

## 빅데이터의 이해와 주요 이슈들

조 영 임

### 〈요 약〉

빅데이터는 저장되지 않았거나 저장되더라도 분석되지 못하고 버리게 되는 방대한 양의 데이터를 말한다. 실제로 빅데이터가 스마트 사회에서는 매우 많이 발생하고 있는데, 이러한 빅데이터를 분석하여 유의미한 지식을 발생시킴으로써 가치창출을 이루고 자원을 절감하며 범죄율을 낮추고 우리의 미래를 예측하여 대처할 수 있는 수단을 제공할 수 있다는 것은 매우 흥미로운 일이다. 따라서 본 논문에서는 스마트 사회의 변화에 따른 빅데이터의 등장배경, 개념적 이론적 이슈를 설명하고 빅데이터 도입과 관련된 논쟁들에 대해 살펴보고, 빅데이터가 지역정보화 발전에 가지는 시사점과 이슈들에 대한 대응방안과 향후 지역정보화차원에서 빅데이터 연구시 고려할만한 사항들에 대해 논의하고자 한다.

〈주제어〉 빅데이터, 빅데이터 플랫폼, 전자지방정부, 지역정보화

## I. 서 론

최근 정보통신 기술의 일상화가 이루어지면서 하드웨어, 소프트웨어에 이어 데이터가 정보화 사회의 새로운 핵심기술로 등장하고 있다. 이는 과거 하드웨어에서 소프트웨어, 그리고 데이터로 관심영역이 점차 확대되고 있다는 것을 의미한다. 이러한 데이터로의 관심영역의 중심에는 빅데이터가 포함되어 있어서 많은 이들과 영역에서 이슈가 되고 있다.

이러한 빅데이터로의 발전은 정보사회의 패러다임을 견인할 정도의 큰 힘을 발휘하고 있음은 누구도 부인하기 어렵다. 스마트 단말기 확산, SNS 활성화, 클라우드, M2M(Machine to Machine) 네트워크의 확산으로 데이터 폭발이 더욱 가속화되어 빅데이터의 기반이 확대되고 있다. 따라서 향후 5년 이내 빅데이터를 위한 고급분석 등 관련기술은 성숙될 것으로 전망하고 있다(한국정보화진흥원, 2011).

이미 알려진 바와 같이 우리나라 전자정부는 2010, 2012년 연속 세계 1위를 차지하고 있으며 2014년에도 1위 달성이 무난한 것으로 전망하고 있다(행정안전부, 2012). 이와 더불어

전자지방정부도 같은 맥락 속에서 추진되고 있다. 특히 전자지방정부에서는 지역정보화를 발전시키기 위한 수단으로 여러 정보화 시스템들이 개발되어 추진하고 있어서 지방정부 운영차원에서 데이터에 근거한 행정을 기대하고 있으나 실제로는 이러한 필요성에 비해 적당한 데이터나 방법이 적시에 제시되지 못하는 경향이 있다. 이미 지역정보화를 추진하는 주체들인 공공과 민간영역에서 많은 양의 유형, 무형의 데이터가 잘 인식하지 못하는 사이에 생성되고 소멸되기도 하지만, 공공과 민간의 융합에 따른 모델창조를 통한 전자지방정부의 구현을 통한 활성화가 필요하다. 따라서 본 논문에서는 빅데이터와 관련된 논쟁적인 이슈에 대한 정리와 지역정보화 차원에서 연구시 고려해야 할 이슈 등에 대한 논의를 통해 지역정보화를 추진하는데 있어서의 대응방안 및 시사점 등을 모색해보고자 한다.

전자지방정부와 지역정보화의 개념상 유사점과 차이점에 대해서는 여러 논란이 있으나(서진완, 2004), 본 논문에서는 빅데이터를 지역정보화에 활용함으로써 전자지방정부의 운영관점에서 시사점과 대응방안을 논의하고자 하는 것이므로 두 가지 개념을 같은 개념으로 정의하여 사용하고자 한다.

본 논문의 구성은 다음과 같다. II장에서는 전자정부 및 스마트 사회의 변화에 따라 필수적으로 등장하는 빅데이터 이슈 및 빅데이터의 개념적·이론적 이슈를 정리하고, III장에서는 빅데이터 도입과 관련된 여러 논쟁들에 대해 살펴보고, IV장에서는 빅데이터가 지역정보화 발전에 가지는 시사점과 이슈들 및 이에 대한 대응방안을 논의하고자 한다. 마지막으로 V장에서는 향후 지역정보화차원에서 빅데이터 연구시 고려할만한 사항들에 대해 제시하고자 한다.

## II. 빅데이터와 관련된 등장배경, 개념적, 이론적 이슈

### 1. 등장배경 및 특징

#### 1) 등장배경

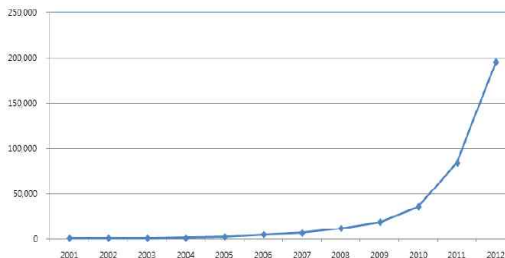
인터넷이 일상화된 지 10여년이 지난 지금 디지털 데이터가 양적으로 증대하고 있어서 전 세계 데이터에 생성될 디지털 정보량이 2011년 1.8 제타바이트<sup>1)</sup>(zetta bytes)의 데이터가

2015년에는 7.9 제타바이트로 증가할 것이며, 2020년에는 50배 급증하게 되어 10배 많은 서버가 필요할 것이라고 전망되고 있어서, 기존의 데이터 관계된 기술로는 감당할 수 없는 규모로 성장하게 될 것이므로 향후 5년 이내 빅데이터를 위한 고급분석 등 관련기술이 예상되고 있다(한국정보화진흥원, 2010; 정보통신산업진흥원, 2011).

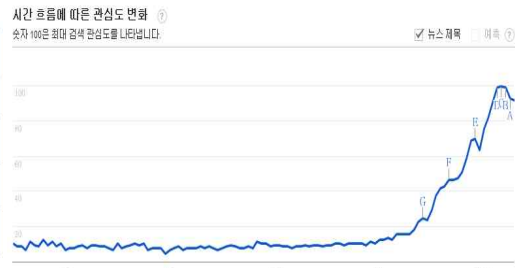
빅데이터를 포함하는 웹문서의 개수를 구글 검색결과로부터 연도별 추이로 표현하면, <그림 1(a)>에서와 같이 2012년에 생성된 빅데이터 관련 웹문서가 2011년에 생성된 빅데이터 관련 웹문서의 두 배가 넘고 있다(양혜영, 2012). <그림 1(b)>에서와 같이 구글 트렌드<sup>2)</sup> 검색기준으로 빅데이터에 대한 검색결과, 2010년까지는 특별한 변화를 보이지 않다가 2011년부터 증가하기 시작하여 특히 2012년에 빅데이터에 대한 검색이 급증하였으며 2013년에 5월달에 최대검색만족도인 100%에 달하고 이후 8월에 약간 낮아졌으나 최대치를 기록하고 있음을 알 수 있다.

### <그림 1> 빅데이터 시대의 시작

(a) 빅데이터를 포함하는 연도별 웹문서의 개수



(b) 빅데이터 검색빈도의 시계열적 변화



2013년 8월 15일 기준으로 <그림 2>는 구글 트렌드에서의 빅데이터에 대한 지역별 관심도를 나타낸 것으로 국가별로는 인도, 대한민국, 싱가포르 순으로 높게 나타났다<sup>3)</sup>. 구글 트렌드에서는 2013년 8월 15일 현재 이슈가 되는 빅데이터 관심 용어를 보면 <그림 3>과 같이 빅데이터 분석가(data analytics) 관련 용어에 대해 관심이 매우 높은 것으로 나타나고 있다. 데이터의 유형 또한 <그림 4>와 같이 매우 다양하게 변화하고 있어서 이러한 비정형 정보들이 향

1) 1.8 제타바이트는 우리나라 모든 사람들이 17만 847년 동안 쉬지 않고 매 분마다 트윗 글 3개를 게시할 경우 생성되는 양이다.

2) <그림 1> <그림 2> <그림 3> <그림 5>의 출처:(<http://www.google.co.kr/trends>)

3) 지역별로는 인도지역(Bangalo, 하이데라바드, 첸나이)이 가장 높고, 다음으로 서울이 높은 것으로 나타났다.

후 10년 동안 생성되는 전체 데이터의 약 90%에 달할 것이라 전망하고 있다(양혜영, 2012).

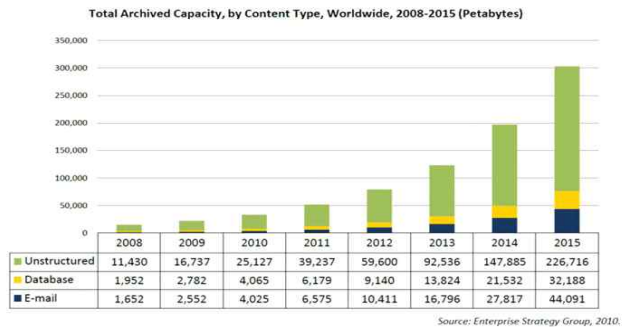
〈그림 2〉 빅데이터에 대한 지역별 관심도(2013.8.15. 기준)



〈그림 3〉 빅데이터 관심용어



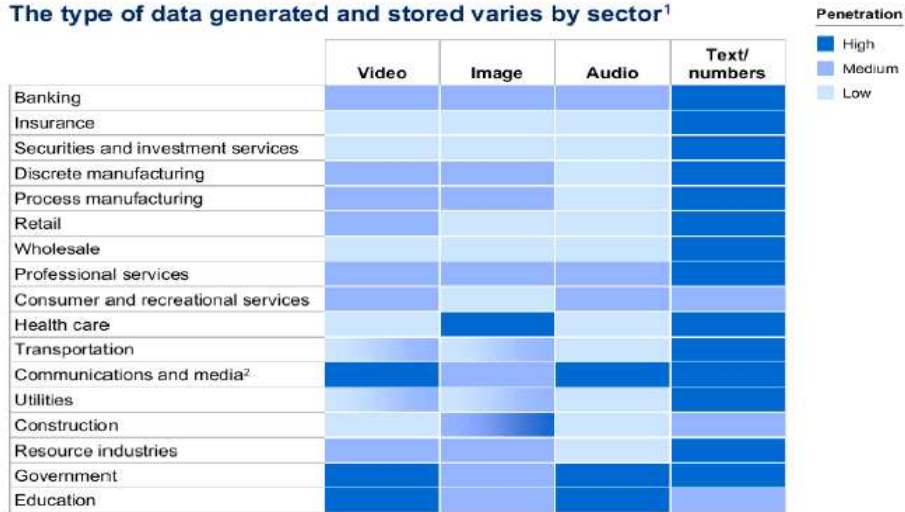
〈그림 4〉 빅데이터 유형전망



Source: Enterprise Strategy Group, 2010.

분야별 빅데이터를 전망하면, 다음 <그림 5>와 같이 비디오, 이미지, 오디오, 텍스트 종류별 은행, 보험회사, 정부 등에서 생성되는 빅데이터의 유형을 구분한 것이다. 특히 정부나 커뮤니케이션 미디어는 비디오, 오디오, 텍스트 데이터가 매우 많고 이미지 데이터도 많은 것으로 나타났다. <그림 6>은 빅데이터를 활용했을 경우 가치가 창출될 수 있는 예상정도는 나타낸 것으로 동그라미 크기가 클수록 예상정도가 높다는 것을 나타낸 것인데, 여기서 정부는 예상정도가 매우 높게 나타남을 알 수 있다.

〈그림 5〉 분야별 빅데이터 종류

The type of data generated and stored varies by sector<sup>1</sup>

1 We compiled this heat map using units of data (in files or minutes of video) rather than bytes.

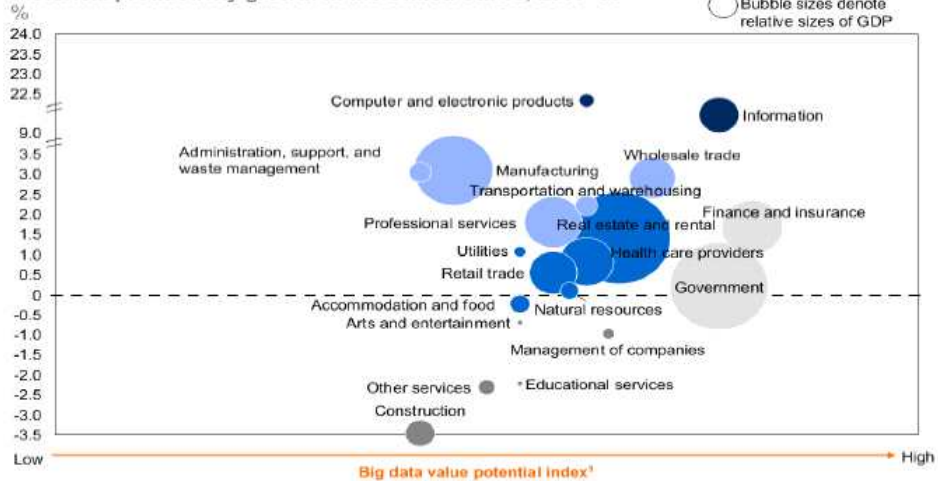
2 Video and audio are high in some subsectors.

SOURCE: McKinsey Global Institute analysis

〈그림 6〉 빅데이터로 부터 가치창출 예상정도

Some sectors are positioned for greater gains from the use of big data

Historical productivity growth in the United States, 2000–08



1 See appendix for detailed definitions and metrics used for value potential index.

SOURCE: US Bureau of Labor Statistics; McKinsey Global Institute analysis

## 2) 특징

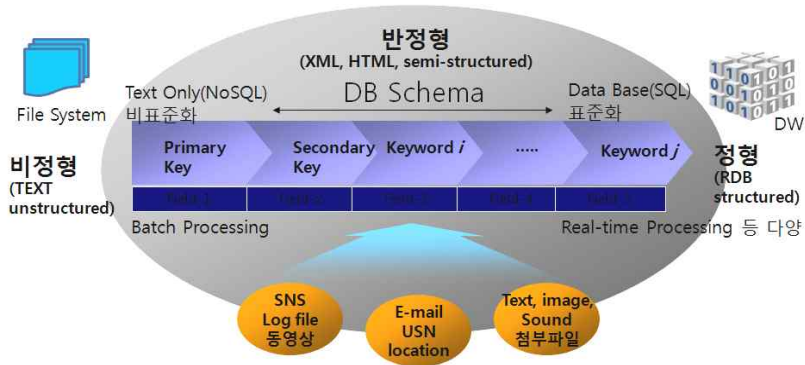
많은 기관들에서 빅데이터를 정의하였는데 특징을 정리하면 <표 1>과 같다.

〈표 1〉 빅데이터 정의 및 특징

기관구분	빅데이터 정의 및 특징
포레스트	<ul style="list-style-type: none"> <li>가치를 얻기 위한 데이터와 무엇을 할 것인지 아는 사람이 기업에게 필요하다는 것을 의미 있는 기술</li> <li>볼륨, 속도, 다양함, 다양성으로 현재의 기술로 감당 어려운 규모의 데이터</li> <li>경제적 가치를 창출하는 데이터</li> </ul>
SERI	<ul style="list-style-type: none"> <li>거대한 데이터 집합으로 대규모 데이터와 관련된 기술 및 도구포함</li> </ul>
가트너	<ul style="list-style-type: none"> <li>3V로 정의 : Volume, Variety, Complexity</li> <li>Volume : 데이터 규모가 엄청남을 의미</li> <li>Variety : 로그기록, 소셜, 위치정보 등 데이터의 종류가 증가로 텍스트와 멀티미디어 등 비정형화된 데이터의 유형이 다양화 되는 것 의미</li> <li>Complexity : 구조화되지 않은 데이터, 데이터 저장방식의 차이, 중복성 문제 등 데이터 종류가 확대되고 외부 데이터의 활용 등으로 관리대상이 증가됨으로써 점차적으로 데이터 관리 및 처리가 복잡화되고 심화되어 새로운 처리 및 관리기법이 요구되는 상황을 의미</li> </ul>
SAS	<ul style="list-style-type: none"> <li>4V로 정의 : Volume, Variety, Velocity, Value</li> <li>Volume, Variety는 가트너 정의와 동일</li> <li>Velocity : 센서나 모니터링 등 사물정보, 스트리밍 정보 등 실시간성 정보가 증가하고 있고, 이러한 실시간성으로 인한 데이터 생성, 이동과 유통의 속도가 증가하고 있으며 대규모 데이터처리 및 가치 있는 실시간성 정보활용을 위해 데이터 처리 및 분석 속도가 매우 중요하게 되었음을 의미</li> <li>Value : 새로운 가치를 창출하는 것을 의미</li> </ul>
노무라 연구소	<ul style="list-style-type: none"> <li>빅데이터를 처리할 수 있는 인재·조직, 데이터 처리·축적·분석기술, 데이터 자원 등을 빅데이터의 3요소로 정의</li> <li>3요소의 조화로운 발전이 데이터의 특성과 컴퓨팅 파워의 발달에 따라 실생활 적용이 빠르게 확산될 것으로 전망</li> </ul>

정형과 비정형은 다음 <그림 7>과 같이 DB 스키마라는 표준방식으로 정의하느냐 일반적 인 파일시스템 형태로 유지하느냐의 차이로 구분된다. 비정형으로 유지하는 이유는 데이터의 저장이 중요하지 않으나 단순한 경향파악이 주요 이유일 경우에 해당되며, 이 경우 일정시간 경과 후 소멸된다. 수많은 데이터를 저장기술을 통해 저장할 필요가 없거나 너무 생성속도가 빨라서 저장기술이 생성속도를 따라갈 수 없을 경우는 비정형 데이터로 유지될 수밖에 없다. 그러나 저장기술이 생성속도를 능가하는 시점이 도래하여서도 빅데이터의 저장여부가 이슈가 되는데 이것은 데이터의 필요성 문제에 관한 것으로 볼 수 있다.

〈그림 7〉 정형과 비정형 데이터의 차이



이와 같이 빅데이터는 IT의 흐름이 하드웨어, 소프트웨어에 이어 데이터로 이동하기 시작하는 정보형태의 변화, 정보관리 및 기대치의 변화에 원인이 있으며, 단순분석보다는 데이터의 의미와 가치에 더 많은 관심을 가지기 시작하면서 발전하기 시작하였음을 알 수 있다(한국정보화진흥원, 2013; 한국정보화진흥원, 2012b).

윤상오는 빅데이터의 장점 뿐 아니라 위험요인에 대한 유형분류를 제시하였으나(윤상오, 2013), 정보화는 위험요인에 대한 인식과 이에 대한 대처방안이 함께 개발되면서 더 나은 방향으로 발전할 수 있을 것이다.

## 2. 개념적 · 기술적 개념적 이슈

### 1) 개념적 이슈

정보화 사회는 매우 급변하게 변화하고 있고 과거 시스템 중심에서 사람중심으로 변화되고 있으며, 소셜네트워크의 파워가 매우 증가하고 있다. 이러한 시대에 빅데이터는 미래사회의 특성인 불확실성, 리스크, 스마트, 융합 등 미래사회의 특성에 대응하는 역할을 수행하며 기회요인을 창출하는 핵심엔진으로 작용하고 있다(정보통신산업진흥원, 2011; 한국정보화진흥원, 2012a; 한국정보화진흥원, 2012b; 한국정보화진흥원, 2012c).

먼저 불확실성에 대한 빅데이터의 역할은 통찰력을 제공하는데 있는데, 주로 사회현상과 현실세계의 데이터를 기반으로 한 패턴분석과 미래전망을 제공하고 있고, 여러 가지 가능성

에 대한 시나리오 시뮬레이션과 다각적인 상황이 고려된 통찰력을 제시하고 있으며 다수의 시나리오로 상황변화에 유연하게 대처하게 한다.

두 번째 특징인 리스크에 대해서는 대응력을 제공하고 있다. 빅데이터는 환경, 소셜, 모니터링 정보의 패턴분석을 통한 위험징후, 이상신호를 포착하고 이슈를 사전에 인지하여 분석하고 빠른 의사결정과 실시간 대응지원을 하고 있으며 기업과 국가경영의 투명성을 제고하여 낭비요소를 절감하게 한다(권대석, 2012).

세 번째 특징인 스마트에 대해서는 경쟁력을 제공하고 있다. 주로 대규모 데이터분석을 통한 상황인지, 인공지능 서비스 등이 기능하도록 하고 있으며 트렌드 변화분석을 통한 제품 경쟁력을 확보할 수 있도록 한다.

네 번째 특징인 융합에 대해서는 창조력을 제공하며, 인과관계 및 상관관계가 복잡한 컨버전스 분야의 데이터 분석으로 안정성을 향상시키고 시행착오를 최소화 할 수 있도록 하며, 방대한 데이터 활용을 통한 새로운 융합시장 창출이 가능하도록 한다.

## 2) 기술적 이슈

Gartner는 2011년 이머징 기술 hype cycle에서 빅데이터를 새롭게 포함시키고 앞으로 주목해야 할 기술로서 빅데이터를 기술발생단계(technology trigger)로 분류하여 전망하였다(한국정보화진흥원, 2012a; 한국정보화진흥원, 2012c).

그러나 빅데이터 플랫폼은 특성상 매우 복잡하기 때문에 NP-hard(Nondeterministic Polynomial-hard) 문제로 볼 수 있다. NP-hard는 특별히 다항식으로 식을 세울 수 없는 어려운 문제들의 집합이므로, 근사추론에 의해 일일이 답을 대입해야만 정답을 유추해야 하는 부류의 문제들이므로, NP-hard 문제에 대한 해결책으로 휴리스틱 접근방법인 인공지능 기술이 대두되고 있다. 빅데이터 플랫폼에 인공지능기술(한국정보화진흥원, 2010; 한국정보화진흥원, 2012c; 조영임, 2012)이 접목되면, 센싱정보로부터 학습, 추론, 인지 등의 기본적 인공지능 과정을 통해 분석결과를 해석하고 의사결정 할 수 있는 지능을 가진 인간과 유사한 시스템을 구축할 수 있는 것이다.

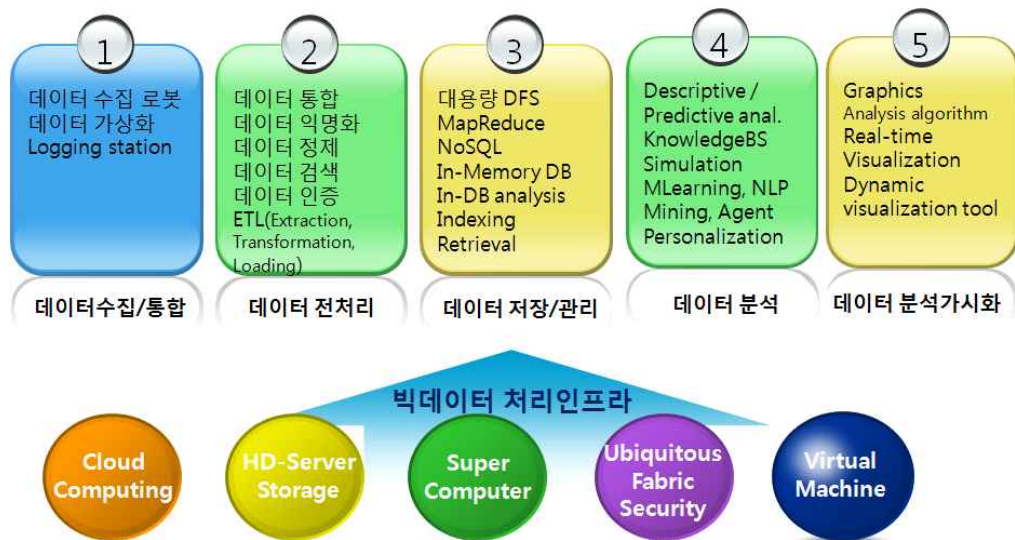
### (1) Human-like 빅데이터 플랫폼

인간과 유사한 human-like 기능의 빅데이터 플랫폼은 <그림 8>과 같이 제시할 수 있다.



즉, 빅데이터 처리인프라에 기반하여 데이터 수집 및 통합에서 데이터 전처리단계를 거쳐 데이터저장 및 관리기술로 이어지고 데이터 분석 및 데이터분석 가시화 단계로 구성된다. Human-like 빅데이터 플랫폼이라고 명명하는 이유는 5단계의 각 기술들에서 인간의 지능을 구현할 기술들이 적용되고 연구되기 때문이다.

〈그림 8〉 Human-like 빅데이터 플랫폼



#### ① 데이터 수집/통합 단계

데이터 수집 및 통합단계에서는 새로운 데이터 생성, 네트워크에 산재해 있는 외부데이터 수집, 내외부 이중데이터 통합 등 데이터의 형태와 소재에 무관하게 데이터를 확보하는 기술들을 말한다. 과거 기술적 가능성, 분석성능의 미비, 분석결과와 타당성 등을 이유로 등한시했던 데이터들을 빅데이터로 명명하고 이를 분석하면서 통계분석기법, 인공지능적 기법, machine learning 기법들을 총체적으로 적용하면서 연관성을 찾아 기업의 경쟁력 강화, 기업의 수익, 소비자 성향과악에 주력해야 한다. 따라서 수집로봇, 데이터 가상화, logging station 등의 기술이 데이터 수집 및 통합단계에서 사용되는 기술들로, 전처리 단계로 진입하기 전에 다양한 원천의 데이터를 확보하는 일이 첫 번째로 수행되는 기술들이 필요하다.

## ② 데이터 전처리 단계

데이터 전처리 단계에서는 첫 번째 단계에서 수집된 정보들에서 센싱정보, SNS 등 지속적으로 발생하는 비정형 스트림 데이터를 정제하여 분석 가능한 형태로 구조화하여 분석의 정확성을 높이고 심층분석을 가능하게 하는 기술들을 말한다. 데이터 통합, 익명화, 정제, 검색, 인증 및 ETL(Extraction, Transformation, Loading) 기술의 개발을 통해 다음 단계로의 진입을 원활하게 해야 한다.

정보검색기술의 목표는 ‘내가 원하는 것을 놓치지 않고 편한하고 빠르게 검색하는 것’이며, 정보검색의 요구사항을 해결하기 위한 관련된 기술로는 인공지능적 기법, 자연어처리, 쿼리 프로세싱, 재현율/정확율 평가방법과 컴퓨터 네트워크가 반드시 지원되어야 한다(조영임, 2013; Christopher D.와, 2009).

빅데이터 플랫폼의 두 번째 단계인 전처리 단계의 정확성 여부가 정보검색의 방향을 결정한다고 해도 과언이 아니므로 수집한 결과를 잘 분류하여 분석이 가능한 형태로 만드는 것이 매우 중요한 기술이 된다.

정보검색에서 문헌을  $d_j$ 라고 하고 색인어  $k_i$ 와의 연관정도라 하면, 문헌  $d_j$ 에서 색인어  $k_i$ 와의 연관도를 구하는 함수는  $g_i(d_j) = w_{ij}$ 에 의해,  $w_{ij}$  값이 높을수록 연관성이 높아서 검색 순위가 높게 나타날 것이다. 재현율(recall)과 정확율(precision)은 검색척도를 나타내는 개념으로 일반적인데, 두 번째 단계의 전처리기술의 정도에 따라 이 비율이 달라지게 되어 정보검색의 효과에 영향을 미치게 되므로 중요하다.

## ③ 데이터 저장/관리 단계

데이터 저장 및 관리 기술은 웹 데이터, 소셜미디어, 비즈니스 데이터, 센싱정보 등의 증가하는 다양한 형식의 데이터를 실시간으로 저장 및 관리할 수 있는 분산 컴퓨팅 기술을 말하는 것으로, 빅데이터 플랫폼의 핵심기술을 말한다. 빅데이터 저장기술인 NoSQL(Not Only SQL)은 RDB(Relational DB)를 넘는 빅데이터 저장을 위한 새로운 DB개념으로 다양한 형태의 인터페이스를 제공하고 있으며 비표준화 상태이고 종류로는 MongoDB, Cassandra, Hbase 등이 있다. 따라서 앞서 언급한 빅데이터 저장의 필요성과 저장기술과의 상관관계에 대해서는 세 번째 단계에서의 연구에 따라 또한 여러 가지 현상적 필요성에 따라 선택할 수 있을 것이다.

최근 많이 사용되는 하둡(Hadoop : High-Availability Distributed Object-Oriented Platform, <http://hadoop.apache.org>)은 데이터 관리단계에 해당되는 기술로서, 오픈소스이며 다음에 설명하는 여러 장점들로 인해 많이 응용되고 있다.

#### ④ 데이터 분석기술

데이터 분석기술은 빅데이터에 내재된 가치를 추출하기 위해 필요한 대규모 통계처리, 데이터 마이닝, 그래프 마이닝 등의 분석기술, machine learning, artificial intelligence 기술을 활용한 다양한 심층분석 기술로서, 빅데이터 플랫폼의 핵심 중의 핵심기술이다. 이 단계의 다양한 기술들의 개발 정도가 human-like 플랫폼을 구축하는 브레인의 지능정도에 해당되므로 이들의 개발정도와 플랫폼의 지능수준은 비례관계를 갖는다.

최근 이러한 이유로 인해 데이터 분석가 양성의 필요성이 대두되고 있다(한국정보화진흥원, 2015d). 구조화되지 않은 대규모 데이터 속에서 숨겨진 정보를 찾아내는 데이터 분석가는 스마트 시대에 최고의 인재를 말하며, 데이터를 관리하고 분석할 수 있는 인력의 중요성이 높아지고 있다. <그림 3>의 빅데이터 관심용어에도 데이터 분석가가 상위에 랭킹하고 있어서 이 분야의 인력양성이 중요함을 알 수 있다.

#### ⑤ 데이터 분석가시화기술

데이터 분석 가시화 기술은 비전문가가 데이터분석을 수행할 수 있는 환경을 제공하는 분석도구기술과 분석결과를 함축적으로 표시하고 직관적인 정보를 제공하는 인포그래픽스 기술을 말한다. ‘빅쿼리’는 2011년 11월 구글의 빅데이터 분석 솔루션 플랫폼으로, 오픈소스 데이터 분석 툴인 하둡을 활용해 신속히 분석하여 시각화할 수 있는 기술이다.

#### (2) 빅데이터 플랫폼 사례

최근 빅데이터 플랫폼 툴로는 WEKA<sup>4)</sup> 빅데이터 분석 툴과 IBM의 InfoSphere 등이 많이 사용되고 있으나 WEKA는 자바기반의 machine learning, 데이터 마이닝 소프트웨어로 데이터분석에 쉽게 사용할 수 있다.

예를 들면, 날씨와 운동과의 관계를 알고 싶을 경우, 날씨에 관련된 데이터는 날씨에 따라

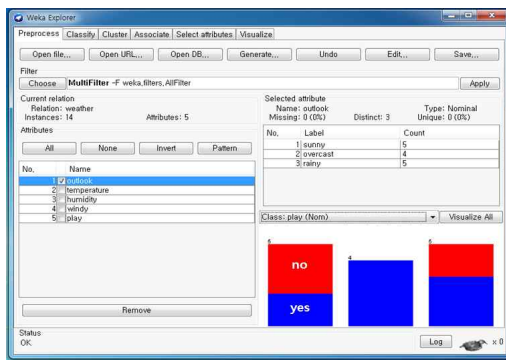
4) <http://www.cs.waikato.ac.nz/ml/weka/index.html>. 무료버전이며 사용법이 쉽기 때문에 지방정부 등에서 빅데이터 분석을 위해 엑셀자료가 쉽게 변환되므로 유의미한 결과를 도출할 수 있을 것이다.

play 여부를 기록한 데이터 즉 날씨의 어떠함에 따라 운동경기를 했는지, 안했는지의 과거 정보들을 기록해둔 데이터를 분석하면 어떤 날씨 조건에서 운동을 하는 것이 좋은가에 대한 유용한 지식을 얻을 수 있다. <그림 9(a)>는 outlook(조망)과 play 관계를 나타낸 그림으로, 파랑색은 play 속성이 'Yes' 값을 갖는 경우이고, 빨강색은 play 속성이 'No' 값을 갖는 경우를 구분한 것이다. 따라서 Outlook이 Sunny(맑음)인 경우는 안하는 경우가 약간 많고, Overcast(흐림)인 경우 100% 운동을 하고, Rainy(비움)인 경우도 경기하는 비율이 약간 높음을 볼 수 있다. <그림 9(b)>는 temperature와 play와의 관계를 나타낸 그림으로, 온도에 따라서는 크게 비율의 차이가 없는 것으로 나타난 그림이다.

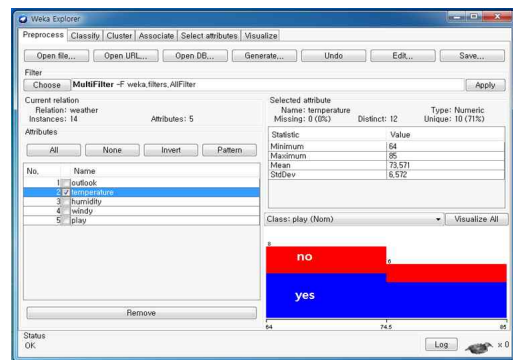
이와 같은 과정을 반복함으로써 <그림 9(c)>와 같은 의사결정트리를 얻을 수 있으며, 이로부터 운동 경기(play)에 영향을 주는 속성은 조망(outlook), 습도(humidity), 풍량(windy)

### <그림 9> WEKA를 활용한 데이터 분석 적용 사례

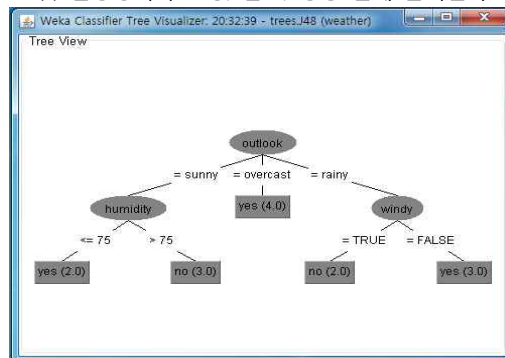
(a) Outlook과 play 관계



(b) Temperature와 play 관계



(c) 운동경기와 조망, 습도, 풍량 관계 분석결과



으로 분석되는데, 가장 중요한 속성은 조망으로 분석된다는 것을 알 수 있다.

앞으로 빅데이터 플랫폼을 지능화하여 융통성 있게 개발함으로써 인공지능과 빅데이터가 상호 시너지효과를 발휘할 수 있도록 하고 결과물의 신뢰성을 향상시켜야 한다. 기존 고객관리나 자료관리나 검색기술의 단순 확장 및 적용이 아니라 창의적 활동이 포함되는 human-like 플랫폼 개발이 필수적이기 때문이다.

### Ⅲ. 빅데이터 도입과 관련된 논쟁

#### 1. 우리정부의 빅데이터 도입과정

빅데이터 도입과 관련되어서는 우리 정부도 지난 몇 년간 정부의 데이터가 중요한 가치창출의 기반임을 인지하고 공공데이터 공개·개방을 지속적으로 추진하였다. 이러한 기반 마련과 함께 <표 2>와 같이 지난 2011년 10월에는 적극적인 빅데이터 활용을 통한 정부혁신과 국가경쟁력 제고를 위해 대통령 직속의 국가정보화전략위원회에서 ‘빅데이터를 활용한 스마트 정부 구현방안’을 마련하였다. ‘빅데이터를 활용한 스마트 정부 구현방안’에서는 공공데이터를 활용하여 범부처간, 정부·민간 융합지식을 도출하고 국내외 경제·사회·질병 등에 대한 실시간 분석·대응으로 고품질 서비스 국가를 실현한다는 것을 목표로 빅데이터 추진역량 강화 및 핵심기반확보를 위한 과제를 제안하고 있다. 이에 안전행정부(구 행정안전부)는 이를 구체화한 ‘스마트 정부 구현을 위한 빅데이터 마스터플랜’을 수립하였다. 정부 차원의 빅데이터 활용방안 외에 방송통신위원회에서는 ‘빅데이터 서비스 활성화 방안’을 마련하여 민간의 빅데이터 활용을 촉진하기 위한 계획을 발표하였다. 빅데이터 활용을 통해 기업과 국가의 경쟁력을 강화하고, 사회 현안 해결과 스마트 라이프 구현이 가능해질 것이라는 기대로 7대 과제를 제시한바 있다(한국정보화진흥원, 2012).

최근 박근혜 정부에서는 정부3.0을 발표하였는데, 정부 데이터를 공개하고 부처간 정보 칸막이를 없애는 정보 행정의 대혁신을 의미하는 것으로, 정부3.0의 키워드는 선제적 정보공개를 통한 개방, 공유, 소통이 핵심 가치이다(관계부처합동, 2013). 정부3.0의 3대 전략 중 ‘일 잘하는 유능한 정부’에는 4개의 중점과제들이 있는데 이 중 하나가 ‘빅데이터를 활용한 과학

〈표 2〉 빅데이터를 활용한 스마트정부 구현과제

추진과제	과제별 주요내용
범정부적 데이터 연계·분석체 계 구축	<ul style="list-style-type: none"> <li>• 기존 행정정보공동이용센터 등을 확대하여 부처 및 공공기관의 정보를 지속적으로 연계·수집할 수 있는 체계 수립</li> <li>• 예측기반의 국정운영 혁신을 위한 국가전방의 데이터 수집·분석 체계 마련</li> </ul>
정부·민간 데이터 융합추진	<ul style="list-style-type: none"> <li>• 소셜 미디어, 포털 데이터 등과 같은 민간 데이터와 공공데이터의 연계·활용을 위한 체계 및 기술확립</li> <li>• 산·학 등이 보유한 각종 정보를 범국가적으로 연계·수집할 수 있도록 저장소 구축·운영</li> <li>• 공공데이터의 단계적 개방으로 민간의 가치창출 및 기업활동 강화 지원</li> <li>• 민간기업의 공공데이터 접근에 따른 데이터 라이선스 규약을 신설하여 데이터 활용의 공익성을 확보</li> </ul>
공공데이터 진단체계 구축	<ul style="list-style-type: none"> <li>• 공공데이터의 효율적 관리를 위한 범정부적 MDM체계 구축MDM(Master Data Management): 다수의 시스템에서 활용되고 중앙에서 통제 관리되어야 하는 표준 참조 데이터</li> <li>• 범부처 데이터 공동관리를 위한 품질관리 기준 및 체계 마련</li> </ul>
법·제도 개선	<ul style="list-style-type: none"> <li>• 범국가적 빅데이터 활용 추진 기본계획 수립 및 관련 법령 개정</li> <li>• 공공부문 빅데이터 분석 서비스 활용촉진방안 및 성과관리 체계 확립</li> </ul>
분석인력양성 및 재교육	<ul style="list-style-type: none"> <li>• 빅데이터 시대 신 수요에 대응한 빅데이터 분석 전문인력 양성</li> <li>• 빅데이터 활용 역량강화를 위한 다양한 재교육 프로그램 제공</li> </ul>
개인정보 익명성 보정체계 확립	<ul style="list-style-type: none"> <li>• 안전하고 신뢰할 수 있는 공공데이터 공개 및 활용을 위해 개인정보와 프라이버시 보호 관련 기본원칙을 체계화</li> <li>• 데이터의 안전한 공유와 유통을 위한 강화된 보안대책 수립</li> <li>• 데이터의 개방·공유·활용에 따른 정부 공공데이터 활용 가이드라인 마련</li> </ul>
기술개발	<ul style="list-style-type: none"> <li>• 빅데이터 관련 인프라 기술개발</li> <li>• 빅데이터 운영·분석 기술개발</li> </ul>

(출처: 국가정보화전략위원회, 빅데이터를 활용한 스마트정부 구현(안), 2011)

행정의 실현’으로 중요한 개념이 되고 있다.

해외에서는 <표 3>과 같이 영국이 2012년 현재 Linked Data 형태로 8,400개의 데이터 세트를 제공하고 있으며 향후 의료, 교육, 세금, 고용, 기상 및 지리 정보 등에 대해 순차적으로 공개를 확대할 예정이다(한국정보화진흥원, 2012).

## 〈표 3〉 영국정부의 공개 데이터 목록 및 공개목적(2014.3까지 목표)

## (a) 공개데이터 목록

구분	공개 내용 및 데이터	최초 공개일	업데이트
고등교육 데이터	학생들의 고등교육 진학시 고려사항	2012.9	매년
기업등록소 데이터	기업 명, 기업 수, 등록주소, 분류상태 및 원문코드 무료 다운로드	2012.7	매년
국립지리원 데이터	국립지리원의 오픈데이터, 국립자연탐방로 데이터 세트	2013.4	계속

## (b) 향후 공개대상 데이터 내용 및 공개목적

구분	공개 내용 및 데이터
의료정보	일반 의료진의 실적정보를 공개하여 환자들이 병원간 환자치료 및 생존률 비교가능
세금 및 자금정보	시민사회 프로그램의 정보를 공개하여 보조금 및 세금지원여부를 확인할 수 있도록 하여 투명성 증대
고용정보	노동연금부의 워크 프로그램 및 고용유지 지원금의 정보를 공개하여 일자리 창출을 위한 정보제공

우리정부의 빅데이터는 이명박 정부에서 이슈가 되어 최근 박근혜정부에서 활성화 되기에 이르렀으며, 빅데이터 도입과 관련되어서는 기술적 개념적 이슈 등이 구체적으로 실현됨으로써 빅데이터 기술이나 응용이 가시화되고 실생활에 접목이 되어야 진정한 가치가 있을 것이다. 빅데이터와 관련되어서는 최근 여러 가지 측면에서 논쟁이 제시되고 있으나 본 장에서는 빅데이터 도입과 관련된 주요 논쟁관점 및 이를 바탕으로 다음 장에서는 지방정부에서 지역정보화 차원에서 빅데이터 시대 준비해야 할 이슈와 사항들에 대해 논의하고자 한다.

## 2. 빅데이터 도입관련 주요 쟁점

## 1) 플랫폼

빅데이터 도입과 관련된 첫 번째 가장 큰 이슈로는 기술적 이슈인 빅데이터 플랫폼에 대한 것이다. 이미 많은 연구자들이 빅데이터의 플랫폼에 대한 필요성을 제시하고 있고 관련 기술들과 응용사례들이 제시되고 있으나 구체적인 스마트사회에 적합한 인간과 유사한 human-like 플랫폼 분석틀이 제시되지는 않는 상황이다.

빅데이터 플랫폼은 민간보다는 공공분야에서 공공분야 정보의 공유를 위해 더 필요하다(한국정보화진흥원, 2012d; 관계부처합동, 2013). 최근 정부3.0에서는 빅데이터는 이러한 플랫폼을 통한 데이터개방을 필요로 하고 있다. 미국은 이미 데이터 생성부터 공개까지 효율적

인 플랫폼과 관리체계가 구축되어 가장 활발한 데이터 공개를 하고 있다. 미국은 원천데이터를 제공하는 서비스 형태가 sheet, chart, map, apps, open API이며, 통계와 지리정보 데이터를 제공하고 있고, link와 다운로드가 가능한 데이터 파일을 공개하고 있다. 영국은 미국에 비해 상대적으로 정보공개현황의 수준이 미흡하나 원천데이터를 중심으로 데이터를 공개하고 있는 점이 차이점이나 미국과 마찬가지로 데이터를 다운로드할 수 있는 데이터 파일형태로 제공하고 있다. 호주도 다운로드 할 수 있는 데이터 파일형태로 제공하고 있다. 그러나 우리나라는 원천데이터의 공개는 이루어지지 않고 있으며 open API 형태로 가공된 프로그램만을 제공하고 있으므로 개발된 데이터의 양이 절대적으로 부족하고, 공개를 위한 자원소모가 상대적으로 크며 이로 인해 공공정보의 활용율이 매우 낮다.

따라서 공공정보의 활용율을 높이고 빅데이터의 효율적 분석과 활용성을 위한 플랫폼의 기본구조와 기능들에 대한 연구가 필요하다. 따라서 빅데이터의 역할이 성공을 거두려면 공공과 민간부문이 통합된 데이터 분석을 위한 플랫폼 개발이 필요하며, 다학제적 이해와 통합적 사고와 직관력을 갖춘 데이터 분석가의 양성이 필요하고, 개인 프라이버시를 위한 기법의 도입과 데이터 자원의 결합과 협력촉진을 위한 신뢰기반형성이 필요하다.

## 2) 정보보안

두 번째 빅데이터 기술적 이슈로 정보보안을 들 수 있다. 정보보안은 빅데이터 도입 이전에도 정보시스템에서 자주 이슈화되는 문제이기도 하다. 기존에도 해결하지 못한 보안 문제를 빅데이터 도입 후 엄청난 양의 데이터를 어떻게 보안유지를 할 수 있을지가 여전히 큰 문제이다. 클라우드 컴퓨팅 환경과 마찬가지로 기업 데이터 및 경영전략에 대한 보안 우려가 빅데이터에서도 역시 화두이다.

정보보안은 개인정보보호와 시스템 보안으로 구분할 수 있다. 따라서 빅데이터에서의 정보보안은 시스템 보안 뿐 아니라 개인정보보호차원에서 철저한 연구가 이루어져야 한다. 만약 공공이든 민간이든 빅데이터 사용할 경우 개인정보보호가 제대로 이루어지지 않으면 심각한 문제를 발생하기 때문이다. 빅데이터는 개인정보뿐만 아니라 위치·의료기록·대출 정보 등이 담겨 있기 때문에 해킹으로 인해 정보가 유출될 경우 피해가 더욱 커질 수 있다. 방대한 데이터 수집과 분석, 활용이 기업의 경쟁력으로 연결되기 때문에 개인정보를 노린 범죄가 더 발생할 수 있다. 실제로 개인정보나 SNS 정보, 온라인 결제 및 GPS 정보, 블로그 등을 분석



하면 얼마든지 범죄에 악용될 수도 있기 때문이다.

그러나 이러한 비정형 데이터의 유형이 전체 데이터의 90%를 차지하고 있는 시점에서 빅데이터의 활용은 매우 중요한 이슈가 되고 있으므로 정보화의 역기능에 해당되는 정보보안문제를 최소화하는 방안을 마련하는 것이 필수적이다. 즉 데이터 암호화, 본인확인기관 검증, 모니터링 강화 등 빅데이터 시대에 맞는 보안체계를 만들어야 한다.

이미 정부3.0에서는 공공정보 개방·공유 및 개인 맞춤형 서비스 확대에 따른 개인정보 침해 요소에 대한 선제적 보호조치 강화 및 안전한 활용 기반을 마련함으로써 정부3.0 추진 기반을 마련하고 개인정보보호 안심사회를 구현하기 위해 준비 중이나 빅데이터 시대에 맞는 보안체계의 수립이 필요하다.

빅데이터 시대 정보보안은 이러한 기술적 이슈 뿐 아니라, 빅데이터 플랫폼 첫단계인 데이터 수집단계에서의 정보수집과 이용과정의 투명성과 정보주체의 명확성을 확립해야 하는 것도 중요하다. 특히 빅데이터가 범죄 등에 악용되지 않도록 법과 제도를 통해 철저하게 관리해야 한다.

### 3) 개인사생활침해

세 번째 이슈로는 정보보안과 같은 맥락이나 개인사생활침해 문제를 들 수 있다. 스마트 시대 사용자들 입장에서 보면 개인사생활침해 문제를 가장 심각하게 받아들이고 있다. 개인정보가 자신도 모르는 사이에 노출되므로 공공이나 민간 입장에서는 정보보안이나 해킹을 염려하였지만, 개인 입장에서는 사생활침해문제가 가장 심각하므로 개인의 사생활이 보호되는 제도적 기술적 개발이 필수적으로 마련되어야 한다.

### 4) 정보의 범위와 깊이

네 번째 이슈로는 빅데이터 플랫폼의 단계별로 이슈화되는 문제에서도 제시하였듯이 어떠한 정보의 수집을 통해 어떠한 형태로 빅데이터 정보가 가공되어 제공하느냐 문제와 어느 범위까지 제공할 수 있느냐 즉, 정보의 범위(coverage)와 정보의 깊이(depth) 관계를 고려해야 한다. 실제로 공공이나 민간 등 빅데이터를 활용하려는 기관에서 필요로 하는 데이터를 적시에 제공하기란 쉽지 않다. 이 문제 역시 앞서 설명한 NP-hard에 속하는 문제이다. 따라서 학습과 추론을 통한 지능형 시스템 개발을 통해 해답을 제안해 나가야 한다.

## Ⅳ. 빅데이터의 지역정보화 발전을 위한 시사점 및 대응방안

최근 안전행정부는 지식행정 활성화를 위해 정보를 공유하고, 협업을 장려하는 것을 주요 내용으로 한 지식행정 도입으로 부처 간 칸막이를 낮추려는 노력을 하고 있으며, 정부가 보유한 공공정보의 개방과 공유를 확대하고 부처간 칸막이를 없애며 중앙과 지방, 정부와 국민간 소통과 협업을 통해 국가 현안을 해결할 수 있는 시스템을 정착하기위해 노력중이다.

정부3.0의 취지가 아니더라도 공공정보와 민간정보의 융합을 통한 정보공유체제를 갖추어서 활용가치를 높여야 하는 것은 최근 빅데이터 추세에 부합하는 일일 것이며, 가치를 추구하는 개방형 스마트 정부에서 마땅히 해야 할 일일 것이다. 따라서 공공과 민간의 정보공유는 향후 정보화 시대에 추진되어야 할 필수적 사항이 되고 있다.

따라서 지방정부에서는 지역정보화를 통해 빅데이터 실현을 위해 노력해야 할 것이다. 서울시와 인천시의 경우 몇 가지 계획을 발표함으로써 지방정부의 빅데이터 실현에 앞장서고 있어서 사례를 제시하고자 한다.

먼저 서울시의 경우<sup>5)</sup> 공공과 민간분야의 빅데이터 공유를 통한 새로운 서비스 발굴을 위해, 2013년 4월 이동통신 기지국 통화량 분석을 통해 인구가 밀집된 지역을 중심으로 심야 시간 버스 노선이 증설할 계획을 수립하고 KT와 공공서비스 개선 협약을 체결하였으며 향후 공공과 민간의 데이터를 활용해 공공서비스 발굴에 노력할 계획이라고 한다. 우선 통신 빅데이터 정보와 공공 교통데이터를 분석해 심야버스 노선을 효율적으로 배치하게 되는데, 이로 인해 교통약자의 안전 귀가를 도모하고 상대적으로 소득이 낮은 심야 경제활동 인구의 교통비를 절감할 수 있을 것이다. 서울시의 사례는 공공과 민간과의 빅데이터 융합, 분석 결과를 시 정책에 반영하는 최초의 사례가 될 것이며, 앞으로도 공공 분야 빅데이터의 주체로 시민에게 보다 유용한 공공데이터를 개방하는 것이 정부3.0의 취지는 물론 빅데이터의 활용 추세에 맞는 일일 것이다.

최근 인천시 등에서도 교통에 빅데이터를 이용하려는 계획<sup>6)</sup>을 발표하는 등 공공과 민간 분야의 융합을 통한 빅데이터 활용이 가속화 될 것으로 보인다.

그러나 지방정부에서 활용된 사례들을 보면, III장에서 빅데이터 도입과 관련되어 주요 쟁

5) <http://www.seoul.go.kr/main/index.html>

6) [http://article.joinsmsn.com/news/article/article.asp?total\\_id=10465948&ctg=1213](http://article.joinsmsn.com/news/article/article.asp?total_id=10465948&ctg=1213)

점으로 논의된 플랫폼, 정보보안, 개인사생활침해, 정보의 깊이와 범위 등에서 구체적인 안이 제시되지 않고 있으며 단순한 응용에 그치는 경향이 있다. 결론적으로 빅데이터는 국가는 물론 지역정보화의 발전에 중요한 역할을 하고 있음에도 불구하고 아직까지 지방정부에서는 빅데이터에 대한 이해와 활용성에 대한 분석이나 연구가 다소 낮은 것으로 보인다. 따라서 이를 해결하기 위한 몇 가지 시사점과 대응방안을 제시하면 다음과 같다.

첫째, 지방정부에서 빅데이터를 활용하여 지역정보화를 달성하려면 공공과 민간의 융합 영역의 창조가 필수적이다. 이는 이미 앞의 서울시와 인천시의 경우를 통해서 제시한 바와 같다. 정보공유 차원에서 빅데이터를 통한 공공과 민간 융합모델의 창조와 이를 통한 새로운 데이터나 서비스모델의 발굴이 필요하다. 단순한 정보의 제공은 또 다른 문제를 야기시킬 수 있으므로 다양한 형태의 공공 민간 융합 모델에 관해 참여하는 범위와 제공되는 정보의 양, 활용처, 주요대상 등 면밀한 분석을 통한 융합 모델이 제시되어야 한다.

둘째, 지역정보화 관점에서 민간분야의 공공화가 필요하다. 즉, 그동안 공공분야에서는 다소 미흡하였으나 민간분야의 전문영역이었던 부분을 공공분야에서 활용할 수 있도록 제도화하는 것이 필요하다.

예를 들면, 민간분야의 취업사이트인 경우 취업정보를 확인할 뿐 아니라 취업정보 이외에 수많은 정보를 제공하기 위한 전문 영역으로 이용자가 증가하고 있으나 공공기관의 취업사이트인 경우 단순 정보제공에 그치는 경향이 많으므로 실제로 이용자 관점에서 보면 유용하지 못한 경우가 많다. 그러나 공공에 비해 민간분야에서는 최근 빅데이터 분석에 의한 서비스를 통해 취업사이트 내에 소셜기능을 적극 도입된 형태로 변화하고 있는데, 실제로 인크루트<sup>7)</sup>에서는 기 개발한 인맥 서비스를 보완하고 이어 인맥끼리 이력서에 첨부하는 ‘추천글’ 서비스를 도입, 소셜을 통한 이력서 신뢰도를 향상시키고 있다. 즉, 관계를 통해 취업의 질과 양의 향상을 동시에 향상시키고 있다. 빅데이터로 부터 합격가능채용정보를 제공하기도 한다. 합격가능채용정보는 기존의 맞춤형채용정보와 다른데, 구직자가 미리 조건을 설정하지 않아도 구직자의 데이터를 분석하고 구직자의 지원 행태와 조건을 동시에 고려하여 선별된 채용정보를 생성함으로써 구직자에게 가장 적합한 취업정보를 제공할 수 있다는 점에서 차별화 된다.

따라서 이러한 민간분야에서 활성화되는 사이트를 공공분야의 취업정보와 연계하여 구축하고 제도화함으로써 상호발전을 꾀하는 것은 물론 지역정보화를 활성화하는데 필요할 것이다.

7) [http://job.incruit.com/jobdb\\_list](http://job.incruit.com/jobdb_list)

셋째, 지방정부에서는 빅데이터 마스터 플랜 및 플랫폼 계획을 먼저 수립해야 하며 정부는 법적근거를 마련해야 한다. 이는 앞서서 빅데이터 플랫폼에서도 제시하였으나 정보보안이나 사생활침해 등의 문제를 적극적으로 해결하면서 전자지방정부의 특성을 살린 특성화된 전자지방정부 빅데이터 플랫폼 구축을 위해 노력해야 할 것이다. 국토해양부에서는 2010년부터 지방정부에서 유비쿼터스 도시 구축을 위해 계획을 수립하고 정부의 승인을 얻어야만 유비쿼터스 도시구축이 가능하도록 법적근거를 마련하였다. 이로서 유비쿼터스 도시를 구축하려는 지방정부들은 반드시 계획수립 후 승인을 얻어야만 구축이 가능하게 되었다. 따라서 향후 빅데이터 활용을 위해서는 이러한 법적근거가 마련되어야 할 것이므로 법을 개정하거나 신규제정이 필요하다.

지방정부에서 빅데이터 활용을 위한 마스터 플랜 수립은 지방정부에서 빅데이터를 활용하여 어떠한 효과를 기대할 수 있는지를 미리 계획하는데 효과적일 것이다.

넷째, 작은 일에도 관심을 갖고 빅데이터 활용을 강화해야 한다. 마이클 래빈의 ‘깨진 유리창 법칙’을 보면 범죄충동은 사람의 개성보다는 환경과 상황에 따른다는 것을 알 수 있다 (마이클래빈, 2006). 이 법칙<sup>8)</sup>이 유명해진 것은 뉴욕의 줄리아니 시장이 범죄와의 전쟁을 선포하면서 이 이론을 적용해 큰 효과를 보면서였다고 하는데, 1994년 선출된 이후 살인이나 강력범죄가 아닌 지하철 낙서와 같은 ‘사소한’ 범죄 근절에 우선 앞장서는 전략을 택했고, GIS를 활용해 살인 등 강력사건이 자주 발생하는 지역에 집중적으로 가로등, CCTV를 설치하고 깨진 창문을 모두 없애고 순찰을 강화한 결과 살인건수가 60%이상 줄었다고 한다.

우리나라도 이와 유사한 사례들을 많이 볼 수 있는데, 지방정부의 지역정보화를 위해서는 빅데이터를 활용하여 미리예측하고 대응함으로써 성공한 사례를 계속 개발하고 작은 일에도 관심을 기울이는 것이 필요하다.

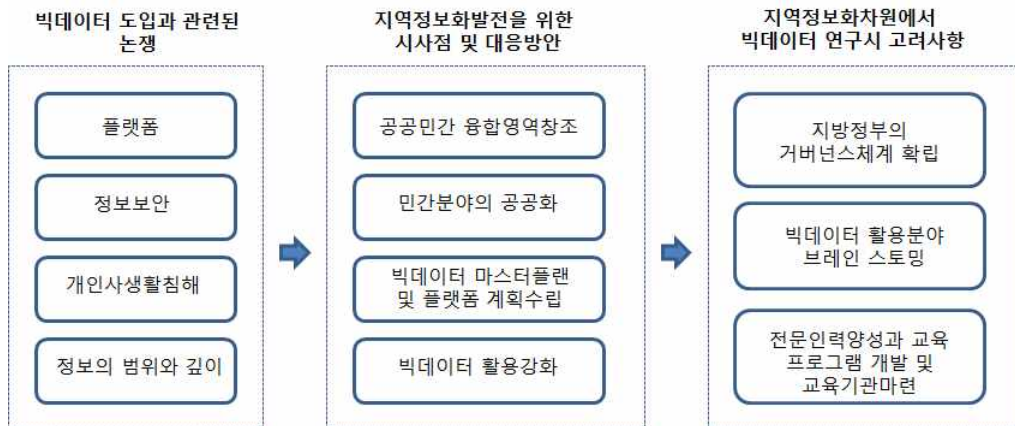
## V. 지역정보화차원에서 빅데이터 연구시 고려사항

III장의 빅데이터 도입과 관련된 논쟁과 IV장의 지역정보화발전을 위한 시사점 및 대응방안은 연관성이 있는 내용으로 지역정보화 차원에서 모두 고려되어야 하는 사항이므로, 이러

8) 깨진 유리창 법칙은 깨진 유리창처럼 사소한 것들이 사람들에게 중요한 메시지를 전달한다고 강조한다. 깨진 유리창을 방치하는 건물은 분명 보안도 허술하고 관리도 제대로 되지 않을 것이라고 생각한다는 것이다.

한 전제하에 지역정보화발전 차원에서 빅데이터 연구시 추가적으로 고려해야 하는 사항을 <그림 10>의 흐름과 같이 제시하여 논의하고자 한다.

〈그림 10〉 빅데이터 도입관련 논쟁 및 지역정보화 발전과의 관계



먼저, 지방정부의 거버넌스 체계 확립이 먼저 선행되어야 한다. 거버넌스 체계(양순애, 2013)가 확립되지 않은 상태에서의 플랫폼과 융합모델은 활성화가 되지 않음은 그동안 이미 여러 경우를 통해 입증되었기 때문이다.

둘째, 다양한 사례에 적용할 수 있도록 하는 빅데이터 활용분야를 브레인스토밍 해야한다. 공공이든 민간이든 여러 분야의 사용자들로부터 니즈를 분석하여 니즈의 클러스터링화를 계속적으로 반복함으로써 bottom-up 방식의 빅데이터 활용분야를 도출해야 한다. 그동안 공공분야에서 top-down 방식의 정보제공이 있었다면 빅데이터는 bottom-up 방식의 접근방법을 통해 주민들에게 직접적으로 다가가는 것이 필요할 것이다. 이것이 또한 공공민간 융합모델의 기본 틀이 되어야 할 것이다.

이미 빅데이터를 통해 정책에 활용하여 성공한 사례는 무수히 많이 제시되고 있다(권대석, 2012). 한국 SAS의 한 예를 들면 하루 1,000만 달러를 생산하는 원유 시추선이 문제가 발생하였을 경우 대처하기 위해서는 평균 5일정도가 소요된다고 한다. 이 경우 약 5000만불의 손해를 입게 되는 것은 당연하다. 그러나 빅데이터 분석을 통해 미리 문제를 예측하고 대처함으로써 이만큼의 손해를 미리 예방하였다고 한다. 따라서 정지일 80%감소, 생산량 5% 증가, 운영비용 매년 7억달러 감소하였음은 잘 알려진 사례이다.

오바마 대통령은 당선되기 2년전부터 66,000번 모의선거를 통해 디너파티 초청장, TV광고 등의 결정에 세심하게 활용하여 와서 당선되었으나 롬니는 빅데이터에 대한 인식이 낮았기 때문에 선거에서 패배하였다고 한다. 또한 구글트랜드에서는 우리나라 7대 대통령 선거에서 이미 당시 후보인 박근혜와 문재인 후보에 대해 52.5 : 47.5로 박근혜 후보의 당선을 예측하였고 실제로도 51.6 : 48로 박근혜 후보가 당선되었다.

셋째, 전문인력양성과 교육프로그램 개발과 교육기관마련이 절실하다. 데이터 분석가나 홍보인력의 필요성은 앞서 바와 같으나 빅데이터 분석이나 우수한 빅데이터 활용사례 홍보 등을 체계적으로 수행할 인력양성을 위한 교육프로그램과 교육기관의 마련이 필요하다. 이것은 어느 대학이나 어느 기관에 맡기는데 그칠 것이 아니라 체계적인 평가와 활용을 통해 구체화되고 실천될 수 있도록 하는 방안 마련이 필요하다.

앞의<그림 4>와 <그림 5>를 통해 정부의 데이터에는 다양한 형태의 데이터가 모여 있고, 또 가치창출 면에서 보면 다른 분야보다 매우 높음을 알 수 있었다. 따라서 정부의 효율적이고 과학적인 운영을 위해서는 데이터 분석가의 양성과 활용이 매우 중요한 일일 것이다.

스마트 사회에서 빅데이터는 필수불가결하게 발생하는 요소이며, 이것을 어떻게 잘 활용하느냐가 스마트사회의 성패를 좌우한다고 해도 과언이 아니다. 우리나라가 자랑하는 전자정부도 빅데이터의 활용으로부터 효율성을 높일 수 있도록 패키지화 하는 것이 향후 지속적인 경쟁우위를 점하는 요소가 될 것이다. 이제는 버리는 데이터가 하나도 없이 다 모아서 분석하여 유의미한 지식을 생산함으로써 가치창출을 이루는 정보화 사회를 이루어나가야 할 때가 도래한 것이다.

전자정부에서는 수많은 데이터들이 쌓이고 버려진다. 앞으로는 어떠한 거버넌스 체계에서 어떠한 형태의 빅데이터를 얼마나 잘 활용하고 운영하느냐가 데이터가 중심인 시대로 도래하면서 경쟁력을 갖추는 핵심요소가 될 것임은 너무나도 분명하다.

## 참고문헌

- 관계부처 합동. (2013). 「정부3.0」 추진 기본계획. 안전행정부.
- 권대석. (2012). 클라우드와 슈퍼컴퓨팅이 이끄는 미래 : 빅데이터 혁명. 21세기북스.
- 마이클래빈. (2006). 깨진 유리창 법칙. 흐름출판.

- 서진완. (2004). 전자지방정부 구현을 위한 모델개발과 적용. 교수논총.
- 양순애. (2013). 지역정보화의 효율적인 추진을 위한 IT 거버넌스 발전방안. 한국지역정보화학회지. 16(1) : 77-96.
- 양혜영. (2012). 빅데이터를 활용한 기술기획 방법론. Korea Institute of Science & Technology Evaluation and Planning.
- 윤상오. (2013). 빅데이터의 위험유형 분류에 관한 연구. 한국지역정보화학회지. 16(2) : 93-122.
- 정보통신산업진흥원. (2011). 빅데이터 확산에 따른 도전과 기회. 최신 IT동향.
- 조영임. (2012). 최신인공지능시스템. 홍릉과학출판사.
- 한국정보화진흥원. (2010). 모바일 시대를 넘어 AI 시대로. IT&Future Strategy.
- 한국정보화진흥원. (2011). 신가치창출 엔진, 빅데이터의 새로운 가능성과 대응전략. IT&Future Strategy.
- 한국정보화진흥원. (2012a). 빅데이터로 진화하는 세상. 빅데이터 연구센터. 2012.
- 한국정보화진흥원. (2012b). 빅데이터 시대 : 효과적인 공공정보 개방을 위한 데이터 플랫폼 구축 방향. IT&Future Strategy. 2012.
- 한국정보화진흥원. (2012c). 빅데이터 시대. AI의 새로운 의미와 가치. IT&Future Strategy. 2012.
- 한국정보화진흥원. (2012d). 빅데이터 시대의 인재, 데이터 사이언티스트의 역할과 가능성. IT&Future Strategy. 2012.
- 한국정보화진흥원. (2012e). 2012 국가정보화백서, 빅데이터 시대의 국가발전 전략. 2012.
- 한국정보화진흥원. (2013). 새로운 미래를 여는 빅데이터 시대.
- 행정안전부. (2012). 주요정책 과제 추진실적 보고서.
- Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, (2009), "An Introduction to Information Retrieval", Cambridge University Press, England.

**조영임(曹永任)**: 1988년 고려대학교 컴퓨터학과 졸업. 1990년 동 대학원 석사. 1994년 동 대학원 박사. 2000년 Univ. of Massachusetts at Amherst, post-doc. 1996년 삼성전자 선임연구원. 현 수원대학교 컴퓨터학과 교수, 현 정부3.0 자문단 및 행정안전부와 소방방재청 자체평가위원, 현 Purdue대 초빙교수, 전 대학산업기술지원단 단장, 전 국가정보화전략위원회 실무위원, 관심 분야는 유비쿼터스 시스템, 인공지능, 정보검색, 빅데이터, 전자정부 등(ycho@suwon.ac.kr).

<논문접수일: 2013. 9. 2 / 게재확정일: 2013. 9. 22>