공공 빅데이터 개방과 표준화

박경현, 차우, 원희선
Electronics and Telecommunications Research Institute (ETRI)

Abstract

공공데이터가 국가 경제발전의 자원으로 인식되기 시작함에 따라 세계 각국에서는 공공데이터 포털을 구축하여 민간에게 공공데이터를 개방하기 시작하였다. 이러한 흐름에 맞추어 오픈소스 진영에서도 CKAN을 선두로 오픈데이터 플랫폼을 개발하여 발전하기 시작하였고, 카탈로그 데이터의 표준 기술을 적용함으로써 타 플랫폼과의 연동도 가능해지게 되었다. 이에 본 논문에서는 공공데이터 유통의 활성화를 위해 국내외 정부기관들이 진행하고 있는 공공데이터 개방 현황과 오픈데이터 플랫폼의 활용 동향에 대해 살펴보고 마지막으로 데이터 카탈로그의 표준화 기술에 대해 살펴본다.

I. 서론

최근 인공지능 기술의 발달로 데이터가 점점 더 중요한 자원으로 인식되면서 세계 각국은 공공데이터 기반의 다양한 비즈니스 기회를 창출하기 위한 오픈데이터 정책을 수립하고 시행하고 있다. 이와 더불어 사회 전분야에 걸친 대용량의 공공데이터를 효과적으로 저장, 관리하고 공유하기 위한 솔루션으로 오픈데이터 플랫폼이 확산, 발전하고 있다.

공공데이터는 공공기관에 의해 생성되고 관리되는 방대한 데이터로 기상, 지리정보로부터 교통, 식품에 이르기까지 사회 전분야에 걸친 방대한 데이터를 일컫는다. 국내에서 공공데이터란 용어는 "공공데이터의 제공 및 이용 활성화에 관한 법률"이 제정된 이후에 공식적으로 사용되었으며 그 이전에는 정보데이터, 행정데이터, 지식자원정보, 공공정보등의 다양한 용어로 사용되었다.

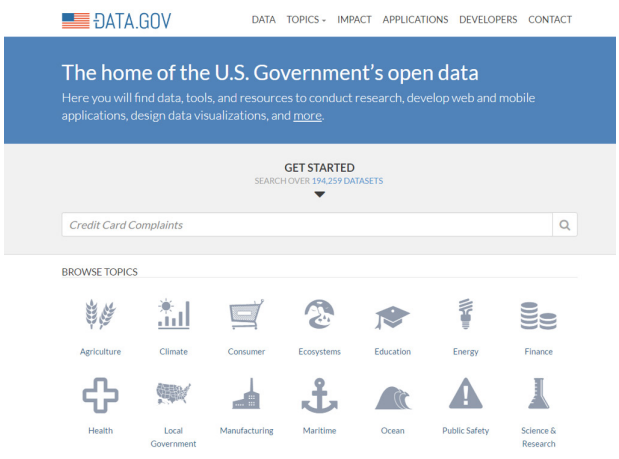
이와 더불어 공공데이터를 효과적으로 공유하고 활용하기 위한 소프트웨어로 오픈데이터 플랫폼이 확산, 발전하고 있는데 대표적인 오픈소스 플랫폼으로 CKAN[7]과 DKAN[8]이 있고 상용 플랫폼으로 Socrata[9]가 있다.

본 논문에서는 정부기관 등에 의해 진행된 공공데이터의 개방 현황과 주요 오픈데이터 플랫폼의 활용 동향을 살펴보고 데이터 포털간의 상호연동을 위한 데이터 카탈로그 표준기술에 대해 살펴보기로 한다.

II. 공공 빅데이터 개방

공공데이터가 경제적 자산으로 인식되어 감에 따라 자국의 경제 활성화를위해 국내외적으로 공공데이터의 개방을 위해 노력하고 있는데 미국의 경우, 2009년부터 정부기관에 의해 운영되고 있는 공공데이터 포털인 data.gov를 중심으로 정부기관의 데이터들을 민간에게 제공하고 있다[4]. data.gov는 공공데이터를 기후, 에너지를 포함하는 21개의 주제로 분류하였고 약 10만여개의 데이터셋(dataset)을 제공하고 있으며 공공데이터의 약 31.9%가 XML, CSV와 같은 오픈 데이터 포맷으로 구성되어 있다.

공공데이터 활용의 활성화를 위해 미국 정부는 2014년에 Open Data Action Plan을 발표하고 미국내 전체 주를 대상으로 오픈데이터 수준평가를 수행하였고, 맥킨지는 OPEN DATA 500 프로젝트를 통해 미국내 500개 기업의 공공 데이터의 활용와 영향에 대해 연구를 진행하고 있다. 이러한 미국내 공공데이터 활성화 정책에 힘입어 Zillow, WeatLaw와 같은 대규모 민간 기업들이 공공데이터를 활용하여 큰 수익을

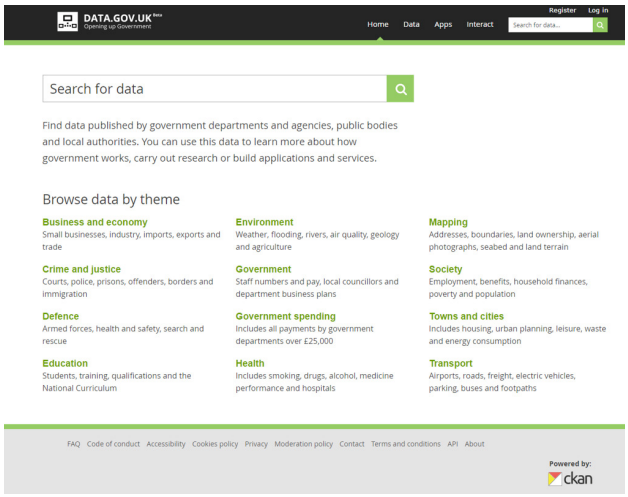


〈그림 1. 미국의 공공데이터 포털(data.gov)[4]〉

창출하기 시작하였고 공공데이터를 활용하는 대규모 기업들이 서서히 증가하기 시작하였다.

공공데이터 개방 및 활성화를 선도하는 대표적인 국가인 영국은 2010년부터 내각사무국을 중심으로 본격적인 오픈데이터 정책을 추진하고 있고 공공데이터 포털인 data.gov.uk를 구축하여 민간에게 공공데이터를 제공하고 있다. data.gov.uk는 약 1,000개 이상의 기관에서 약 15만개의 데이터셋을 제공하고 있고 공공데이터의 44.8%를 오픈포맷 형태로 제공하고 있다[5].

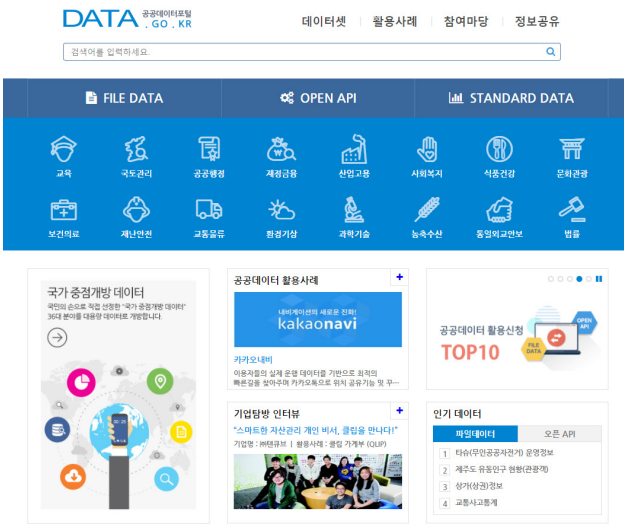
또한 영국은 2014년에 Open Data Strategy를 통해 오픈데이터의 5단계 발전방향을 제시하여 데이터 오픈원칙을 세우고 이를 기준으로 평가결과를 공개함으로써 공공데이터의 활성화에 노력하고 있다.



〈그림 2. 영국의 공공데이터 포털(data.gov.uk)[5]〉

국내에서는 공공데이터의 개방 및 활성화를 위해 2013년에 공공데이터 개발 로드맵을 수립하였고 공공데이터 포털인 data.go.kr를 구축하여 민간에게 공공데이터를 개방하였다[6]. 공공데이터 포털은 2017년 현재 파일데이터 약20,000건, 오픈API 약 2,300건, 표준데이터 44건을 공개하고 있으며, 행정자치부는 기업과 대국민 수요조사를 통해 인공지능, IoT, 빅데이터와 관련된 데이터들을 계속 발굴하고 제공하기 위해 노력하고 있다.

2013년에는 WWW 재단이 오픈데이터 현황 보고서인 ODB(Open Data Barometer)를 발표하였는데 보고서에서 3개 평가항목(준비도, 실행력, 효과)을 기준으로 국가별 순위를 평가하였다. 그 결과, ODB 보고서에서 한국은 ODB 평점 54.12로 12위를 차지하였고 이에 한국은 2014년에 "오픈데이터 5대 강국도약을 위한 공공데이터 개방 발전전략"을 수립하여 ODB(Open Data Barometer) 2017년에 데이터 오픈 순위 5위, 국가중점 공공데이터를 30개 이상, 오픈포맷 데이



〈그림 3. 한국의 공공데이터 포털(data.go.kr)[6]〉

터를 50%이상으로 목표를 설정하고 이를 달성하기 위한 방안으로 민간-공공간의 데이터 생태계 조성을 위한 정책과 공공데이터포털을 오픈데이터 플랫폼으로 구축하기 위한 정책을 수립하여 추진하고 있다.

III. 오픈데이터 플랫폼 기술

3.1 오픈데이터 플랫폼 정의

오픈데이터 플랫폼은 데이터 공급자가 소유하고 있는 데이터를 공개하여 데이터 수요자가 효율적으로 검색하고 활용할 수 있도록 하기 위한 데이터 유통 플랫폼으로 오픈데이터를 효율적으로 관리하기 위해 다음과 같은 기능들을 포함하고 있다.

• 데이터 등록

데이터 등록이란 데이터 소유가자 데이터의 개방을 위해 플랫폼에 데이터를 업로드 및 저장하는 것을 말한다. 오픈 데이터 플랫폼은 데이터 등록을 위해 판독이 가능한 데이터 형태를 정의하고 XML, CSV, RDF와 같은 오픈 데이터 포맷을 지원한다. 또한 데이터 등록을 위한 사용자 인터페이스 및 API를 제공한다.

• 데이터 발행

데이터 발행이란 데이터 사용자들이 데이터에 접근하여 활용할 수 있도록 데이터를 제공하는 기능을 말한다. 오픈데이터 플랫폼은 데이터 등록과 마찬가지로 판독이 가능한 데이터 형태와 오픈 데이터 포맷을 제공해야 하고 데이터 제공에 따

른 데이터 카탈로그 표준, 데이터 보안, 데이터 품질관리 기능을 추가로 제공해야 한다.

• 데이터 현황

데이터 현황은 데이터에 대한 정보를 관리하는 기능으로 일반적으로 메타데이터 관리 또는 데이터 카탈로그 관리라고도 말한다. 오픈데이터 플랫폼은 데이터의 출처, 분류, 내용 및 데이터의 이력정보등을 제공해야 하며 데이터의 원활한 사용을 위해 메타데이터를 기반으로 다양한 검색 기능을 제공한다.

• 데이터 포털

데이터 포털은 데이터 개방 및 활용을 위해 웹 브라우저를 통해 오픈데이터 플랫폼에 접근할 수 있도록 웹 기반의 사용자 인터페이스를 제공한다. 오픈데이터 플랫폼은 웹 인터페이스를 통해 사용자들이 데이터 유통에 필요한 전반적인 기능들을 사용할 수 있는 환경을 제공해야 하고 그 외에도 고객센터 및 운영 관련 서비스 등을 지원해야 한다.

• 시각화 기능

시각화 기능이란 도표, 그래프등을 이용하여 데이터 및 데이터 분석 결과를 시각적으로 제공하는 기능을 말한다. 시각화 기능은 플랫폼마다 기술적이 차이를 보이는 기능이며 일반적으로 오픈소스 플랫폼에 비해 상용 플랫폼이 풍부한 시각화 기능을 제공하고 있다.

3.2 주요 오픈데이터 플랫폼

공공데이터 유통을 위한 공공데이터 포털은 자체 개발하거나 오픈데이터 플랫폼을 기반으로 구축된다. 현재 전세계적으로 활용되는 대표적인 오픈데이터 플랫폼으로는 CKAN[7], DKAN[8], OGPL[10], Socrata[9]등이 있다.

CKAN(Comoprehensive Knowledge Archive Network)은 오픈데이터 플랫폼중 가장 많이 사용되는 오픈소스 기반의 오픈데이터 플랫폼으로 비영리단체인 OKF(Open Knowledge Foundation)에 의해 개발되었다. 현재 CKAN은 영국, 미국, 캐나다등 40여개국에서 사용하고 있으며 콘텐츠 관리, 시각화, API 추출등과 같이 CKAN에서 제공하지 않는 기능들은 타 오픈소스와의 통합 인터페이스를 통합되어 발전하고 있다. CKAN은 자체적으로 제공하는 플랫폼의 기능 외에 CKAN API를 제공하고 있어 별도의 서비스 구축도 가능하다. 또한 기본적으로 데이터 검색 기능, 키워드 유사 매칭 기능, CSV 데이터 시각화 기능을 제공하고 기능확장을 위한 목적으로 다양한 플러그인을 추가할 수 있게 설계되었다.

DKAN은 Drupal을 기반으로 만들어진 오픈데이터 플랫폼으로 미국의 NAUM에 의해 개발되었다. DKAN은 CKAN의 기능 대부분을 포함하고 있고 프로그램 언어와 데이터베이스 지원 같은 기술적인 측면에서 차이점을 보이고 있다.

OGPL(Oopen Government Platform)은 미국과 인도 정부에서 공동으로 개발한 오픈데이터 플랫폼으로 미국과 인도 정부의 오픈데이터 플랫폼의 장점만을 모아 개발하고 있는 오픈데이터 플랫폼이다.

Socrata는 클라우드 기반의 오픈데이터 플랫폼으로 data.gov와 미국의 여러주에서 사용중인 데이터 플랫폼이다. Socrata는 SODA API와 SoQL(SODA Query Language)를 지원하고 성능향상을 위해 Truth 스토어를 지원한다. 이외에도 Socrata는 오픈소스 기반의 데이터 플랫폼에 비해 풍부한 분석 및 시각화 기능을 제공하고 있다. 하지만 오픈소스인 CKAN이 전세계적으로 사용이 확대됨에 따라 Socrata도 오픈소스로 개방하려는 노력을 진행중이다.

IV. 오픈데이터 플랫폼 활용 현황

본 장에서는 각 국가별로 공공데이터 포털의 기반이 되는 오픈데이터 플랫폼의 활용 현황에 대해 살펴보고자 한다.

미국은 2009년에 Socrata를 기반으로 공공데이터 포털인 data.gov를 구축하여 운영하고 있고 현재는 CKAN, Drupal Hybrid와 같은 오픈소스의 도입을 진행중이다. data.gov는 다른 오픈데이터 플랫폼과와 연동은 지원하지 못하지만 분석 및 통계 서비스를 통해 데이터 시각화 기능을 제공하고 있다.

영국은 2010년 CKAN을 기반으로 공공데이터 포털인 data.gov.uk를 구축하였고 CKAN의 사용이 전세계적으로 확대되자 공공데이터 포털(data.gov.uk)의 소스코드를 CKAN 커뮤니티를 통해 공개하고 오픈데이터 플랫폼의 표준화 작업을 선도하고 있다.

한국은 2011년 오픈데이터 플랫폼인 공공데이터 포털(data.go.kr)를 구축하고 '공공데이터의 제공 이용 활성화에 관한 법률'에 따라 각 기관별로 생성, 보유하고 있는 공공데이터를 통합관리하고 있다. 공공데이터 포털은 자체개발하였으며 메타데이터 관리를 위해 CKAN 카탈로그를 활용하고 있다.

공공데이터 포털은 데이터를 파일 데이터, 오픈 API, 표준 데이터로 구분하여 민간에게 제공하고 국내 다른 데이터 포털

들과 연동을 지원하고 있지만 각 기관이 제공하는 원문 데이터를 단지 링크하는 수준으로 단순한 데이터 연계만을 지원하고 있는 실정이다.

V. 데이터 카탈로그 표준 기술

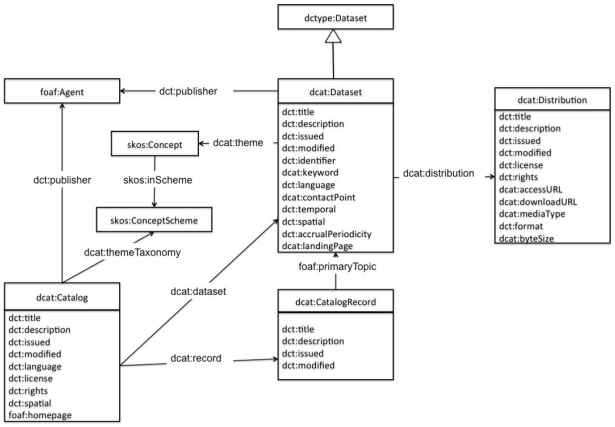
오픈데이터 플랫폼은 데이터 카탈로그를 통해 데이터를 관리하고 다양한 검색기능을 제공한다. 또한 다른 플랫폼과의 정보교환 및 검색도 데이터 카탈로그를 통해 이루어지기 때문에 오픈데이터 플랫폼의 활용에 있어서 데이터 카탈로그의 표준화는 가장 중요한 요소중의 하나이다.

데이터 카탈로그 표준기술 중 가장 대표적인 기술로는 웹상의 데이터를 통합 관리하기 위한 메타데이터 표준인 DCAT이 있다. DCAT(Data Catalog Vocabulary)은 웹상에 존재하는 카탈로그 데이터 간에 상호연동성을 제공하기 위한 W3C 표준으로 웹상의 다양한 데이터 소스로부터 메타데이터를 읽어 데이터의 접근 및 활용을 가능하게 할 수 있도록 RDF 포맷로 정의되어진다[11].

DCAT은 유연한 확장성으로 인해 CKAN, DKAN, Socrata와 같은 많은 오픈데이터 플랫폼에서 적용하고 있고, 실제 data.gov와 data.gov.uk와 같은 많은 공공데이터 포털에서 데이터 연동을 위해 활용되고 있다. DCAT은 그림 4와 같이 3개의 주 클래스인 Catalog, DataSet, Distribution와 1개의 중요 클래스인 Distribution을 중심으로 구성된다.

DCAT-AP는 DCAT을 기반으로 유럽의 데이터 포털이 공공데이터를 발행 및 연동을 위한 DCAT 응용 프로파일이다. ODS(Open Data Support)는 데이터 포털간의 연동을 위해 필요한 최소한의 속성들을 프로파일에 정의하고 있다[12].

DCAT의 대표적인 활용 사례로, CKAN을 개발한 OKFNLab은 유럽 데이터 포털인 publicdata.eu 프로젝트를 통해 DCAT을 기반으로 CKAN 카탈로그를 확장하고 있으며 데이터 상호운용성을 지원하기 위한 표준화 작업을 진행중에 있다. 또한, CKAN은 DCAT을 지원하기 위해 RDF기반의 직렬화 기능을 포함하는 플러그인을 개발하여 추가함으로써 CKAN이 외부의 메타데이터와 상호운용성을 가지도록 하였다.



〈그림 4. DCAT 데이터 모델 [11]〉

VI. 결 론

전 세계적으로 정부기관을 중심으로 하는 공공데이터 개방이 이루어지고 있으며 주요 몇몇 국가들은 개방된 공공데이터의 다양성과 품질 및 사용자 활용 측면에서 상당한 수준에 이르렀다. 국내에서도 정부 3.0을 중심으로 공공데이터의 민간 개발에 대한 법률을 제정하고 로드맵을 작성하는 등 공공데이터의 활성화에 박차를 가하고 있다.

하지만 제한된 범위의 공공데이터 공개, 데이터 표준화 및 데이터 품질관리의 부재, 민간과의 협업 부족, 공공데이터 포털간 상호연동 부재등은 아직 정부가 풀어가야 할 숙제로 남아있다. 현재 많은 국가들이 CKAN을 중심으로 공공데이터 포털을 구축하여 공공데이터를 개방하는데 주력하고 있다. 이에 국내 공공데이터 포털들도 CKAN과 같은 오픈데이터 플랫폼을 도입하여 적용하고 데이터 표준화에 집중함으로써 한국의 공공데이터 포털이 세계의 공공데이터 개방 및 활성화에 앞장서기를 기대한다.

Acknowledgement

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(2017-0-00253, 국제표준 기반 오픈 데이터 유통 플랫폼 확장 기술 개발)

References

[1] 정국환, 문정욱, 이시직, 유지연, 한은영, 왕재선, 서혁준, " 공공데이터 개발, 활용 성과측정을 위한 평가모델 연구", 정보통신정책연구원, 2013.
[2] 백인수, “오픈데이터 플랫폼과 국가 데이터 전략방향”, IT & Future Stragegy 제16호, 한국정보화진흥원, 2013
[3] 정유철, 서동준, 이혜진, 김광영, “오픈 데이터 플랫폼 동향”, 정보처리학회지 23권 5호, 2016년
[4] 미국 공공데이터 포털, https://www.data.gov/
[5] 영국 공공데이터 포털, https://data.gov.uk/
[6] 한국 공공데이터 포털, http://data.go.kr/
[7] CKAN, The open source data portal software,http://

ckan.org/
[8] DCKAN, http://getdkan.com/
[9] Socrata, The open data platform for digital government, https://socrata.com/
[10] Open Government Platform(OGPL), http://ogpl.github.io/index-en.html
[11] W3C Data Catalog Vocabulary(DCAT), https://www.w3.org/TR/vocab-dcat/
[12] DCAT Application Profile for Data Portals in Europe, https://joinup.ec.europa.eu/asset/dcat_application_profile/description

Biographies



박 경 현

1999년: 충북대학교 컴퓨터공학 학사
2001년: 충북대학교 전산학 석사
주요 관심분야: 대용량 데이터관리시스템, 빅데이터 플랫폼, 클라우드 컴퓨팅
e-mail: hareton@etri.re.kr



차 우

2009년: 베트남 호치민대학교 전산학 학사
2013년: KAIST 전산학 석사
주요 관심분야: 빅데이터 관리, 분산 시스템, 소트프웨어 아키텍쳐
e-mail: chau@etri.re.kr



원 희 선

1990년: 연세대학교 전산과학과 학사
1992년: KAIST 전산학 석사
2016년: KAIST 전산학 박사
주요 관심분야: 빅데이터, 클라우드 플랫폼, 오픈 데이터 거버넌스
e-mail: hswon@etri.re.kr

빅데이터 거버넌스와 표준화 동향

조완섭
충북대학교 경영정보학과

Abstract

실시간으로 쏟아지는 빅데이터를 분석하여 국가나 기업의 경쟁력을 혁신하는 사례가 급증하면서 빅데이터가 4차 산업혁명의 원동력이 되고 있다. 최근 몇년간 공공분야는 물론 제조업이나 보건의료 등 전체 산업분야를 혁신하는데 있어 빅데이터가 중요한 역할을 하고 있다. 각국은 정부가 보유한 데이터를 적극적으로 개방하여 데이터 기반의 혁신을 이루어가는 경쟁을 하고 있다. 그러나 빅데이터에 대한 활용이 증가하고, 중요한 의사결정에 빅데이터가 사용될수록 빅데이터에 대한 체계적인 관리(거버넌스)가 중요하다. 빅데이터 거버넌스는 데이터의 품질보장, 프라이버시 보호, 데이터 수명관리, 데이터 전담조직을 통한 데이터 소유 및 관리권의 명확화 등 체계적인 데이터 관리를 의미한다. 본 고에서는 빅데이터 거버넌스에 관해 살펴보고 빅데이터 관련 표준화 동향을 소개한다.

I. 서론

빅데이터는 한대의 컴퓨터에는 저장하거나 처리할 수 없을 정도로 큰 데이터이면서 비정형적이고 빠르게 생성되는 특징을 가진다. 이러한 빅데이터를 수집, 관리, 분석하여 기업은 물론 국가의 경쟁력을 강화하는 사례가 속속 발표되면서 범국가적으로 생성되는 빅데이터를 원동력으로 하여 4차 산업혁명을 선도하려는 경쟁이 치열하다. 미국과 영국 및 독일 등 선진국들은 공공 데이터를 적극적으로 민간에 개방하여 정부와 공공의 효율성 제고는 물론 민간 산업의 경쟁력을 획기적으로 높이고 있다. 이들 데이터 선진국들은 공공행정은 물론 금융분야, 보건의료분야, 제조분야, 농업분야 등에 빅데이터와 인공지능 기술을 접목하여 전체산업 분야를 혁명적으로 업그레이드하고 있어 우리나라와의 격차를 더욱 벌이고 있는 실정이다.

우리나라에서도 정부 3.0을 시작으로 공공데이터 개방 정책을 강력히 시행하면서 지난 몇 년 동안 데이터 공개가 급속히 증가하고 있다. 또한 공공기관 등에서 빅데이터의 중요성을 인식하고 있다는 점에서 일단은 긍정적인 변화이다. 그러나 공개된 데이터가 부실하고, 품질이 낮으며, 표준화가 되어 있지 않고, 여러 기관들에서 공개한 데이터를 연계 (메쉬업)하여 가치가 높은 분석을 하는 것이 불가능하며, 개인정보보호 등을 이유로 개인수준의 가치높은 데이터가 공개되지 않고 있어 사용자들의 불만이 높은 실정이다[8]. 실제 공공데이터 포털에 공개된 데이터의 12%만 활용된다는 조사결과도 있다. 빅데이터 전문인력 양성측면에서도 대학 등에 체계적인 지원이 부족하여 인력수요에 효과적으로 대응하지 못하고 있다. 지난 수년동안 우리나라의 전자정부에 대한 국제기관들의 평가는 상위권에 위치해 있지만 데이터 부문의 평가는 10위권에서 20위권 후반으로 평가되어 4차 산업혁명의 원동력이 될 데이터 문제가 심각함을 보여주고 있다.

이러한 데이터 부문에서의 총체적인 문제는 데이터에 대한 체계적인 관리(데이터 거버넌스, Data Governance)에 대한 인식부족에서 비롯된 것이다[5,11]. 최근에 선진국들은 데이터를 조직의 중요한 자산 (예를들어 현금이나 주요시설 등)으로 간주하여 체계적인 관리에 공을 들이고 있다. 일본의 경우 유용한 데이터셋에 대하여 특허권 인정을 검토할 정도로 데이터에 관한 중요성을 강조하고 있다. 그러나 우리나라는 아직 HW/SW만 잘 갖추면 데이터는 자동으로 해결되는 것으로 착각하는 경우가 많다. 그러나 최신의 고가 HW/SW를 갖추고도 데이터를 쓰레기 같이 관리하는 조직도 있고, 그 반대인 조직도 많다. 다시 말하면, 일정 수준의 HW/SW 시설을 갖추는 것은 빅데이터를 잘 관리하기 위한 필요조건이지 충분조건은 아니라는 의미이다.

데이터 거버넌스(Data Governance)란 전사적으로 보유하고 있는 데이터에 대해 관리 정책, 지침, 표준, 전략 및 방향을 수립하고, 데이터를 관리할 수 있는 조직 및 서비스를 구축하는 데이터 관점에서의 IT 관리 체계(IT 거버넌스)를 의미한다[5]. 빅데이터 시대가 되면서, 데이터가 4차 산업혁명의 원동력으로 인식되면서 데이터 거버넌스의 중요성은 더욱 강조될 것으로 보인다. 조직내 데이터 거버넌스가 확립되지 못하면 중요한 의사결정에 품질이 낮은 데이터를 사용함으로써 오류를 유발할 수 있고, 개인 프라이버시 관련 데이터로 인해 빅브라더의 우려가 현실화될 수 있으며, 폭증하는 데이터의 수명관리 소홀로 인해 IT 비용의 급증과 유출의 위험을 초래하기도

한다. 기존의 금융권이나 글로벌 기업들은 전사적인 데이터베이스 구축과 활용 뿐 아니라 데이터 거버넌스 구축을 통해 효과적이고도 완전한 데이터 관리체계를 실현하여 데이터를 조직의 중요한 자산으로 관리해 나가고 있다. 특히, 빅데이터 시대에는 크고(Volume), 다양하며(Variety), 빠른(Velocity) 데이터를 다루어야 하므로 데이터 거버넌스 문제가 더욱 어렵고, 복잡하게 된다. 따라서 기존 데이터 거버넌스보다 확장된 데이터 거버넌스의 구축을 통해 지속 가능한 빅데이터를 실현하는 것이 중요한 과제이다.

II. 빅데이터 거버넌스와 표준화 현황

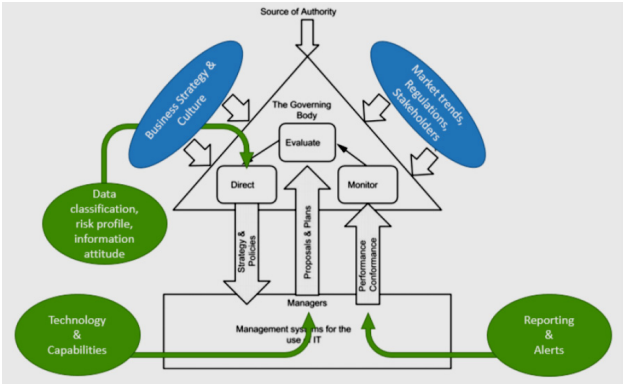
여기서는 데이터 거버넌스에 관한 소개후에 빅데이터 및 빅데이터 거버넌스의 표준화 동향을 간단히 소개한다.

2.1 빅데이터 거버넌스

데이터 거버넌스(data governance)란 전사적인 차원에서 보유하고 있는 모든 데이터에 대해 관리 정책, 지침, 표준, 전략 및 방향을 수립하고, 데이터를 관리할 수 있는 조직 및 서비스를 구축하는 데이터 관점에서의 IT 관리 체계(IT 거버넌스)를 의미한다[5]. 데이터 거버넌스는 관한 정의는 기관마다, 문헌마다 다양하게 정의하고 있으며, 데이터 거버넌스를 IT 거버넌스의 일부로 보는 시각도 있지만 분명한 차이도 존재한다. 데이터 거버넌스에서는 데이터셋을 조직의 중요한 자원으로 관리한다. 한편, 빅데이터 거버넌스는 기존의 데이터 거버넌스로부터 빅데이터의 특징 (volume, velocity, variety)을 반영하여 확장한 것으로 정의할 수 있다[5]. 특히, 빅데이터의 3가지 특징으로 인해 기존 데이터 거버넌스가 더욱 확장되고 중요한 이슈로 부각되고 있다[5].

데이터 거버넌스의 궁극적인 목적은 고품질 데이터의 확보와 관리 및 적극적인 활용을 통해 조직의 다양한 가치 창출에 지속적으로 기여하는 것이다. 이것은 데이터로 인한 리스크를 미리 예측하여 대비하고, 데이터 관리 비용의 절감과 데이터 활용도 증대를 통해 데이터의 가치를 향상시키며, 고품질 정보가 조직의 비즈니스 목적에 부합하고 최적의 서비스를 지속적으로 제공할 수 있도록 효과적으로 관리하는 것이다.

데이터 거버넌스를 조직의 데이터 관리활동 (데이터베이스 구축과 관리, 데이터의 품질관리, 메타 데이터 관리, 마스터 데이터 관리, 프라이버시 관리, 데이터 수명관리 등)과 연관시켜 설명하는 것이 이해도를 높이는데 도움이 된다. 그림 1은 국제표준화기구(ISO, International Standardization Organization)에서 제시하는 데이터 거버넌스 개념을 보여

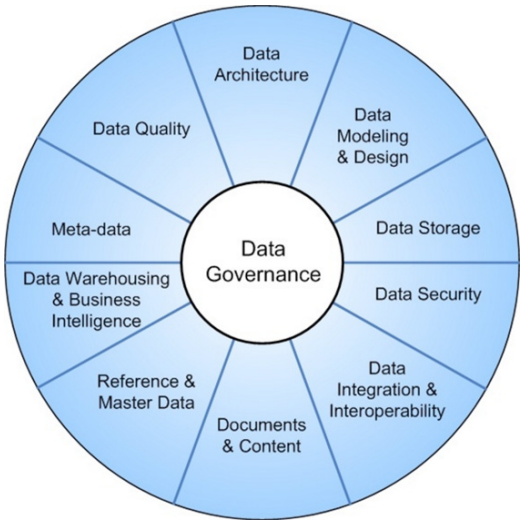


〈그림 1. 국제표준화기구의 데이터 거버넌스 개념[6]〉

주고 있다[6] (아직 공식 표준으로 확정되지는 않았으며, 표준화가 진행중에 있음). 그림 1에서 데이터 거버넌스는 조직의 데이터 관리기능이 효과적으로 완전하게 이루어지도록 평가, 감독, 모니터링(EDM: Evaluation, Direction, and Monitoring)하는 체제를 의미한다.

- 조직의 각 부문에서 현재와 미래의 데이터 관련업무와 성과를 평가함(Evaluation)
- 데이터 관련 부문이 조직의 비즈니스 목적에 부합하도록 전략과 정책을 수립하고 시행하는지를 지휘감독 함 (Direction)
- 데이터와 관련된 전략과 정책이 조직 전반에서 합당하게 시행되는지 모니터링함 (Monitoring)

그림 1의 하단 박스는 조직의 데이터 관리기(data managers)를 의미한다. 조직의 데이터 관리기능은 다양한 세부 업무들로 구성되며, 그림 2는 이들 관리기능들을 보여주고 있다. 각 데이터 관리업무들은 방대한 기술적인 내용을 포함하고 있으며,



〈그림 2. 조직의 데이터 관리 업무〉

SW 패키지로 구현되거나 혹은 사람(제도 등)에 의해 그 목적이 달성되고 있다. 데이터 관리부문을 지원하기 위해 다양한 SW 도구들이 개발되어 보급되고 있다. 빅데이터 시대에는 이러한 각각의 데이터 관리기능이 빅데이터의 특징을 반영하여 더욱 확장되는 추세이다. 예를들어, 데이터 스토리지 기능은 하둡(Hadoop)이나 NoSQL과 같이 분산시스템의 장점을 충분히 활용하는 형태로 확장되고 있으며, 데이터 품질 관리도 빅데이터의 다양성 특징으로 인해 더욱 복잡하게 발전하고 있다.

2.2 빅데이터 관련 표준화

빅데이터가 다양하게 활용됨에 따라 국제표준화 기구의 여러 하부 그룹들에서 빅데이터에 관한 표준화 작업을 진행하고 있다. 여기서는 국제기구에서 수행되고있는 빅데이터 표준화 동향을 살펴본다.

국제표준화기구(ISO/IEC JTC 1/WG 9)에서는 빅데이터에 대한 전반적인 표준화를 수행하고 있다. WG 9에서 추진중인 빅데이터 표준화 과제의 주제와 일정은 다음과 같다[13].

- ISO/IEC 20546: Big Data – Overview and Vocabulary (Dec. 16, 2015 – Oct. 9, 2018)
- ISO/IEC 20547 Part-1: Big Data Reference Architecture – Framework and Application Process (May 3, 2016 – May 5, 2019)
- ISO/IEC 20547 Part-2: Big Data Reference Architecture – Use Cases and Derived Requirements (May 3, 2016– May 5, 2019)
- ISO/IEC 20547 Part-3: Big Data Reference Architecture – Reference Architecture (Dec. 16, 2015 – Dec. 15, 2018)
- ISO/IEC 20547 Part-4: Big Data Reference Architecture – Security and Privacy Fabric (under SC27/WG4 with collaboration with SC 27/WG 5 & JTC 1/WG 9)
- ISO/IEC 20547 Part-5: Big Data Reference Architecture – Standards Roadmap (May 3, 2016 – May 5, 2019)

이와 함께 WG 9에서는 데이터 거버넌스와 데이터 품질관련 표준화를 각각 SC 40 등과 연계하여 추진하고 있다. 데이터 거버넌스 표준화는 SC 40에서 IT 거버넌스와 동일한 수준에서 표준화를 추진하고 있으며, 곧 표준규격에 제정될 것으로 보인다. 그러나 여기서는 빅데이터 보다는 기존 데이터에 대한 거버넌스에 초점을 맞추고 있어 향후 확장이 필요할 것으로 보인다. 또한 데이터 품질 관련 표준화는 ISO/TC 84/SC 4/WG 13과 ISO/TC 84/SC 4 /WG 23과 연계하여 추진하고 있으며, 여기서도 기존 데이터에 관한 품질 표준화를 근간으로 빅데이터의 특성을 반영하여 확장하는 문제가 제기될 것으로 보인다.

- JTC 1/SC 40 on IT and Data Governance work
- ISO/TC 84/SC 4/WG 13 and ISO/TC 84/SC 4 /WG 23 on Data Quality Management

마지막으로 기존의 데이터베이스 표준을 담당하고 있는 ISO JTC 1 SC32에서도 빅데이터에 관한 다양한 측면들을 표준화하고 있다.

- SQL과 NoSQL 사이의 정보교환 및 연계를 지원하기 위한 SQL 확장
- 메타 데이터 표준
- 데이터 마이닝을 지원하는 SQL 및 NoSQL
- 데이터 프로비넌스
- 대용량의 복잡한 데이터 구조를 지원하는 SQL 혹은 SQL/MM
- 차세대 분석과 빅데이터 (Next generation analytics and big data)

또한, ITU에서도 2015년 다음과 같은 빅데이터 관련 표준문서를 발간하였다.

Recommendation ITU-T Y.3600 “Big data – cloud computing based requirements and capabilities”

III. 결론

빅데이터의 활용이 강조될수록 빅데이터에 대한 체계적인 관리(거버넌스)가 중요한 문제로 부각될 것이다. 이는 자동차가 널리 활용되기 위해 관련 법규나 조직 및 인력이 필수적인 것과 유사한 것이다. 실시간으로 쏟아지는 빅데이터의 분석결과로부터 미래에 관한 통찰력을 얻고, 중요한 의사결정에 활용하기 위해서는 데이터의 품질보장, 프라이버시 보호, 데이터 수명관리, 데이터 소유 및 관리권의 명확화 등이 함께 이루어져야 하기 때문이다. 빅데이터가 중요한 의사결정에 사용될수록 더 큰 리스크를 가질 수 밖에 없으며, 이러한 리스크는 적절한 데이터 거버넌스를 통해서 관리될 수 있을 것이다.

빅데이터 거버넌스는 시스템 도입으로 한 번에 이루어지는

것이 아니라 CEO의 의지와 함께 지속적인 노력으로 각 분야에서 점진적으로 개선해 나가는 것이 중요하다. 빅데이터가 활성화될수록 조직에서 사용하는 데이터 소스가 다양해지고, 그 형태는 비정형적이며, 데이터의 생성과 분석 및 활용주기도 빨라지게 될 것이다. 또한, 빅데이터가 활용되는 분야도 점차 확대되어 조직의 수많은 업무들과 관련성을 가지게 될 것이다. 더구나 빅데이터 거버넌스의 요소들 하나하나도 간단한 문제가 아니다. 예를들어 품질문제만 보아도 다양한 빅데이터의 품질을 어떻게 측정할지, 품질 기준을 어떻게 마련할지, 통합시 어떠한 표준을 사용할지 등등 복잡한 문제에 직면하게 된다. 데이터 품질을 높이기 위해 전문 인력을 유치하거나 기존 인력의 재교육이 필요할 수 있으며, 새로운 시스템이나 장비가 추가로 설치할 필요도 있다. 나머지 빅데이터 거버넌스 요소들도 복잡한 문제를 내포하고 있으므로 지속적인 노력과 교육 및 시스템 도입을 통하여 빅데이터 거버넌스를 강화해 나가는 것이 중요하다.

특히, 빅데이터 거버넌스의 상당부분을 사람에 의존하는 것보다 시스템으로 구축하여 자동화해 나가는 노력도 병행해야 한다. 예를들어, 최신의 데이터 수집기에는 데이터 품질을 개선하는 다양한 기능과 메타 데이터를 자동으로 관리하는 기능이 포함되어 있으므로 사람의 개입을 최소화하는 방향으로 빅데이터 거버넌스를 강화해 나가야 할 것이다.

Acknowledgement

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술연구개발사업 지원에 의하여 이루어진 것임(과제고유번호: HI16C1094)

References

[1] 문흥근, IT Governance의 개념과 추진방향, SAMSUNG SDS Consulting Review, No. 3, 2006
[2] Information technology - Governance of IT for the organization, ISO/IEC 38500
[3] Board Briefing on IT Governance 2nd Edition, ITGI, 2003
[4] COBIT 모델로 살펴본 IT Governance의 이해, 2014.04.14 (http://blog.lgcns.com/473)
[5] 조완섭 외, 빅데이터 거버넌스, 빅데이터 활용의 전제요건, 홍릉과학출판사, 2014
[6] Information Technology - Governance of IT - The Application of ISO/IEC 38500 to the governance of data, ISO/IEC JTC1/SC40 N199 (2015.2.3.).
[7] http://nosql-database.org/
[8] 개인정보보호에 꽉 막힌 빅데이터 산업… 걸음마도 못 떼는 한국, 매일경제 2017.03.26
[9] http://legaltalkmedia.com/LTN/TDD/TDD_051412_BPOil.mp3
[10] http://www.cgrey.be/rm_map/Germany.html
[11] 조완섭, 빅데이터 거버넌스 - 지속가능한 빅데이터 활용의 전제요건, 주간기술동향, 한국정보통신산업진흥원, 2013.11.13
[12] Mars Climate Orbiter Fact Sheet, http://mars.jpl.nasa.gov/msp98/orbiter/fact.html
[13] Wo Chang, ISO/IEC JTC 1/WG 9 Status Report to JAG, July 24, 2017.

Biographies



조 완 섭

1985년: 경북대학교 통계학과 학사
1987년: KAIST 전산학과 석사
1996년: KAIST 전산학과 박사
주요 관심분야: 빅데이터, 비즈니스 인텔리전스, 데이터베이스 보건의료 빅데이터
e-mail : wscho@chungbuk.ac.kr

www.etri.re.kr



미래를 창조하는
ICT Innovator

ICT Innovator, For a Great Tomorrow

ETRI의 최첨단 기술은
새로운 영역과 새로운 가치를 만들어내고 있습니다.
꿈꾸는 모든 것을 현실로 구현하는 기술,

ETRI의 첨단ICT기술은
창조경제를 견인하는 성장동력입니다.

ETRI 40th Anniversary 1976 - 2016