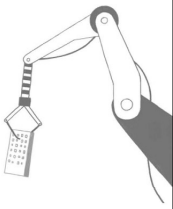
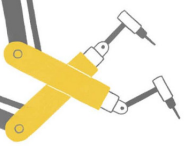


빅데이터

김수연 · 도지훈 · 김보라





Contents

 제1장 개요	3
 제2장 기술동향	6
 제3장 산업동향	13
 제4장 정책동향	17
 제5장 R&D투자 동향	21
 제6장 결론	25

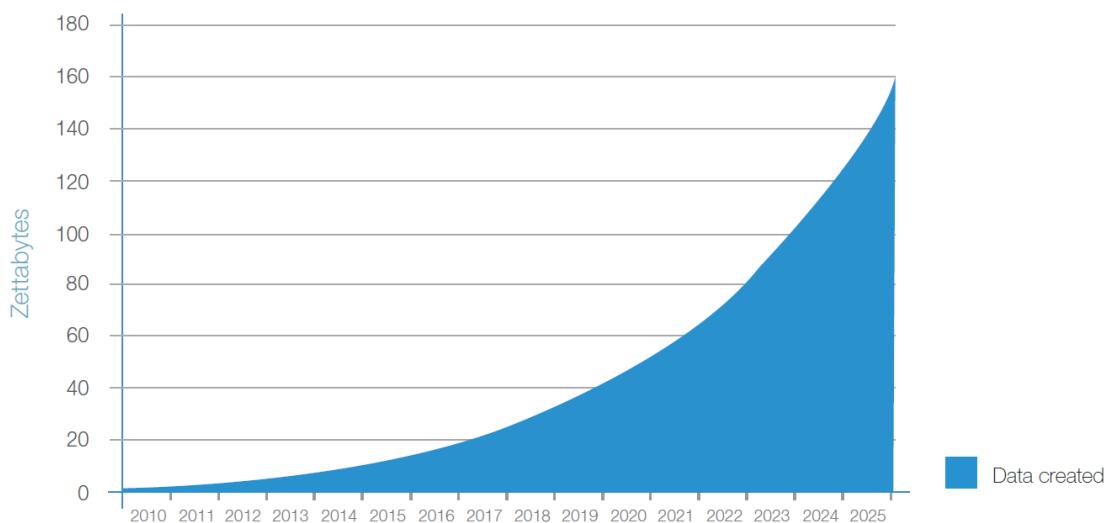


제1장 개요

1.1 작성배경

빅데이터 기술은 초연결 사회, 4차 산업혁명 등 기존 사회에 획기적인 변화를 가져오는 기술의 진보를 위한 기반 기술로 관련 수요가 증가하고 있음

- 모바일 기기의 확산 및 SNS의 활성화 등으로 전 세계 데이터량이 증가하기 시작하였고, 사물인터넷의 발달로 데이터량은 기하급수적으로 증가할 것으로 예상되고 있어 데이터로부터 새로운 가치와 혁신을 창출하기 위한 관련기술이 중요하게 부상하고 있음



출처 : Reinsel et al.(2017)

[그림 1] 연도별 세계 데이터 예측 규모

- 2016년 다보스포럼에서 논의한 4차 산업혁명을 견인할 기술로 제시하고 있는 인공지능, 로봇 공학, 사물인터넷, 자율주행차 등은 빅데이터 기술과 밀접하게 연관되어 있으며, 가트너(Gartner)는 빅데이터가 21세기 원유라고 발표하여 빅데이터의 중요성을 강조함
- 엔비디아(NVIDIA) CEO 젠슨 황은 2018년 GPU Technology Conference 키노트 스피치에서 '데이터가 곧 소스코드'라고 언급하며 데이터 중심으로 컴퓨팅 기술의 패러다임 전환을 시사함

최근 우리나라 정부는 지능화 기반 구축을 위해 빅데이터 기술을 혁신성장동력으로 지정¹⁾하고, 2022년까지 분석·예측의 정밀도 향상, 양질의 전문 빅데이터 구축을 목표로 지원 방안을 마련하여 해당분야 조기 상용화 여건을 조성 중임

- 금융, 통신, 교통 등 분야별 전문 빅데이터 구축과 데이터 개방·유통 및 연계·활용을 촉진하고, 현장에서 활용 가능한 실증·시범 사업을 추진하여 성공사례를 확산하는 등의 지원방안 마련

본 브리프에서는 혁신성장동력 시행계획²⁾에 제시된 빅데이터 분야 세부기술을 중심으로 기술, 산업 및 정책 동향을 분석하여 향후 정부 정책 및 R&D 투자방향 마련에 참고할 수 있도록 시사점을 도출하고자 함

1.2 기술 정의 및 범위

빅데이터는 기존의 통상적으로 사용되는 데이터 수집, 관리 및 처리 SW의 수용 한계를 넘어서는 거대한 규모의 데이터로 양(Volume), 속도(Velocity) 및 다양성(Varity) 측면에서 아래와 같이 전통적인 데이터와 구분되는 정보 자산을 의미함

〈표 1〉 전통적 데이터와 빅데이터의 특성 비교

구분	전통적 데이터	빅데이터
양	기가바이트(GB)	테라바이트(TB), 페타바이트(PB)
생산 주기(속도)	시간, 일	실시간
구조	구조적	반구조 및 비정형
데이터 원천	중앙집중	분산
데이터 통합	쉬움	어려움
데이터 저장	RDBMS ³⁾	HDFS ⁴⁾ , NoSQL ⁵⁾


1) 관계부처 합동, 혁신성장동력 추진계획, 2017.12.22.

2) 관계부처 합동, 혁신성장동력 시행계획, 2018.5.28.

3) 관계형 데이터베이스 관리시스템(Relational Database Management System) : 모든 데이터가 테이블에 저장된 관계 모델에 의해 저장 관리되는 데이터베이스 시스템으로 전통적인 대부분의 데이터베이스는 이 모델을 기반으로 함

4) 하둡 분산 파일 시스템(Hadoop Distributed File System) : 범용 하드웨어에서 실행되도록 설계된 분산 파일 시스템으로 내결함성이 높으며 저가의 하드웨어에 배치되도록 설계되고, 응용 프로그램 데이터에 대한 높은 처리량의 액세스를 제공하여 대용량 데이터셋 응용 프로그램에 적합

5) '데이터를 저장하는 방식이 SQL만 있는 것은 아니다(Not Only SQL)'라는 뜻으로, 클러스터에서 실행할 목적으로 만들어져 전통적 데이터베이스시스템인 RDBMS에서 사용하는 관계형 모델을 사용하지 않고, 스키마 없이 동작하며 구조에 대한 정의를 변경할 필요없이 DB 레코드에 필드를 자유롭게 추가할 수 있음

 **혁신성장동력 시행계획에 제시된 빅데이터 기술의 범위는 데이터 수집·저장·처리 등의 플랫폼기술, 이와 연계된 분석기술 및 새로운 통찰력과 비즈니스 가치를 창출하는 활용기술을 포괄하고 있음**

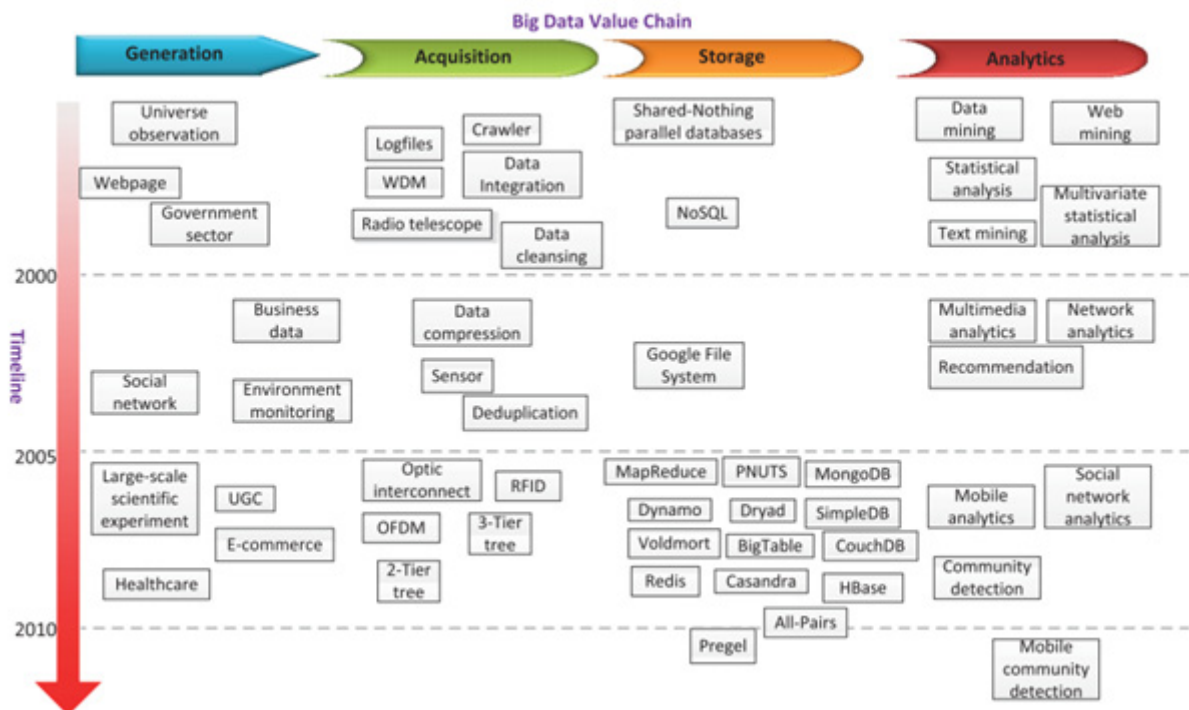
〈표 2〉 혁신성장동력 빅데이터 분야 기술 범위

기술분류	핵심기술	기술정의
플랫폼	데이터 자가 증식 및 수집·정제 기술	데이터의 양적인 확대를 위하여 데이터 증식 알고리즘을 활용한 자가 증식을 하거나, 유효하지 않은 데이터를 필터링하거나 샘플링, 정제, 수집하는 기술
	다양한 응용 패턴 통합 지원 기술	데이터가 실제 사용되는 시점에 데이터 사용 목적에 따른 데이터 모델에 맞추어 실시간으로 데이터를 구성하여 제공하고, 다양한 응용 패턴 (배치, 대화형, 스트림 등)을 통합하여 동시 수행을 지원하는 멀티 타입 빅데이터 처리 프레임워크
	멀티모델 데이터 통합, 고신뢰 데이터 관리 및 다각도 분석 기술	분석 목적에 맞게 다양한 모델의 데이터를 통합하고, 데이터의 신뢰성을 확보하면서 통계적으로 중요도를 갖는 결과의 자동 탐색 및 실시간 다각도 분석 기술
	초연결 데이터 관리 및 협업 기술	초연결 인공지능 구현을 위하여 물리적인 데이터 위치나 종류와 무관하게 데이터를 제공할 수 있는 초연결 데이터를 관리하고, 최적의 분석 결과 도출을 위한 다수의 다양한 지능 객체간 집단 협업 지능 플랫폼 기술
	빅데이터 처리 및 저장·관리 기술	질의 처리 성능 향상을 위해 성능 가속 HW, 통신 가속, 인메모리 컴퓨팅 기술을 활용하는 고속 빅데이터 처리 및 저장·관리 기술
분석	지능형 예측 분석 기술	데이터에 숨겨진 패턴을 찾아 과거와 현재의 상황 이해를 바탕으로 미래상황을 예측함으로써, 선제적인 의사결정을 지원
	이종 소스 심층 융합 분석 기술	비정형 텍스트, 관계형 DB 저장 데이터와 더불어 이미지/비디오 및 IoT 스트림 데이터 등 복합형 데이터를 대상으로 통합 분석
	엣지 분석 및 협업 분석 기술	초연결 시대에 발생하는 패스트 데이터에 대한 엣지 분석과 영역별로 산재하는 다수의 엣지 분석 플랫폼들이 연계하여 하나의 글로벌 문제를 분석하고 해결하는 분산/협업형 데이터 분석 기술
	모사현실 모델링 프레임워크	복잡한 실세계를 모사현실로 구현하는 대규모 개방형 모델링 프레임워크 및 최적화 기술
활용	빅데이터 유통 플랫폼 기술	공공·민간의 자유롭고 편리한 데이터 등록, 검색, 활용을 지원하는 플랫폼과 데이터 익명화와 같은 개인정보 보안성을 제공하는 빅데이터 유통 인프라 구축
	워크플로우 기반 적용 시나리오 구현기술	빅데이터 플랫폼의 적용 범위 확대를 위해 응용 분야별 특화된 적용 시나리오를 워크플로우 기반으로 제공함으로써 빅데이터 플랫폼 및 분석 기술의 활용성 제고
	데이터 품질 정량화 및 최적화 기술	데이터 가치 향상(Value-up)을 위한 데이터의 체계적 축적 및 지속적 관리 체계를 구축하는 데이터 라이프 사이클 관리 기술과 데이터의 품질 진단 및 개선 기술
	빅데이터 응용·서비스 기술	누적된 데이터 또는 실시간 데이터를 발생시키는 다양한 산업분야(의료·건강, 소비·거래, 에너지, 재난안전 분야 등)의 Domain Knowledge와 융합하여 빅데이터 플랫폼 및 분석 기술을 적용, 활용하는 응용·서비스 기술

제2장 기술동향

2000년대 이후 플랫폼 및 분석과 관련된 다양한 기술이 개발되었고, 2010년 이후 빅데이터 기술이 주목을 받기 시작함


- 2000년대 초반 일부 기업이 내부 데이터 분석을 통한 고객관리를 시도한 이후(김광섭, 2016) 빅데이터의 생성, 수집, 저장 및 분석 관련 기술이 개발되고 있음(Hu et al., 2014)
 - 생성단계 기술은 데이터 소스를 다양화하기 위해 점차 데이터의 비정형성을 고려할 수 있는 기술로 발전하고, 수집단계 기술은 수집, 사전처리 및 전송 관련 기술이 각각 다른 시기에 개발됨을 알 수 있음
 - 저장 단계는 대부분의 연구가 2005년 이후 시작되었으며, 기본적인 분석기술은 2000년 이전에 만들어지고, 이후에는 특정 도메인 문제를 해결하기 위한 시도들이 이루어짐



출처 : Hu et al.(2014)


[그림 2] 빅데이터 기술 발전 추세

- 가트너(Gartner)는 2011년 빅데이터 기술을 기술촉발(Technology Trigger) 단계⁶⁾로 제시하였고(Fenn and LeHong, 2011), 2012년 세계경제포럼에서도 새롭게 주목해야할 기술(Emerging Technology)로 선정함(WEF, 2012)

 **현재 빅데이터 기술은 전 산업에 영향을 미치는 기반 기술로 자리매김하고 있으며, 지능형 분석을 통한 ‘변화 예측형 빅데이터’ 기술로 발전하고 있음**

- 2015년부터 가트너(Gartner)는 신기술 하이프 사이클에 더이상 빅데이터 기술을 포함시키지 않지만, 2017년 자료에 따르면 2~5년 이후 공급망(Supply Chain) 관련 주요한 기술이 될 것으로 예상함(Gartner, 2017)
- 최근 글로벌 주요 50개 기업 대상으로 설문조사를 한 결과, 95% 이상의 기업이 빅데이터 기술에 투자하고 있으며, 비용절감이나 혁신을 위한 새로운 방향 마련 등에서 성과를 거두고 있다고 응답함(NVP, 2017)
- 현재 데이터 간 연계활용, 실시간 분석 능력 향상, 지능·예측형 분석 기술 확보, 빅데이터 수집·처리·저장·분석에 최적의 성능을 얻을 수 있는 기술이 개발되고, 유의미한 정보 획득이라는 기존 빅데이터 기술의 목적이 변화 예측으로 바뀌고 있음(IITP, 2016a)
 - 사물인터넷(IoT) 관련 기기로부터 시계열에 따라 생성되는 다양한 트랜잭션, 고객행동, 센서기록, 보안 위협 등이 2018년 이후에는 빅데이터 추세를 결정할 중요한 요소가 될 것으로 전망되어(Foote, 2017) 실시간으로 분석할 수 있는 기술들이 개발되고 있음
 - 빅데이터와 인공지능 기술이 결합하여 비즈니스 혁신이 이루어지고 있으며, 인공지능의 판단 및 학습능력을 가능하게하기 위해 빅데이터의 수집·분석·활용 기술은 필수적임

2.1 플랫폼기술

 **수집, 저장과 관련된 오픈소스 소프트웨어 솔루션이 활발히 개발되고 있으며, 클러스터 기반 대용량 데이터 처리 및 분산 응용 프로그램을 지원하는 프레임 워크인 하둡(Hadoop)⁷⁾이 가장 널리 활용되고 있음**

6) 기술의 성숙도를 시각적으로 표현하기 위해 가트너에서 개발한 하이프 사이클(Hype Cycle)은 잠재적 기술이 관심을 받는 기술촉발(Technology Trigger) 단계, 대부분의 기업이 관망하는 부풀려진 기대의 정점(Peak of Inflated Expectations) 단계, 관심이 시들해지는 환멸(Trough of Disillusionment) 단계, 많은 기업이 투자하게 되는 계몽(Slope of Enlightenment) 단계 및 시장의 주류가 되는 생산성 안정(Plateau of Productivity) 단계로 구분됨


7) 분산파일시스템(Hadoop Distributed File System, HDFS), 분산프로그래밍 모델(MapReduce), HDFS 분산 기반 데이터베이스(HBase)와 그 외 빅데이터 처리와 관련된 여러 솔루션들을 함께 제공하고, 클릭스트림(Click Stream) 분석, 로그(Log) 처리, 페타바이트(Petabyte) 규모 분석 등에 적용됨(<http://hadoop.apache.org/>)

- 빅데이터의 주요 소프트웨어 솔루션인 수집, 저장(DB), 처리 소프트웨어 등은 2017년도에도 30~300명 수준의 활발한 참여자(Contributor)에 의해 중요한 소프트웨어 버전 업데이트(Major Version Update)가 진행되고 있음(Boncea et al, 2017)


〈표 3〉 주요 빅데이터 오픈소스 솔루션의 개발 현황(2017년도 기준)

	Flume(수집)	MongoDB(저장)	Redis(저장)	HBase/Hadoop (저장/처리)
주요 공개(출시) 횟수	1	2	2	2
지난 1년간 소규모 업데이트 (Minor Release) 횟수	1	14	9	2
활발한 참여자 (Contributor) 수	32	280	220	104/87
중요한 문제 해결에 걸리는 평균 시간(일/문제)	-	2.4	18	7

출처 : Boncea et al.(2017)

 사물인터넷 기술이 확산되면서 생성되는 빅데이터에 대한 수집·저장·처리를 위해 클라우드 기반의 플랫폼, 실시간 처리 기술 등이 개발 및 활용되고 있음

- 구글(Google), 아마존(Amazon), 마이크로소프트(Microsoft), 오라클(Oracle) 등의 기업에서 클라우드형 빅데이터 플랫폼을 제공하는 서비스가 증가하고 있음
- 기존 하둡 시스템으로는 실시간 분석을 해결하기 어려워 데이터를 메모리에 올려 처리하는 인-메모리(In-Memory) 기술이 적용된 스파크 프레임워크⁸⁾, 실시간 컴퓨팅 시스템(Storm)⁹⁾, 분산 스트리밍 플랫폼(Kafka)¹⁰⁾과 같은 기술들이 개발되고 있음(Yadranjiaghdam et al., 2016)
- 트랜잭션 기반의 운영 데이터를 직접 분석할 수 있는 운영·분석 일체형 데이터관리 플랫폼 기술개발이 진행되었음(IITP, 2016a)

 국내 빅데이터 플랫폼 관련 기술 수준은 낮은 편이나, 빅데이터 실시간 처리 기술의 상용화 등이 최근 이루어지고 있음

- 2017년 빅데이터 수집기술 관련 논문 평가(활동도 및 영향력 고려)결과 최고기술국 미국 대비 30.6% 수준으로 조사 대상국(미국, 중국, 일본, 유럽¹¹⁾, 한국) 중 최하위였으며, 저장·처리 기술은 최고기술국 미국 대비 29.3%로 일본과 중국보다는 근소하게 앞서는 것으로 조사됨(IITP, 2017a)

8) <https://spark.apache.org/>


9) <http://storm.apache.org/>

10) <https://kafka.apache.org/>

11) 독일, 프랑스, 스위스, 네덜란드, 영국, 이탈리아, 스웨덴 등 29개 국가


- 2017년 빅데이터 수집기술 관련 특허 평가(활동도, 영향력 및 시장성 고려)결과 최고기술국 일본 대비 37.9% 수준으로 조사 대상국(미국, 중국, 일본, 유럽, 한국) 중 최하위였으며, 저장·처리 기술은 최고기술국 일본 대비 43.9%로 중국보다는 근소하게 앞서는 것으로 조사됨(IITP, 2017a)
- 엘지씨엔에스(LG CNS)는 클라우드 기반으로 인공지능을 활용해 데이터 수집에서 분석, 결과 시각화까지 일련의 과정을 빠르게 수행하여 기업들의 데이터 기반 의사결정을 돕는 ‘답(DAP)’ 서비스를 출시하여 고객사인 LG화학, LG이노텍, LG디스플레이 등에서 활용 중임(조선일보, 2018)
- 한국전자통신연구원(ETRI)은 대규모 트랜잭션 처리와 복합 분석을 통합하여 데이터 이동 없이 단일 시스템에서 데이터 운영·분석 일체화 개발을 진행하였음(IITP, 2016a)

2.2 분석기술

 빅데이터 실시간 분석을 위해 하둡(Hadoop) 기반의 오픈소스 플랫폼인 스파크(Spark)¹²⁾가 많이 활용되고 있음

- 스파크는 디스크 액세스가 필요하지 않는 병렬 인-메모리(In-Memory) 프로세싱 기능을 기반으로 결과를 빨리 제시할 수 있는 장점이 있어 컴퓨터 메모리 및 SSD 가격이 하락됨에 따라 2013년 이후 빅데이터 실시간 분석 툴로 활발하게 확산됨

※ 2017년 기준 스파크는 2번의 주요 공개 및 5번의 소규모 업데이트가 있었고, 오픈소스 수정을 위한 60명의 활발한 참여자가 활동하고 있음

 글로벌 IT 기업 중심으로 인공지능 등이 결합된 빅데이터 분석 플랫폼을 개발하여 서비스하고 있으며, 빅데이터에서 유의미한 정보를 획득할 뿐만 아니라 미래 상황을 예측하고자 하고 있음

- 통계기법, 머신러닝 등으로 빅데이터를 분석하여 정확도가 높고, 사용자의 편의성을 높이는 서비스를 제공함(NIPA 및 NCIA, 2017)
 - 구글(Google)은 통계적 기계번역(Statistical Machine Translation)을 통해 수억 개의 번역 문서에서 패턴을 찾아 언어 간 번역 규칙을 컴퓨터 스스로 발견하여 번역의 정확도를 향상시킨 번역 서비스, 검색데이터와 머신러닝에 기반한 AI비서 서비스 등을 제공하고 있음

12) 즉각적인 고급 분석을 위해 고속 쿼리 수행 툴, 머신 학습 라이브러리, 그래프 프로세싱 엔진, 스트리밍 분석 엔진 등이 포함되어 있음(<https://spark.apache.org/>)

- 마이크로소프트(Microsoft)는 파워포인트에 인공지능 번역기 기능을 탑재하고, 엑셀에 과거 데이터를 기반으로 미래의 변화를 예측해주는 'forecast' 기능을 추가하고, 머신러닝 기반의 문서정리서비스 'MS 델브(Delve)' 및 챗봇 서비스 '팀즈(Teams)' 등을 공개함
- 의료, 금융 등 전문 분야에 특화된 비정형 데이터를 예측·분석할 수 있는 아이비엠(IBM)의 왓슨(Watson), 사물인터넷 데이터를 인공지능 기반의 자동화된 모델로 분석할 수 있는 썬(SAS) 솔루션 등이 개발되어 서비스되고 있음(조선비즈, 2018a 및 2018b)

영상인식, 스트림 분석 등 이종 소스 심층 융합 분석 기술은 해외 IT 관련 기업 위주로 기술개발 및 상품 출시가 이루어짐(IITP, 2016a)

- 영상인식기술은 아이옴니사이언트(iOmniscient), 에이전트비(AgentVi), 보쉬(Bosch) 등 해외 업체들이 강세를 보이고, 스트림 분석 플랫폼으로는 오라클(Oracle), 마이크로소프트(Microsoft), 아이비엠(IBM) 등의 상용 솔루션이 주를 이루고 있음
- 데이터 스트림 마이닝을 위한 오픈소스 소프트웨어는 래피드마이너(RapidMiner), 모아(MOA) 등이 있고, 데이터 시각화 기술은 아이비엠(IBM), 썬(SAS), 오라클(Oracle) 등의 대형 IT 기업과 타블로(Tableau)와 같은 시각화 전문 업체들이 관련 솔루션을 출시하였음

국내 빅데이터 예측·분석, 이종소스 분석 등의 기술은 초기단계임

- 2017년 빅데이터 분석 및 예측 관련 기술의 논문 평가(활동도 및 영향력 고려)결과 최고 기술국 미국 대비 28.1% 수준이고, 특허 평가(활동도, 영향력 및 시장성 고려)결과 최고 기술국 일본 대비 38.7% 수준으로 조사 대상국(미국, 중국, 일본, 유럽, 한국) 중 최하위임
- 비정형적 데이터와 관련된 예측 연구는 시작단계이고, 시각화 기술은 데이터마이닝 작업에 기초한 정보 전달보다는 메시지 전달을 위한 시각 표현 위주의 작업이 이루어졌음(IITP, 2016a)
- 다음, 솔트룩스, SK텔레콤 등이 자체기술로 공간 빅데이터 서비스를 제공하나, 관련 연구는 아직 초기단계이며, 스트림 분석은 대학 및 연구소 중심으로 연구가 이루어졌음(IITP, 2016a)

2.3 활용기술

제조, 유통 산업 관련 도메인 지식(Domain Knowledge)에 빅데이터 플랫폼 및 분석 기술을 적용하고 있음

- 비용 효율성과 결점관리, 새로운 제품이나 고객에 대한 서비스 방향 등을 발견하기 위해 제조업 설비와 센서를 통해 수집된 빅데이터를 적시에 분석하고 있음(Auschwitzky, Hammer and Rajagopaul, 2014)
 - 구글(Google), 테슬라(Tesla)¹³⁾, 엔비디아(NVIDIA) 등은 자율주행차 개발과 관련하여 실제 주행 환경에서 카메라, 레이더, 라이다 등 약 12개의 다양한 센서에서 1일 당 약 4GB의 데이터를 수집하여 새로운 머신러닝 알고리즘 개발에 활용함(한겨레, 2018, Bernasch, 2017)
- 이베이(eBay)의 ‘Shop your Interests’, 마이크로소프트(Microsoft)의 Microsoft Audience Network 플랫폼, 디즈니(Disney)의 Luminata는 빅데이터를 활용하여 소비자의 선호 스타일을 파악하여 맞춤형 서비스를 제공함(Ong. 2018, Throp-Lancaster, 2018, Munson, 2018)
- 아마존(Amazon)은 온라인 쇼핑몰 고객 데이터에서 패턴을 파악하여 서비스 및 물류 최적화를 하였고, 최근 추가적인 개인화 서비스 발굴을 위해 오프라인 고객 데이터를 구축함
 - 온라인 쇼핑몰에 ‘도서추천 시스템’을 개발하여 고객이 읽을 것으로 예상되는 책을 추천 하면서 할인쿠폰을 지급하거나, 고객이 구매하기 전 배송을 준비하여(Anticipatory Shipping) 고객 주소지 근처 물류창고로 배송함(NIPA 및 NCIA, 2017)
 - 최근 오프라인 무인 식료품 매장인 ‘아마존 고(Amazon Go)를 확대하고, 20주간 고객의 체형과 사이즈를 측정하여 체형변화를 이해하려고 하는 등 오프라인 고객 데이터를 구축함 (NYT, 2018, The Seattle Times, 2018, WSJ, 2018)

국내 빅데이터 활용기술 수준은 낮은 편이고, 주로 금융 및 통신 산업 관련 도메인 지식 기반 빅데이터 기술을 활용하고 있음

- 2017년 빅데이터 응용·서비스 기술의 논문 평가(활동도 및 영향력 고려)결과 최고기술국 미국 대비 30.0%로 조사 대상국(미국, 중국, 일본, 유럽, 한국) 중 최하위 수준이고, 특허 평가(활동도, 영향력 및 시장성 고려)결과 최고기술국 미국 대비 42.2% 수준으로 4위권을 기록함(IITP, 2017a)


13) <https://www.tesla.com/autopilot/>

- 신한, KB국민, 비씨카드 등의 카드사는 자사가 보유한 빅데이터로 고객의 구체적인 소비행태를 파악하여 고객 맞춤형 서비스를 제공하고(디지털타임스a, 2018), SK플래닛은 빅데이터 기반 추천 플랫폼 ‘콜로세오’로 11번가의 실시간 맞춤형 상품 추천 서비스를 고도화함(디지털타임스c, 2018)
- 케이티(KT)는 고객 정보와 사물인터넷 기술 등을 활용하여 감염병 확산 방지를 위한 해외 입국자의 이동 경로 파악, 심야버스 노선, 에너지 예측 및 관리(GIGA energy GEN-태양광 Operation & Management) 등 신사업을 발굴하고 있음(매일경제, 2018a, 동아일보, 2018b 및 2018c)
- SK텔레콤은 가입자의 정보 분석을 통해 청소년, 군인, 외국인 등 고객 맞춤형 요금제를 출시하고, SK텔레콤이 보유한 빅데이터를 분석하여 경기도의 따복버스¹⁴⁾ 맞춤형 버스운행 스케줄 마련 및 대구시의 관광·교통·복지시설 등의 행정 수요 예측에 활용함(디지털타임즈, 2018b, 동아일보, 2018a 및 2018b)


14) 경기도 벽지와 오지 산업단지 관광지 등 대중교통이 불편한 지역에서 운행되는 교통복지 수단

제3장 산업동향


3.1 해외 산업동향

 세계 빅데이터 시장은 현재도 성장 중이며, 시장조사기관마다 규모의 차이는 존재하나 공통적으로 높은 성장률을 전망하고 있음

- 위키본(Wikibon)은 2017년 세계 빅데이터 시장규모가 350억 달러 규모였고, 2018년에는 약 20% 성장한 420억 달러, 2027년에는 1,030억 달러에 이를 것으로 전망함(Statista, 2018)
- 2017년 전세계 빅데이터 매출액은 2016년에 비해 12.4% 증가한 1,508억 달러에 달할 것으로 예측하고, 2020년까지 연평균 11.9%의 성장률(CAGR)로 성장하여 매출액이 2,100억 달러를 초과할 것으로 예측함(IDC, 2017)

 2017년 전세계 빅데이터 관련 기업 수는 전체 ICT 기업 중 15% 수준의 비중을 차지하고, 국가별로는 미국이 전체의 48.4%, 중국 10.1%, 일본 2.4%, 한국 1.5%로 조사됨(IITP, 2017b)

- 전세계 빅데이터 관련 기업의 시가총액 중 상위 20대 기업이 차지하는 비중은 84.5%, 상위 10대 기업이 차지하는 비중은 72.4% 수준이고, 국가별로는 미국은 38.8%, 중국 26.5%, 일본 2.6%, 한국 0.1%의 수준으로 나타남

 2018년 상반기 기준으로 전세계 빅데이터 관련 약 3,500여개의 스타트업이 존재하고, 우수 기술을 보유한 스타트업은 꾸준히 M&A가 이루어지고 있음

- 8월 1일 기준 빅데이터 분야에 3,538개의 스타트업과 19,323명의 엔젤 투자자가 있으며 기업 당 평균 투자금액은 460만 달러에 이룸¹⁵⁾

15) <https://angel.co/big-data-analytics>

- 벤처캐피탈리스트의 빅데이터 기업 투자는 꾸준히 이루어지고 있으며, 2017년 대규모 자금이 투자된 기업은 Looker(\$81M), InsideSales(\$50M), DataRobot(\$54M), Confluent(\$50M), DataBricks(\$60M) 등임(Matt Turck, 2017)
- 우수 기술을 보유한 스타트업 기업은 2017년에도 꾸준히 M&A가 이루어지고 있으며, 주요 기업으로는 인텔(Intel)에 인수된 모바일아이(Mobileye), 시스코(Cisco)에 인수된 앱다이너믹스(AppDynamics) 등이 있음(Matt Turck, 2017)


〈표 4〉 주요 스타트업 M&A 현황(2016 및 2017년)

연도	스타트업		인수 기업	인수 금액
	기업명	주력 기술		
2017	Mobileye	비전기반 자율주행 기술	Intel	\$15.3B
	AppDynamics	애플리케이션 성능 관리 기술	Cisco	\$3.7B
	Nimble Storage	예측 플래시 저장 기술	HPE	\$1.1B
	Kaggle	빅데이터 예측분석 플랫폼	Google	-
2016	Qlik	비즈니스 인텔리전스	Thoma Bravo	\$3B
	Cruise Automation	자율주행 기술	General Motors	\$1B
	Apigee	API 매니지먼트	Google	\$625M
	OPower	클라우드 기반 에너지 관리 플랫폼	Oracle	\$532M

빅데이터 관련 주요 기업들의 동향을 살펴볼 때 클라우드, 인공지능과 연계한 사업 확장이 주요 이슈로 논의됨


- 아마존(Amazon)은 아마존 웹서비스(AWS)를 통해 클라우드와 연계한 빅데이터 제품의 시장 지배력을 강화하고, 구글(Google)은 폭넓은 빅데이터 제품¹⁶⁾을 구축하고, 기계학습 관련 커뮤니티인 캐글(Kaggle)을 인수하여 새로운 기계학습 API 구축을 진행함(Lardinois, 2017)
- 마이크로소프트(Microsoft), 오라클(Oracle) 등은 클라우드, AI와 연계된 빅데이터 서비스를 추진하고, 아이비엠(IBM)과 세일즈포스(Salesforce), 에스에이피(SAP)와 구글(Google)처럼 ‘데이터(리포지토리) 보유 기업’과 ‘인공지능기술 보유 기업’간의 협력관계가 형성되고 있음 (Matt Turck, 2017)

16) BigQuery, Dataflow, Dataproc, Datalab, Dataprep 등

 플랫폼기술 관련 산업의 주요 이슈는 SQL¹⁷⁾의 재등장, 데이터 가상화¹⁸⁾ 등이 있고, 분석기술 관련 산업에서는 자동화된 분석도구가 주요화됨

- 최근 SQL 데이터베이스 기술인 구글(Google)의 스패너(Spanner) 클라우드 버전, 아마존(Amazon)의 아테나(Athena)등이 출시되었고, 구글(Google) 빅쿼리(BigQuery), 스파크(Spark) SQL 및 프레스토(Presto)가 많이 활용되고 있음(Matt Turck, 2017)
- 데이터 가상화의 급속한 증가로 인해 차세대 분석업체들은 데이터 가상화 및 처리를 제공하고 있으며, 안전하고 쉽게 데이터에 접근하기 위한 솔루션들을 인포매티카(Informatica), 콜리브라(Colibra), 에이레이션(Alation) 등에서 제공하고 있음(Matt Turck, 2017)
- 아마존(Amazon), 구글(Google), 마이크로소프트(Microsoft)는 데이터센터를 건립하는 등 데이터 수집에 315억 달러를 투자하여 빅데이터 저장 능력을 증가시킴(조선일보, 2017)
- 아직까지는 빅데이터 기술 전문가가 수요에 비해 많지 않아 기업에서 관련 인재를 영입하기 어려운 실정인데 최근 자동화된 분석도구가 등장하여 빅데이터 분석기술에 능하지 않더라도 상대적으로 쉽게 접근할 수 있게 되었음(Matt Turck, 2017)


3.2 국내 산업동향

 2017년 국내 빅데이터 시장규모는 4,547억 원으로 추정되고 이는 전년대비 32.2% 성장한 수치로 지속적인 상승세를 유지하고 있음(NIA, 2018)

- 국내 빅데이터 시장은 스토리지(26.8%), 소프트웨어(23.2%), 서버(22.1%), 서비스(20.9%), 네트워크(7.1%) 순으로 서버, 스토리지, 네트워크 등의 인프라 투자가 소프트웨어나 서비스에 비해 상대적으로 집중되어 있음
- 시장영역별로 공공시장은 빅데이터 도입, 활용에 대한 예산증액 등으로 인해 34% 성장한 1,338억 원으로 시장비중이 29.4%로 나타났고, 민간시장은 금융, 유통, 통신업과 일부 대기업에 한정된 수요로 인해 31.5% 성장한 3,209억 원에 그치며 시장비중이 70.6%로 소폭 하락함

17) 관계형 데이터베이스 관리시스템의 데이터를 관리하기 위해 설계된 특수 목적의 프로그래밍 언어(<https://ko.wikipedia.org/wiki/SQL>)

18) 데이터를 클라우드에 저장하고 가상환경에서 분석하는 것


 국내 기업의 빅데이터 관련 투자가 확대되고, 매출도 증가하는 등 산업 기반이 갖추어지고 있으나, 아직은 관련 인프라 및 글로벌 선도 기업이 부족한 실정임

- 국내 기업의 빅데이터 시스템 도입률은 2016년 약 50개 기업에서 2017년 약 65개 기업으로 증가하였고, 빅데이터 도입 기업의 평균 투자금액도 2016년 5억 2,100만원에서 2017년 6억 100만원으로 증가함(NIA, 2018)
- 2016년 데이터 산업 직접 매출 시장 규모는 5조 8,803억 원으로 2015년 대비 3.3% 성장한 것으로 조사됨(KDATA, 2017)
- 빅데이터 기술발전의 기반이 되는 10PF급 이상 슈퍼컴퓨터는 중국 2개, 일본 3개, 미국이 4개를 보유한 반면 국내에는 전무하고(과학기술정보통신부, 2018.5.), 글로벌 100대 빅데이터 기술혁신 기업 중 국내기업은 포함되지 못함¹⁹⁾


19) <https://www.crn.com/news/applications-os/300084738/the-2017-big-data-100.htm>

제4장 정책동향

4.1 해외 정책동향

 (미국) 2012년 ‘빅데이터 R&D 이니셔티브’²⁰⁾를 발표한 이후 2016년 ‘빅데이터 R&D 전략계획’²¹⁾을 발표하여 빅데이터 방법론, 데이터 신뢰성 확보 등을 위한 R&D 전략을 제시함(IITP, 2016b)

- 미래에 요구될 새로운 빅데이터 기술 관련 방법론 개발 등의 차세대 기술 개발, 데이터 신뢰성 확보를 통한 빅데이터 기반 의사결정 R&D, 사이버 인프라 구축, 데이터 공유 및 관리 촉진 정책, 개인정보 보호 및 보안 등 7대 전략과 18개의 세부과제가 포함됨


 (EU) 2018년 4월 25일 데이터를 기반으로 신제품 및 신서비스 개발이 가능하도록 유럽 내 공통된 데이터 공간을 창출하기 위한 중요한 조치들을 채택함²²⁾

- 이번에 채택한 ‘공통 유럽 데이터 공간’ 관련 조치는 2014년 7월 채택한 ‘데이터 주도 경제’와 2017년 1월 채택한 ‘유럽 데이터 경제 구축’의 후속조치로 볼 수 있음
- 서로 다른 부문, 국가 및 분야의 데이터 통합을 제안한 본 조치는 공공 부문 정보의 재사용 지침에 대한 검토 제안, 과학적 정보의 접근 및 보존에 관한 2012년 권고안 개정 및 B2B 및 B2G 환경에서의 민간부문 데이터 공유에 대한 지침 등이 포함됨
- 또한, 데이터 경제의 지속 가능한 발전을 위해 일반정보보호규정 등의 데이터 보호 규칙은 완전하게 수행되어야 함을 전제로 하고 있음


20) 과학기술정책국(Office of Science and Technology)에서 빅데이터 기술개발 및 활용, 빅데이터 전문인력 양성 등을 주요 목적으로 발표

21) 미국 내 부처 및 기관들이 협력하는 ICT 분야의 R&D를 추진하는 범부처 R&D 프로그램인 NITRD(Networking and Information Technology Research and Development)를 통해 발표함


22) <https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy>

 (EU) 2018년 5월 25일 일반정보보호규정(General Data Protection Regulation, GDPR)²³⁾이 시행되어, 향후 데이터 관리와 개인정보보호를 위한 국제적 프로토콜이 될 가능성이 있음

- EU 회원국 시민들의 개인정보는 적절한 관리와 보안의 대상으로 시민들은 데이터 관리자로 부터 개인 데이터 처리방식을 투명하게 제공받을 수 있고, 데이터 관리자에게 개인 데이터를 삭제하여 더 이상 사용하지 못하도록 요청할 권리 등이 있음을 제시함
- 기업 등이 해당 규정을 위반할 경우 72시간 내에 규제 기관에 보고해야하며, 데이터 유출이 개인정보리와 자유에 높은 위험을 초래할 가능성이 있을 시 개인에게 통보하도록 하고, 규정을 위반하는 기업에 대해 글로벌 매출의 4%에 달하는 벌금 부과가 가능함

 (영국) ‘빅데이터 역량강화 종합전략(2013)’²⁴⁾ 이후, 2017년 ‘정부 변화 전략’²⁵⁾을 통해 2020년까지 정부데이터의 효과적인 활용을 위한 전략을 제시함

- 공공서비스를 위해 정부 기관 간 경계를 허물어 데이터를 효과적으로 사용할 수 있도록 하고, 디지털 경제 법안(Digital Economy Bill)이 의회에서 통과되고 나면 데이터 공유 조항을 통해 전체 정부의 데이터를 효과적으로 사용할 수 있도록 함
- 새로운 데이터 담당 책임자(CIO)를 임명하고, 새로운 데이터 자문위원회(Data Advisory Board)를 구성하여 정부 간 데이터 활용을 극대화하기 위한 조정에 힘씀
- 데이터의 효과적인 활용을 위해 정부 차원에서 데이터 과학과 분석 능력 구축, 데이터 공유 관련 윤리 교육, 국가 데이터 인프라 구축 및 보완성 확보, 데이터 검색 도구 개선, 주요 리포지토리 저장 및 관리 방식 변화 등을 추진함

 (중국) 공업정보화부는 ‘빅데이터 발전 촉진을 위한 행동요강’²⁶⁾의 구체적 실행 방안으로 2016년 ‘빅데이터 산업 발전 계획’을 발표하고 관련기업 육성을 목표로 함 (IITP, 2017b)

- 2020년까지 빅데이터 관련 제품 및 서비스 매출을 1조 위안까지 확대하기 위해 10개 이상의 빅데이터 글로벌 선두기업과 500개의 빅데이터 응용서비스 기업을 육성하고자 함


23) <https://www.eugdpr.org/eugdpr.org-1.html>

24) 본 계획(Seizing the data opportunity: a strategy for UK data capability)은 전문인력양성, 인프라 SW 개발 및 R&D 협력체계 구축, 공공데이터 접근성 및 안정성 확보 등을 제시함(NIPA, 2014)

25) Government Transformation Strategy의 5개 전략 중 하나로 데이터 사용 개선에 대한 전략을 제시하고, 2020년까지 추진하고자 하는 세부과제를 제시함(UK Government, 2017)


26) 빅데이터의 개발 및 응용을 통한 ‘대중창업, 만민혁신’을 추진하고, 경제성장을 위한 신동력 개발 그리고 정부의 관리 감독 서비스를 개선하고자 함

- 빅데이터 기술개발, 제조업 생산과정에 빅데이터를 전면 도입하는 등 응용분야 확대, 산업 역량 및 보안시스템 강화 등을 추진함


 (일본) ‘일본재흥전략 2016’²⁷⁾과 ‘미래투자전략 2017’을 통해 빅데이터 활용의 중요성을 강조하고, 관련 기반과 제도를 구축하고자 함(최해옥, 2017)

- (일본재흥전략) 사물인터넷이 빅데이터 수집의 중요한 수단임을 고려하여 ‘산업경쟁력회의’를 중심으로 민관전문가 위원회를 구성하여 규제완화 도입, 데이터 공유·이용 촉진, 일본 내 혁신 창조, 인적자원개발 등에 초점을 맞추어 추진함
- (미래투자전략) 2020년까지 ‘관민데이터 활용추진 전략회의’를 중심으로 공공데이터 개방 집중 실시, 데이터 활용 권한을 명확히 하는 기업 간 계약 가이드라인 전개, 데이터 유통을 위한 법제정비, 정보은행·데이터 거래시장의 제도정비 등을 실시함

4.2 국내 정책동향

 2017년 11월 발표된 ‘4차 산업혁명 대응계획’은 지능화 기반 산업 및 사회·공공 분야에 빅데이터 기술을 활용하여 지능화 혁신을 이루고자 함²⁸⁾

- 의료기관 간 진료정보 전자교류 전국 확대 및 보건의료 데이터 구축·개방을 추진하여 데이터 인프라 확보, 신약개발주기 및 비용 단축을 위한 빅데이터 활용 등을 추진
- 제조업, 에너지, 농수산업, 스마트시티 등의 분야에 빅데이터 기술을 활용하거나, 빅데이터를 구축하여 지능화 혁신을 추진

 우리나라 정부는 빅데이터를 혁신성장동력으로 지정하고(’17.12)²⁹⁾, 2022년까지 해당분야 조기 상용화를 위해 제도개선, 연구개발, 인프라 구축 등의 지원 방안을 발표함(’18.5)³⁰⁾

- 개인정보의 안전한 활용을 위해 개인정보 범위 명확화, 비식별 조치 법제화 등의 제도개선 논의를 추진하여 합의(안)을 도출하고, 정보 주체 중심의 데이터 활용 촉진을 지원


27) 2016년 발표한 정책으로 4차 산업혁명의 핵심기술로 IoT, 빅데이터, 인공지능 로봇 기술을 지정하고 2020년까지 30조 엔이 부가가치 창출을 목표로 함

28) 관계부처 합동, 혁신성장을 위한 사람중심의 4차 산업혁명 대응계획, 2017.11.

29) 관계부처 합동, 혁신성장동력 추진계획, 2017.12.22.

30) 관계부처 합동, 혁신성장동력 시행계획, 2018.5.28.


- 인공지능 및 초연결 사회를 위한 빅데이터 연구개발을 지원하고, 빅데이터 분석 플랫폼과 유통 플랫폼의 구축 및 고도화 지원

 최근 대통령 직속 4차산업혁명위원회(2018)에서는 데이터 산업 활성화 전략을 발표하여 개인정보 보호와 활용의 균형 및 데이터 경제 활성화를 추진하고자 함


- 정보주체인 개인이 데이터 통제·활용권을 가지도록 하고, 개인정보의 안전한 활용을 위한 법제화·안심존 구축·신기술 적용 등을 지원할 뿐만 아니라 데이터 구축·개방, 저장·유통, 분석·활용의 지원체계를 마련하고, 핵심기술 확보 및 전문인력 확보 등을 통해 기반 조성을 추진함
- 또한, 국회 4차산업혁명특별위원회는 개인정보보호법과 정보통신망법, 신용정보법 등 중복 조항 정비 및 강력한 사후규제를 도입하되 가공정보에 대한 안전한 활용이 가능하도록 입법 권고(안)를 발표함(전자신문, 2018)
 - 법률상 개인정보와 가명정보의 개념을 구체화하고, 가명정보 개념 신설 시 목적 외 이용 및 제3자 제공가능 상황을 규정, 가명화된 개인정보 결합 추진기관 법제화, 개인정보보호 위원회 위상강화 등을 추진하도록 함

제5장 R&D투자 동향


5.1 해외 R&D투자 동향

 (미국) 2019년 연구개발 정부예산(안)에 빅데이터에 대한 직접적인 언급은 없으나, 인공지능과 고성능 컴퓨팅 투자를 우선순위로 제시하여 빅데이터 분야도 투자가 증가될 것으로 예측됨³¹⁾

- 국립과학재단(NSF)은 인공지능 관련 기초연구 지원을 하고, 교통부(Department of Transportation)에서 자율 및 무인시스템 추가 개발을 지원하며 국방부(Department of Defense)는 빅데이터 기반 AI 알고리즘 개발(warfighting system)을 할 예정임
- 국립보건원(National Institute of Health)에서는 암연구 관련 빅데이터의 빠른 분석을 위해 고성능 컴퓨팅에 투자하고, 에너지부(Department of Energy) 예산으로 국립연구소인 아르곤(Argonne) 및 오크릿지(Oak Ridge) 슈퍼 컴퓨팅 인프라 업그레이드를 할 예정임

 (일본) 2018년 빅데이터 연구개발은 주로 ‘Society 5.0(초스마트사회) 실현 플랫폼’과 관련하여 투자가 이루어지고 있음

- Society 5.0 실현플랫폼과 관련하여 2018년 ICT R&D에 1,295억엔을 투자하고(IITP, 2017c), 특히 총무성은 IoT/BD/AI 정보통신 플랫폼 사회실장 추진사업에 2.6억 엔을 투자하고, 문부과학성은 인공지능/빅데이터/IoT/사이버시큐리티 통합프로젝트로 85.6억 엔을 투자함(유종태, 2018)

 (EU) 최근 정확하고 자원 최적화된 빅데이터 기술에 대해 투자함

- 유럽연합 집행위원회(European Commission)는 매우 방대한 양의 빅데이터를 정확하게 분석하고, 최적화된 전산 자원 사용으로 에너지 효율성을 높일 수 있는 예측 소프트웨어 솔루션을 찾기 위해 최근 200만 유로의 호라이즌 상을 제시함³²⁾

31) https://www.whitehouse.gov/wp-content/uploads/2018/02/ap_18_research-fy2019.pdf

32) <https://ec.europa.eu/digital-single-market/en/news/eur-2-million-prize-big-data-forecast->

5.2 국내 R&D 투자동향

2016년 증가하였던 빅데이터 분야 정부 R&D 투자 비중이 2017년 전년대비 소폭 감소하였고, 연구비 기준으로 개발단계에 많이 투자되고 있음

- 전체 정부R&D³³⁾ 대비 빅데이터 분야 지원현황은 과제 수 기준으로 2015년 1.4%, 2016년 2.8%로 증가하다 2017년 2.6%로 감소하였고, 예산 기준으로는 2015년 1.6%, 2016년 2.3%로 증가하던 예산이 2017년 2.2%로 감소함
- 연구개발단계로 살펴볼 때, 우리나라 정부R&D는 개발단계에 많은 투입을 하고 있으며, 기초연구는 개발 및 응용 연구에 비해 평균적으로 소규모 예산으로 수행되는 과제가 많음³⁴⁾

〈표 5〉 빅데이터 투자 현황(2015년~2017년)

구분	과제 수			예산(백만원)		
	2015년	2016년	2017년	2015년	2016년	2017년
기초연구	320	491	711	64,891	97,334	118,683
응용연구	123	196	216	81,943	73,601	88,551
개발연구	259	760	490	137,599	239,873	187,848
기타	36	104	174	17,296	32,949	31,769
소계	738	1,551	1,591	301,729	443,757	426,851

정부R&D의 연구수행 주체별 예산비중으로 살펴볼 때, 산·학·연 중 기업 지원 비중이 높음

- 2015년~2017년 산·학·연 정부R&D 예산 투입 비중을 살펴보면, 연도별 차이는 있으나 기업 투자 비중이 높으며, 기업 중에서도 중소기업 투자 비중이 높음
- 2016년 및 2017년의 경우, 과제당 평균 예산을 살펴보면 대기업의 평균 예산이 가장 높았고, 그 뒤로 중견기업, 연구소, 중소기업 및 학교 순으로 나타남³⁵⁾

ing-system-will-enable-more-efficient-use-energy-resources

33) 정부 R&D 과제 수 및 예산은 2015년 54,433건 및 18.9조원이고, 2016년 54,827건 및 19조원, 2017년 61,280건 및 19.4조원임

34) 기초, 응용 및 개발 단계 과제당 평균 예산은 2015년 각각 202백만원, 666백만원, 531백만원이고, 2016년 198백만원, 375백만원 및 315백만원이며, 2017년 166백만원, 410백만원 및 383백만원임

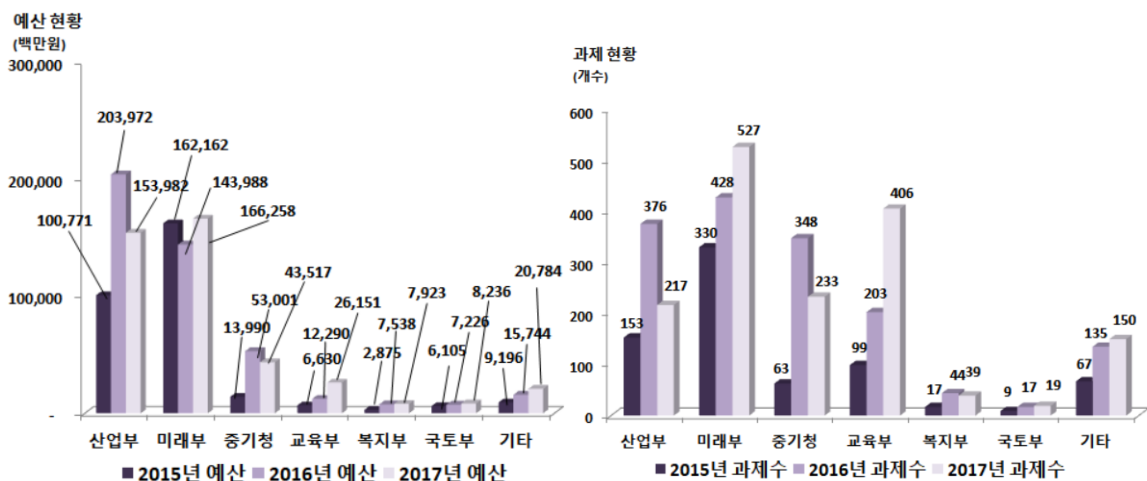
35) 과제당 평균 예산의 경우, 2015년은 2016년 및 2017년과 상이한 패턴이 있었고, 과제당 평균 예산이 높은 순서는 연구소, 대기업, 중견기업, 중소기업 및 학교 순으로 나타남

〈표 6〉 연구수행주체별 현황(2015년~2017년)

(단위 : 백만원)

구분		2015년			2016년			2017년		
		과제 수	예산		과제 수	예산		과제 수	예산	
			금액 (백만원)	비중		금액 (백만원)	비중		금액 (백만원)	비중
산	대기업	6	6,590	35.0	7	8,446	52.2	7	9,144	45
	중견기업	9	7,785		15	12,695		9	7,736	
	중소기업	210	91,178		703	210,433		484	175,137	
	소계	225	105,553		725	231,596		500	192,019	
학		415	79,141	26.2	662	105,922	23.9	909	123,253	28.9
연	출연연구소	62	91,480	30.6	81	65,279	15.0	92	76,330	18.8
	국공립연구소	7	888		20	1,466		29	3,977	
	소계	69	92,368		101	66,746		121	80,307	
기타	정부부처	1	42	8.2	1	100	8.9	7	649	7.3
	기타	28	24,626		62	39,394		54	30,623	
	소계	29	24,668		63	39,495		61	31,272	
총합계		738	301,729	100	1,551	443,758	100	1,591	426,851	100

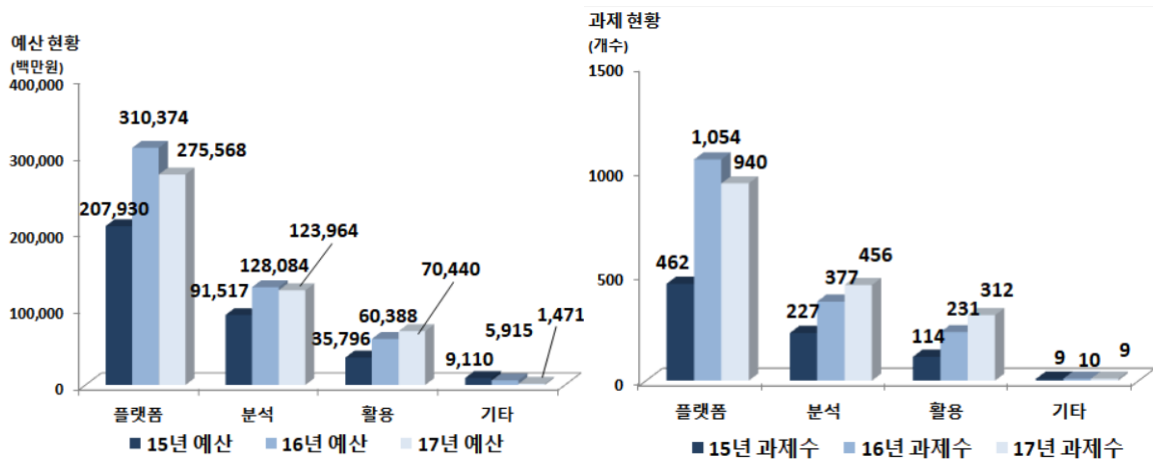
부처별로는 (구)미래창조과학부, 산업통상자원부, (구)중소기업청 순으로 지원 비중이 높으며, 2017년 교육부의 지원 과제수가 2배 증가함



[그림 5] 부처별 지원 현황(예산(좌), 과제 수(우))

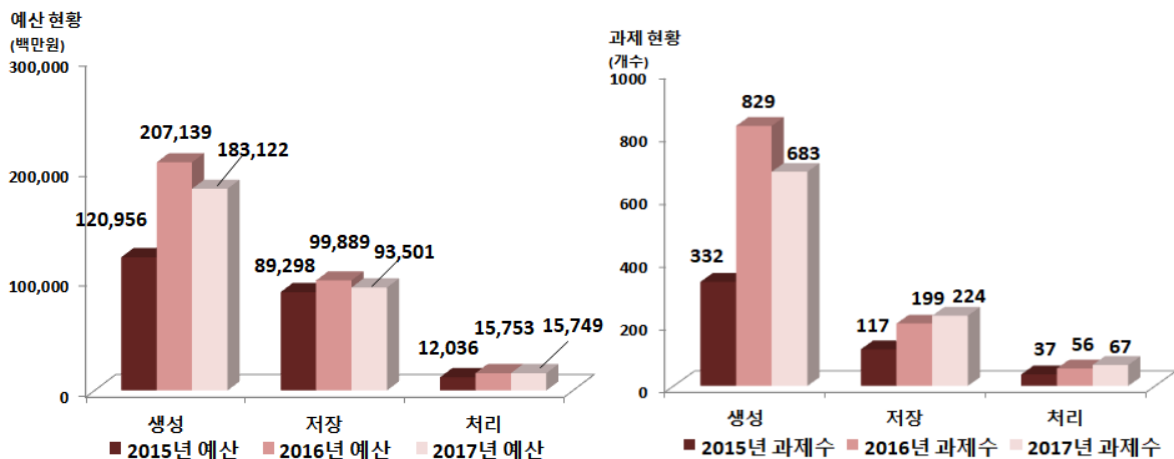
혁신성장동력 기술범위로 지원현황을 살펴보면, 분석 및 활용기술에 비해 플랫폼기술 지원 비중이 높고, 플랫폼 기술 중에서 빅데이터 생성기술의 지원 비중이 높음

- 빅데이터 분야 전체 예산 대비 플랫폼 기술의 예산 현황은 각각 2015년 2,079억원(69%), 2016년 3,104억원(70%) 및 2017년 2,757억원(65%)이고, 2017년 기준 과제 수는 플랫폼기술 940개, 분석기술 456개, 활용기술 312개 및 기타 9개임



[그림 6] 혁신성장동력 기술범위별 현황

- 플랫폼기술 중에서는 데이터 생성 기술의 예산 비중이 높고³⁶⁾(2015년 58%, 2016년 67% 및 2017년 66%), 데이터 생성 기술 중에서는 사물인터넷 관련 기술의 비중이 높음³⁷⁾




[그림 7] 플랫폼 기술의 하위 기술별 현황

36) 생성, 저장 및 처리 기술이 함께 포함된 과제가 존재하여 분류별로 중복을 허용하여 산출함


37) 빅데이터 생성과 관련된 사물인터넷 과제 수 및 예산은 2015년 각각 263건 및 848억원, 2016년 734건 및 1,830억원, 2017년 381건 및 1,221억원임

제6장 결론


6.1 요약

 빅데이터 분야는 오픈소스 소프트웨어 솔루션들이 널리 사용되고, 클라우드, 사물인터넷, 인공지능 등의 기술과 밀접하게 연관되어 기술이 개발되고 있음

- 플랫폼 및 분석 관련 오픈소스 소프트웨어 솔루션이 활발히 개발되고 있고, 사물인터넷 기술의 확산에 따라 생성되는 빅데이터 수집·저장·처리를 위해 클라우드 기반의 플랫폼, 실시간 처리 기술이 개발되고 있음
- 글로벌 IT 기업 중심으로 인공지능 등이 결합된 빅데이터 분석 플랫폼을 개발하여 서비스하고, 미래 상황 예측에도 활용하고 있음
- 제조업의 설비와 센서 등에서 수집된 빅데이터를 적시에 분석하고, 유통산업 고객 데이터를 활용한 최적화된 서비스를 제공하고자 하고 있으며, 우리나라는 금융 및 통신기업 중심으로 고객 맞춤형 서비스 개발, 신사업 발굴 등에 적용하고 있음


 전 세계적으로 빅데이터 시장은 높은 성장률을 보이고 있으며, 세계 주요 기업들은 클라우드, 인공지능 등과 연계하여 사업을 확장하고 있음

- 세계 빅데이터 시장은 현재도 높은 성장률로 성장하고 있으며, 2017년 기준으로 전체 ICT 기업 중 빅데이터 관련 기업은 약 15%정도로 파악됨
- 빅데이터 관련 주요 기업들은 클라우드, 인공지능과 연계하여 사업을 확장하고 있으며, 우수 기술을 보유한 스타트업들의 꾸준한 M&A가 이루어지고 있음
- 우리나라는 정부 및 민간의 빅데이터 관련 투자가 확대되고, 매출도 증가되는 등 산업의 기반을 마련하고 있음


 해외 빅데이터 정책 동향과 유사하게 우리나라도 개인정보보호 및 데이터 활용에 관한 정책이 발표되고 있으며, 빅데이터 분야 R&D지원은 증가 추세임

- 미국은 빅데이터 기술 관련 차세대 방법론 개발, 데이터 신뢰성 확보를 통한 의사결정 R&D 등을 지원하고, EU는 최근 데이터 관리 및 개인정보 보호 의무를 제시하였으며, 영국, 일본 및 중국은 빅데이터 활용을 촉진하고자 함
- 우리나라는 2016년 빅데이터 분야 R&D 투자가 전년대비 급격하게 증가한 이후 2017년 소폭 감소한 상태이며, 분석 및 활용기술에 비해 플랫폼기술에 투자 비중이 높음

6.2 정책제언

 혁신성장동력 정책 개선 시 13대 혁신성장동력 중 차세대통신(사물인터넷), 인공지능 분야 정책과 긴밀한 연계와 우수기술 기반 스타트업 육성을 고려할 필요가 있음

- 빅데이터 기술은 사물인터넷, 클라우드, 인공지능 등의 기술 발전과 사회 수용도 등을 고려하여 발전하고 있어 정책 마련 시 종합적인 고려가 필요한 상황으로 혁신성장동력 13대 분야 중 사물인터넷 및 인공지능 분야를 상호 연계할 수 있는 방안을 마련할 필요가 있음
 - 혁신성장동력 정책 중 빅데이터 분야의 민간·공공 데이터 개방이나 분석플랫폼 구축 등은 인공지능 및 사물인터넷 분야에서 제시하고 있는 정책들과 연계 추진을 고려할 필요가 있음
- 우수기술 보유 스타트업과 글로벌 IT 기업의 M&A를 통해 사업 확장이 이루어지고 있음을 고려할 때, 원천기술 확보를 위한 R&D 지원을 통해 참신한 기술 또는 활용도가 높은 기술을 보유한 스타트업 육성이 필요함
 - 빅데이터 분야 원천기술 확보를 위해 다양한 접근법으로 문제 해결을 시도하는 것이 중요하므로 연구자 주도로 과제 제안 및 수행하는 상향식 과제 지원 비중을 늘리는 방향을 검토할 필요가 있음

 빅데이터 기술이 향후 기반기술로 자리매김하기 위해 관련기술 전반에 걸친 지속적인 투자와 더불어 제도개선이 함께 추진되어야 할 것으로 사료됨

- 빅데이터 분야 인력현황과 플랫폼, 분석 및 활용 기술별 필요한 예산 규모 등을 종합적으로 고려하되 지속적으로 빅데이터 전반의 기술에 대한 R&D 투자를 할 필요가 있음
- 개인정보 보호와 활용의 조화를 위해 정부에서 관련 정책을 마련하고, 국회에서 논의된 입법 권고(안)들이 발표되었으므로, 빠른 시일 내에 관련 법제도 개선을 추진할 필요가 있음

| 참고 문헌 |

● 해외 문헌

- Bernasch, J. (2017). Machine Learning & Active Safety Using Autonomous Driving and NVIDIA DRIVE PX. GPU Technology Conference 2017. Silicon Valley , USA.
- Boncea et al. (2017). A Maturity Analysis of Big Data Technologies, Informatica Economică vol. 21, no. 1/2017, DOI: 10.12948/issn14531305/21.1.2017.05
- Fenn, J., and LeHong, H. (2011). Hype Cycle for Emerging Technologies, 2011. Gartner.
- Hu, H., Wen, Y., Chua, T., and Li, X. (2014). Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. IEEE Access Vol. 2.
- NVP. (2017). Big Data Business Impact: Achieving Business Results through Innovation and Disruption.
- Reinsel, D., Gantz, J., and Rydning, J. (2017). Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on the Data That's Big. [IDC White Paper Sponsored by SEAGATE].
- Yadraniaghdam, B., Pool, N., and Tabrizi, N. (2016) A Survey on Real-Time Big Data Analytics: Applications and Tools. 2016 International Conference on Computational Science and Computational Intelligence. DOI: 10.1109/CSCI.2016.0083

● 국내 문헌

- 과학기술정보통신부. (2018). I-Korea 실현을 위한 인공지능(AI) R&D 전략. 2018.5.
- 관계부처 합동. 혁신성장을 위한 사람 중심의 4차 산업혁명 대응계획. 2017.11.
- 관계부처 합동. 혁신성장동력 추진계획. 2017.12.22.
- 관계부처 합동. 혁신성장동력 시행계획. 2018.5.28.
- 대통령 직속 4차산업혁명위원회 및 관계부처 합동. (2018). 데이터 산업 활성화 전략. 2018.6.26.
- 관계부처 합동. 데이터 산업 활성화 전략. 2018.6.26.
- 김광섭. (2016). 빅데이터산업의 부상 및 시사점. 산은조사월보.
- 유종태. (2018). 일본의 연구개발 동향. KISTEP 기술동향브리프. 한국과학기술기획평가원.
- 정보통신산업진흥원(NIPA). (2014). 영국 정부의 빅데이터 역량 강화 전략. 해외 ICT R&D 정책동향.
- 정보통신산업진흥원(NIPA) 및 차세대융합콘텐츠산업협회(NCIA). (2017). 스마트콘텐츠 월간 동향보고 2017년 11월호.

- 정보통신기술진흥센터(IITP). (2016a). ICT 중장기 기술로드맵 2022.
- 정보통신기술진흥센터(IITP). (2016b). 미국의 빅데이터 R&D 전략계획. 해외ICT R&D 정책 동향.
- 정보통신기술진흥센터(IITP). (2017a). 4차 산업혁명을 선도하는 주요 기술 대상 기술수준평가 및 기술수준 향상방안.
- 정보통신기술진흥센터(IITP). (2017b). 중국의 AI·IoT·빅데이터 정책 및 산업동향. 주간기술동향.
- 정보통신기술진흥센터(IITP). (2017c). 미국과 일본의 R&D 예산(안) 분석.
- 최해옥. (2017). 일본 국가전략인 「미래투자전략 2017」 대응 정책과 시사점. 과학기술정책연구원.
- 한국데이터진흥원(KDATA). (2017). 2017 데이터산업 백서.
- 한국정보화진흥원(NIA). (2018). 2017년 Big Data 시장현황 조사.

● 언론기사

- 디지털타임스. (2018a). “고객을 읽는다” ... 카드사 ‘빅데이터’ 확산. 2018.1.2.
- 디지털타임스. (2018b). 빅데이터 적용 SKT 맞춤형요금제 ‘잘 나가네’. 2018.1.24.
- 디지털타임스. (2018c). 소비자 추천상품 콕 집어내는 ‘콜로세오’. 2018.4.2.
- 동아일보. (2018a). 대구시 “모바일 빅데이터 활용해 유동인구 분석”. 2018.1.5.
- 동아일보. (2018b). 공공사업 핵심자원으로 뜨는 빅데이터. 2018.1.29.
- 동아일보. (2018c). KT, 중소형 태양광 발전소 관리사업 7월 진출. 2018.4.30.
- 매일경제. (2018a). 황창규 회장, 빅데이터 활용 감염병 확산 방지. 2018.1.16.
- 전자신문. (2018). 정책권고 105건·입법권고 47건 ... 4차 산업혁명 나침반 기대. 2018.5.29.
- 조선일보. (2017). “앞으로 100년은 빅데이터 싸움” 구글·MS·아마존 年36조원 투자. 2017.7.25.
- 조선일보. (2018). 빅데이터 분석 기간 크게 줄인 ‘DAP’으로 시장 공략. 2018.3.5.
- 조선비즈. (2018a). “왓슨은 수십, 수백가지로 변신 가능한 AI”. 2018.6.21.
- 조선비즈. (2018b). 올리버 샤벤버거 SAS 수석부회장 “디지털 경제는 분석의 경제”. 2018.5.16.
- 한겨레. (2018). 구글 자율주행차, 지구 200바퀴 돌다. 2018.3.2.
- The New York Times. (2018) Inside Amazon To, a Store of the Future. 2018.1.21.
- The Wall Street Journal. (2018) Amazon wants to know your waistline. 2018.5.3.
- The Seattle Times. (2018) Amazon Go targets Chicago, San Francisco for new store. 2018.5.14.

● 웹 사이트

- Auschitzky, E., Hammer, M., and Rajagopaul, A. (2014). How big data can improve manufacturing. <https://www.mckinsey.com/business-functions/operations/our-insights/how-big-data-can-improve-manufacturing>
- Choudhury, S. (2014) How P&G uses Data Visualization to uncover new opportunities for growth? <https://www.fusioncharts.com/blog/how-pg-uses-data-visualization-to-uncover-new-opportunities-for-growth/>
- Foote. 2017. Big Data Trends for 2018. <http://www.dataversity.net/big-data-trends-2018/>
- Gartner. 2017. Gartner's Hype Cycle Reveals the Digitalization of the Supply Chain. September 11. <https://www.gartner.com/newsroom/id/3798863>
- IDC. (2017) Big data and business analytics revenue forecast to reach \$150.8 billion this year, led by banking and manufacturing investments, according to IDC. <https://www.idc.com/getdoc.jsp?containerId=prUS42371417>
- Lardinois, F. (2017). Google is acquiring data science community Kaggle. <https://techcrunch.com/2017/03/07/google-is-acquiring-data-science-community-kaggle/>
- Matt Turck. (2017) Firing on All Cylinders: The 2017 Big Data Landscape. <http://matrturck.com/bigdata2017/>
- Munson, B. (2018) Disney intros Luminate, a new data-driven targeted ad product. <https://www.fiercecable.com/video/disney-intros-luminate-a-new-data-driven-targeted-ad-product>
- Ong, T. (2018) eBay's adding a customized homepage to feature products based on your interests. <https://www.theverge.com/2018/5/17/17364318/ebay-personalized-homepage-shop-your-interests-feature/>
- Statista. (2018) Big data market size revenue forecast worldwide from 2011 to 2027 (in billion U.S. dollars). <https://www.statista.com/statistics/254266/global-big-data-market-forecast/>
- Thorp-Lancaster, D. (2018) Microsoft Audience Network gives advertisers a boost with AI, LinkedIn data. <https://www.windowscentral.com/microsoft-audience-network-gives-advertisers-boost-ai-linkedin-data>
- UK Government. (2017). Government Transformation Strategy: better use of data. <https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy-better-use-of-data>
- WEF. (2012). "The top 10 emerging technologies for 2012". Last modified Feb. 15. <https://www.weforum.org/agenda/2012/02/the-2012-top-10-emerging-technologies/>

- <http://hadoop.apache.org>
- <https://spark.apache.org>
- <http://storm.apache.org>
- <https://kafka.apache.org>
- <https://www.tesla.com/autopilot>
- <https://angel.co/big-data-analytics>
- <https://www.eugdpr.org/eugdpr.org-1.html>
- <https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy>
- https://www.whitehouse.gov/wp-content/uploads/2018/02/ap_18_research-fy2019.pdf
- <https://ec.europa.eu/digital-single-market/en/news/eur-2-million-prize-big-data-forecasting-system-will-enable-more-efficient-use-energy-resources>
- <https://www.crn.com/news/applications-os/300084738/the-2017-big-data-100.htm>
- <https://ko.wikipedia.org/wiki/SQL>

| KISTEP 기술동향브리프 발간 현황 |

발간호	제목	저자 및 소속
2018-01	블록체인	유거송(KISTEP), 김경훈(KISDI)
2018-02	독일의 연구개발 동향	이주석·김승연(KISTEP)
2018-03	휴먼 마이크로바이옴	황은혜·김은정(KISTEP) 남영도(KFRI)
2018-04	신육종기술(NPBTs)	박지현·홍미영(KISTEP) 한지학(㈜툴젠)
2018-05	2차원소재	함선영(KISTEP)
2018-06	이산화탄소 포집·저장·활용기술	김한해·배준희·정지연(KISTEP)
2018-07	줄기세포	김주원·김수민(KISTEP)
2018-08	일본의 연구개발 동향	유종태(KISTEP)
2018-09	AR/VR 기술	임상우(KISTEP), 서경원(UBC)
2018-10	미국의 연구개발 동향	강문상(KISTEP)
2018-11	빅데이터	김수연·도지훈·김보라(KISTEP)



| 저자 소개 |

김 수 연

한국과학기술기획평가원 미래성장전략센터 부연구위원

Tel: 02-589-2828 E-mail: suyoun@kistep.re.kr

도 지 훈

한국과학기술기획평가원 성장동력사업센터 연구원

Tel: 02-589-6115 E-mail: jhd0330@kistep.re.kr

김 보 라

한국과학기술기획평가원 미래성장전략센터 연구원

Tel: 02-589-2205 E-mail: kbora@kistep.re.kr

한국과학기술정보연구원(KISTI) 김성찬 박사님, 정보통신기술진흥센터(IITP) 조일구 팀장님
및 SK텔레콤 김정선 부장님의 자문으로 원고가 작성되었습니다.

※ 본 KISTEP 기술동향브리프의 내용은 필자의 개인적 견해이며, 기관의 공식적인 의견이 아님을 알려드립니다.

KISTEP 기술동향브리프 | 2018-11호

빅데이터