

빅데이터 기법을 활용한 Data Technology의 키워드 분석[†]

Keyword Analysis of Data Technology Using Big Data Technique

박성욱(Sung-Uk Park)*

목 차

- | | |
|------------------------------|-----------|
| I. 서 론 | IV. 분석 결과 |
| II. 빅데이터와 Data Technology 개요 | V. 결 론 |
| III. 키워드 분석 | |

국 문 요 약

경제가 성장하고 인터넷이 발전되면서 사람들의 경제형태와 소비는 많이 바뀌었다. 중국 알리바바 그룹은 모바일, 온라인, 오프라인, 인공지능을 결합한 플랫폼으로 약 28조의 매출을 창출하고 있다. 이는 1초에 약 25만건을 처리하는 수준이며, 2016년 대비 40% 증가했다. 이를 가능하게 한 핵심 기술은 소위 Data Technology라고 불리는 빅데이터와 클라우드 컴퓨팅이 융합된 기술이다. 기술의 발전속도에 비해 Data Technology에 관한 정확한 개념적 정의는 부족하다.

이에 본 논문은 빅데이터 분석기법인 TextTom을 활용하여 구글과 네이버의 최근 3개년(2015년 11월~2018년 11월) 신문기사를 데이터 마이닝 및 정제하여 'Data Technology' 키워드로 한정하여 관련 핵심 키워드를 도출하였다. 그 결과 빅데이터, O2O, 인공지능, 사물인터넷, 클라우드 컴퓨팅의 핵심 키워드 기술이 Data Technology와 관계가 있음을 알수 있었다. 본 연구의 분석결과는 향후 Data Technology 시대가 도래되면 참고할 수 있는 유용한 정보로 활용될 수 있다.

핵심어 : 빅데이터 기법, Data Technology, 네트워크 분석, 키워드 분석, 클라우드 분석

※ 논문접수일: 2019.2.18, 1차수정일: 2019.3.11, 게재확정일: 2019.3.20

* 한국과학기술정보연구원 정책연구실 선임연구원, supark@kisti.re.kr, 042-869-0925

† 본 연구는 한국과학기술정보연구원(KISTI) 주요사업 과제[K-19-L04-C03-S01]의 지원을 받아 수행하였으며, 박성욱 외 (2018)의 연구보고서를 요약, 재정리하였음.

ABSTRACT

With the advent of the Internet-based economy, the dramatic changes in consumption patterns have been witnessed during the last decades. The seminal change has led by Data Technology, the integrated platform of mobile, online, offline and artificial intelligence, which remained unchallenged.

In this paper, I use data analysis tool (TexTom) in order to articulate the definitfite notion of data technology from Internet sources. The data source is collected for last three years (November 2015 ~ November 2018) from Google and Naver. And I have derived several key keywords related to 'Data Technology'. As a result, it was found that the key keyword technologies of Big Data, O2O (Offline-to-Online), AI, IoT (Internet of things), and cloud computing are related to Data Technology. The results of this study can be used as useful information that can be referred to when the Data Technology age comes.

Key Words : Big Data Analysis, Data Technology, Network Analysis, Keyword Analysis, Cloud Analysis

I. 서 론

제4차 산업혁명 시대를 맞이하여, 빅데이터, 인공지능, 자율주행차 등 미래사회를 변화시킬 다양한 기술들이 사회에 적용되면서 혁신적인 변화를 유발하고 있다. 제4차 산업혁명 시대 핵심기술을 확보하고, 혁신성장동력을 육성하여 기술기반 창업을 활성화하여 과학기술이 경제성장과 일자리 창출을 견인하고 있다. OECD 국민 1인당 산업부가가치 순위는 2016년 18위에서 2022년 12위로 과학기술과 정보통신기반 일자리 창출은 2022년까지 26만개를 창출하는 계획을 세웠다(과학기술정보통신부, 2018). 제4차 산업혁명의 핵심은 사물인터넷, 인공지능, 3차원 인쇄, 로봇공학, 나노기술, 무인운송수단과 같은 6대 분야에서 혁신을 추진하고 있다. 제4차 산업혁명은 디지털, 생물학적, 물리적 세계를 빅데이터 활용으로 통합시키고 산업·경제 등 전 분야에 영향을 미치는 새로운 기술이다. 특히 과학기술적 측면에서는 사물인터넷, 인공지능, 클라우드기술, 빅데이터, 모바일 인터넷 등의 기술이 과학기술을 변화시키는 동인으로 될 것으로 예측하고 있다(Schwab, 2016; World Economic Forum, 2016; Wikipedia, 2018).

마윈은 Data Technology(DT)시대에는 IT기업도 전통적인 기업이 될 수 있음을 역설하며 IT에서 DT로의 변화의 중요성을 부각하였다. 미래 제조업은 제품을 생산할 뿐만 아니라 기계 역시 사고하고 소통할 수 있어야만 한다고 밝히며, 앞으로 모든 제조업은 인터넷과 빅데이터의 단말기업이 될 것으로 예측하였다(마윈, 2017). 마윈의 주장을 부각하지 않아도, 한국도 데이터가 산업발전과 가치창출에 촉매 역할을 하면서 새로운 가치를 창출하는 디지털 서비스의 원천이 되고 있음을 의식하고, 다양한 정책을 통해 데이터 자본의 부상을 지원하고 있다.

전통적 산업혁명 시대에서는 금전적 자본투자가 중요했지만 4차 산업혁명에서는 데이터가 핵심자본으로 각 기관, 기업이 공간에 쌓아둔 데이터를 자본으로 꺼내어 비즈니스 및 사회현안 해결에 적극 활용할 것을 주문하고 있으며, 또한 스타트업 및 벤처기업들이 혁신경제로 진출할 수 있도록 모든 산업분야에서 수집된 데이터를 재이용 가치자본으로 개방 및 유통 공유하여 인공지능(AI), 빅데이터 분석·가공 기업이 등장하도록 적극 나서야 함을 언급하고 있다(과학기술정보통신부, 2018).

아울러, 국가 과학기술정보인프라의 발전과 운영을 책임지고 있는 한국과학기술정보연구원에서도 ‘2015년 미래과학기술 정보 정책과 전략 연구 및 KISTI 성과확산 기반구축’ 연구를 통해 DT시대의 기관 발전 방안과 환경변화를 제시하였다(한국과학기술정보연구원, 2015). 이 보고서에서는 기존 IT의 의미적 한계를 넘어 데이터 중심의 주도적 역할 변화를 내세우고, 국가 및 사회적 기여측면에서 기관의 전략적 대응의 필요성을 도출하였다.

이에 본 논문에서는 구글과 네이버 뉴스 기사에서 언급된 ‘Data Technology’ 키워드가 실제

문헌에서 언급된 빅데이터 기술등과 연관성이 있는지 보고자 하였다.

본 논문의 제2장에서는 빅데이터와 Data Technology 개요에 대해 개략적으로 설명한다. 제3장에서는 키워드 분석방법에 대해 설명한다. 여기서는 빅데이터 분석기법인 Textom을 활용하여 구글과 네이버의 최근 3개년(2015년 11월~2018년 11월) 신문 기사를 데이터 마이닝 및 정제하여 ‘Data Technology’ 키워드로 한정하여 관련 핵심 키워드를 도출할 것이다. 제4장에서는 그 결과를 바탕으로 Data Technology의 핵심 키워드 기술을 정의하고, 마지막 제5장에서는 향후 정책적 시사점을 논의할 것이다.

II. 빅데이터와 Data Technology 개요

1. 빅데이터

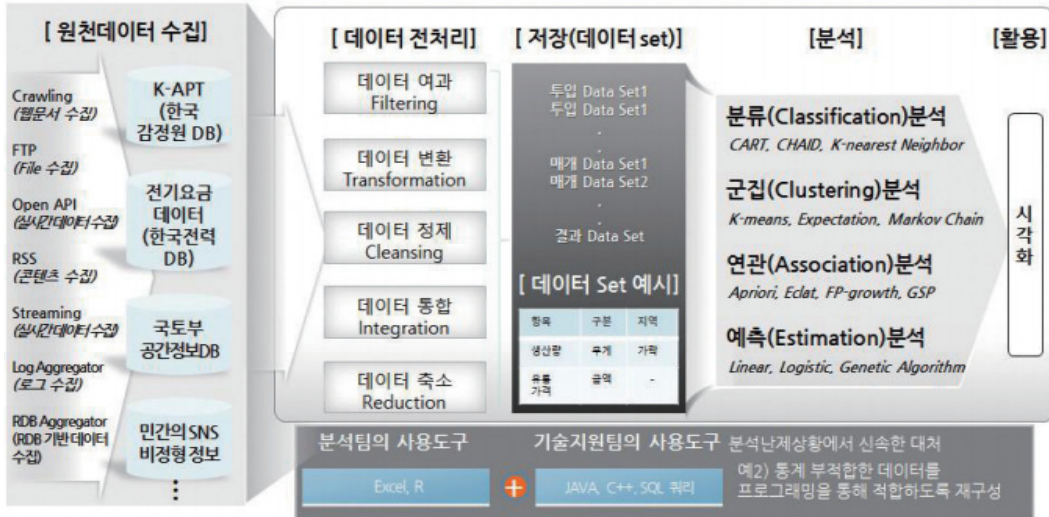
마윈 회장이 수차례 언급한 내용을 떠올리지 않아도 빅데이터(Big Data)는 기술 분야뿐 아니라 여러 다양한 분야에서 수많은 의미로 사용되고 있다. 특히 근래에는 Big Data가 전세계의 연계·통합된 환경에서 상거래 및 경제에 막대한 영향을 미칠 수 있어, 비즈니스 분야로 주목을 받고 있다.

빅데이터란 “기존 데이터베이스 관리 도구로 데이터를 수집, 저장, 관리, 분석 할 수 있는 능력을 넘어서는 대량의 정형 또는 비정형 데이터베이스 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술”이다(Sunil Soares, 2013; Marr, 2016; 지원철, 2017; 한국소프트웨어기술인협회 빅데이터전략연구소, 2017). 국가 정보화 전략 위원회에서는 빅데이터를 대용량 데이터를 활용·분석하여 가치 있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 기술이라고 정의 하였다(한국정보화진흥원, 2017).

빅데이터에 관한 다양한 정의를 종합해 보면 단지 거대한 데이터라기보다 형식이 다양하고 순환속도가 매우 빨라서 기존 방식으로는 관리·분석이 어려운 데이터를 의미한다(한국정보화진흥원, 2017). 빅데이터는 기업들이 오랫동안 고민해온 난제에 대한 해결책을 제공해 줄 뿐 아니라, 프로세스와 조직, 산업 전반, 심지어 사회 자체를 변화시킬 수 있는 새로운 방법까지 제시하고 있다(IBM 비지니스가치연구소, 2012). 또한, 빅데이터를 이용하여 고객 중심 성과를 추구하고, 내부 데이터를 활용하고, 바람직한 정보 생태계를 구축하고자 노력하고 있는 것으로 밝혀지고 있다(IBM 비지니스가치연구소, 2012).

이러한, DT 시대의 가장 핵심인 빅데이터 기술은 그간 학술적 정의와 시장에서의 정의 등으

로 구분되어져 왔으며, 기존 데이터 관리 방식과의 차이점과 분류를 통해 다음 (그림 1)과 같은 핵심 기술이 존재한다.¹⁾



자료 : 박성욱 외(2018)

(그림 1) 빅데이터의 지식도출 Framework

빅데이터는 규모(volume), 다양성(variety), 속도(velocity)의 특징을 가지고 있다. 여기서, 3V는 빅데이터의 핵심적인 특징을 포함하고 있지만, 기업들이 정확성(veracity)이라는 네 번째

<표 1> Big Data 특징

구분	내용
규모 (Volume)	데이터 양. 빅 데이터 하면 가장 먼저 연상되는 규모는 기업들이 전반적인 의사 결정 능력을 향상시키기 위해 활용하려 노력하는 데이터의 양을 의미
다양성 (Variety)	데이터의 다양한 형태와 소스. 다양성은 정형, 반정형, 비정형 데이터를 전부 포함하는 복잡하고 다양한 형태의 데이터를 관리하는 것과 관련
속도 (Velocity)	데이터의 이동. 데이터가 생산, 처리, 분석되는 속도도 지속적으로 증가하고 있으며, 데이터가 실시간으로 생성된다는 점과 스트리밍 데이터를 비즈니스 프로세스와 의사 결정 과정에 도입해야 한다는 점이 속도를 높이는데 기여
정확성 (Veracity)	데이터의 불확실성. 정확성은 일정 유형의 데이터에 부여할 수 있는 신뢰수준을 의미. 높은 데이터 품질을 유지하는 것은 빅 데이터의 중요한 요구사항임

자료 : IBM 비즈니스가치연구소, 2012(재정리)

1) 빅데이터 기술은 대용량의 데이터를 다룰 때, 여러 과정을 거치게 되는데, 데이터 수집 및 데이터 전처리, 저장, 분석, 활용(시각화)까지의 과정을 거치게 되며 각 과정별로 핵심 기술이 존재한다.

중요한 차원을 하나 더 고려하여 추가한 특징을 일반적으로 <표 1>로 통용하고 있다(IBM 비즈니스가치연구소, 2012).

미래 경쟁력의 핵심은 부동산과 화폐와 다름없는 가치를 지닌 데이터의 확보와 활용에 있음을 시사하고 있으며, 이에 데이터를 매개로 한 신생태계(New Ecosystem)인 ‘데이터 경제(Data Economy)’의 중요성이 부각되고 있다. 세계 시가총액 10위권 기업은 2018년 7월 기준으로 애플, 알파벳(구글의 모회사), 아마존, 페이스북, 알리바바, 텐센트 등 데이터 기반 혁신 기업이 대다수를 차지하였다. 이는 통신(차이나 모바일, AT&T)과 에너지(페트로차이나, 엑스모빌, 가스프롬)가 세계 10위권 안에 들었던 10년 전이나 20년전에 제조업(GE, 코카콜라) 우위의 산업 지형과 다른 양상을 보이고 있다. 기업뿐 아니라 공공 부문이 ‘열린 정부(Open Government)’와 ‘오픈 데이터(Open Data)’를 표방하며 데이터에 기초한 행정은 물론 데이터 공유와 활용을 통한 부가가치 창출과 경제 활성화에 시황을 걸고 있는 이유도 데이터 경쟁력 확보에서 공공 부문도 예외일 수 없음을 의미하고 있다(정용찬, 2018).

2. Data Technology

2015년 6월, 마윈(马云) 알리바바 회장은 “지난 20년간 지속된 IT의 시대가 저물고, 앞으로 30년간 Data Technology(DT, 数据技术) 혁명에 기반한 새로운 인터넷 시장이 열리게 될 것이며, 이제는 방대한 고객 데이터를 활용해 개별 고객의 요구에 부응할 줄 아는 기업이 성공하는 DT시대가 될 것”이라고 말했다. 더불어 마윈은 “정보기술시대를 넘어 앞으로 30년간 데이터 기술에 기반을 둔 새로운 시장이 열릴 것”이라고 했다(강준영, 2018).

DT를 특정 시대의 트렌드라고 부르기에는 DT가 일으킨 지각 변동이 너무나도 크다. 산업 전반, 사회 곳곳에서 게임의 규칙을 송두리째 바꾸고 있기 때문이다(WIKITREE 비즈, 2016).

미국과 OECD, IMF등의 국제기구는 2016년부터 최근 기술변화에 따른 실물 경제 변화를 측정하기 위한 시도로 ‘디지털 경제’라는 개념에서 추진해 왔다. DT 산업은 미국 상무성 경제 분석국(Bureau of Economic Analysis)의 연구결과를 통해 시장 동향을 파악할 수 있다. 아래 <표 2>에서 보듯이 디지털 경제는 경제의 디지털 변화를 초래하는 디지털 촉진(digital-enabling) 인프라와 전자상거래와 디지털 콘텐츠를 포함하고 있다. 디지털 촉진 인프라는 하드웨어, 소프트웨어, 통신, 지원 서비스 등 4가지 부분으로 이루어진다. GDP에서 차지하는 디지털 경제의 비중 변화를 살펴보면 2005년 6.1%에서 2016년 6.5%로 0.4%p 상승했다.

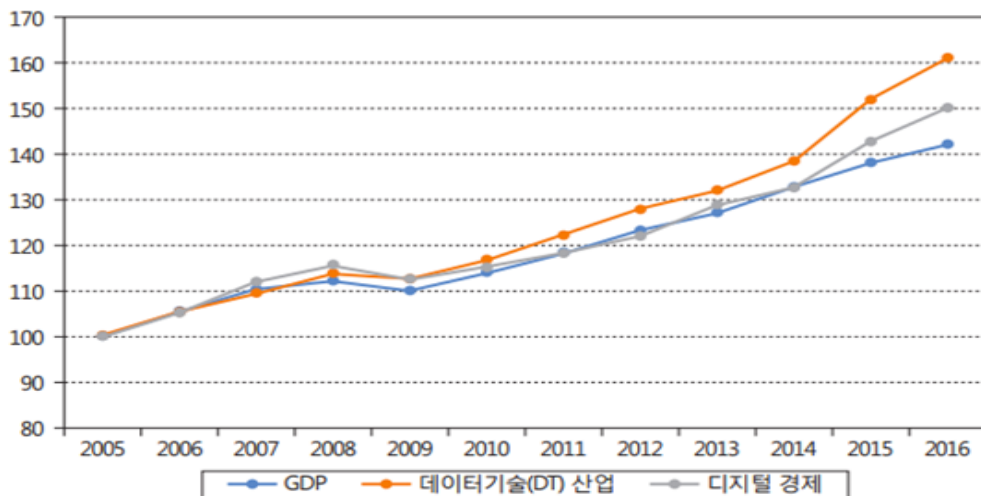
〈표 2〉 미국 디지털 경제의 변화 추이(부가가치기준)

단위 : 십억달러

	금액 기준			GDP 구성비		
	2005	2010	2016	2005	2010	2016
경제전체(GDP)	13,094.7	14,964.4	18,624.5	100.0	100.0	100.0
디지털경제	803.8	931.7	1,209.2	6.1	6.2	6.5
디지털축진인프라	742.6	847.1	1,072.6	5.7	5.7	5.8
H/W	104.0	113.9	131.3	0.8	0.8	0.7
S/W	160.0	185.2	258.8	1.2	1.2	1.4
통신	276.3	300.4	320.4	2.1	2.0	1.7
지원서비스	201.8	247.6	363.2	1.5	1.7	1.9
전자상거래와디지털 콘텐츠	61.2	84.6	136.5	0.5	0.6	0.7

자료 : Barefoot et al.(2018)

DT 산업의 발전이 어느 정도인지 평가하기 위하여 GDP, DT 산업, 디지털 경제의 2005년 부가가치가 100일때의 변화를 비교해 보면 (그림 2)에서처럼, DT 산업의 성장속도가 2009년부터는 지속적으로 빠른 것으로 보인다. 여기서 눈여겨봐야 할 점은 2015년부터 성장속도가 더 빨라졌다는 것이며, 이것은 미국의 경우 2015년부터 4차 산업혁명을 선도하는 DT 산업이 양적으로는 전체 경제성장을 주도 하고 있으며, 질적으로는 기존산업의 변화를 촉진하고 있다



자료 : Barefoot et al.(2018)

(그림 2) 전체 산업과 Data Technology산업의 성장추세 비교(2005=100)

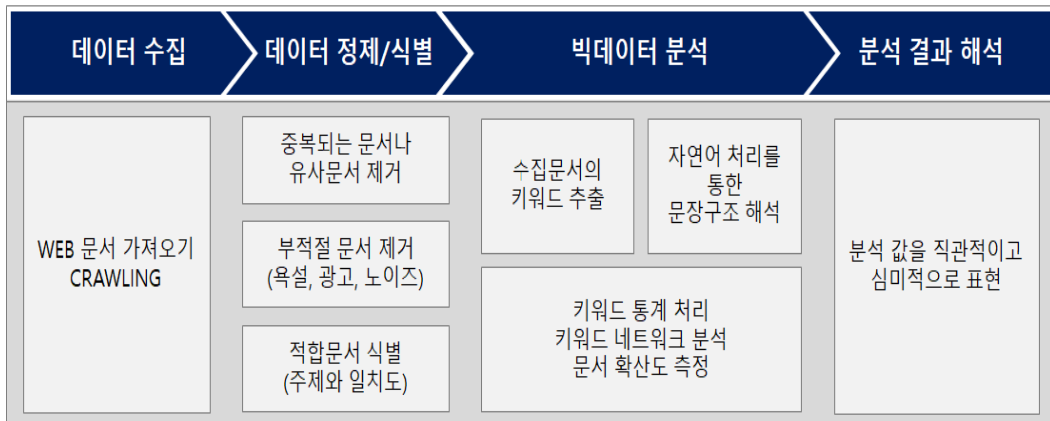
는 것을 알 수 있다. 한편, 2015년 부티의 급속한 성장을 구성 요소별로 살펴보면 2014년 대비 2016년의 경우 H/W는 25%, S/W는 15.4%, 지원 서비스는 14.7%가 증가했다.

이에 따르면 예상과 달리 H/W가 다른 두 분야보다 10% 포인트 이상의 큰 차이로 성장률이 높은 것으로 나타난다. S/W와 지원서비스는 15%대로 비슷한 성장률을 나타내고 있지만 GDP와 디지털 경제 전체보다는 더 높은 성장률을 기록 했다는 면에서 DT 산업의 역동성은 크다고 할 수 있다(장윤중, 2018).

III. 키워드 분석

본 논문에서는 Big Data를 통한 국내·외 트렌드 및 핵심 키워드 확인을 위해 텍스트(Textom)을 활용하며, 구글(Google)과 네이버(Naver)의 최근 3개년(2015년 11월 ~ 2018년 11월) 신문 기사를 데이터 마이닝 및 정제하여 결과를 확인 하였다.

텍스트는 수집된 데이터 중 연구의 주제와 맞지 않는 데이터의 삭제가 가능하며, 같은 의미의 단어를 정리하는 정제화를 수행한다(김해원·전채남, 2014). 본 논문에서는 수집된 데이터에서 Data Technology와 관련 없는 단어를 정제하였다.



자료 : 박성욱 외(2018)

(그림 3) 빅데이터 분석 방안

텍스트를 통해 대량의 데이터를 수집, 빈도분석, 의미연결망 분석 및 네트워크 중심성분석 등을 수행한 연구들은 다수 존재하나 Big Data에 의한 농촌 환경 의료 관심 트렌드 분석 등

‘Data Technology의 추세 및 트렌드’를 분석한 자료는 전무하다. 하지만, 수요자들의(소비자들의) 유인을 이끌어 내기 위한 노력으로 소비자들의 인식과 태도 및 트렌트 조사에 민감한 스포츠 분야에서 활발한 연구가 진행 중이다. 2018년 평창올림픽에 대한 빅데이터 키워드 분석을 통해 평창올림픽의 다양한 키워드 및 문제점을 파악했으며, 한진욱 외(2015)은 프로야구 SNS 밈(meme)분석을 통해 한국프로야구 SNS 밈이 무엇인지 규명하고 적합도를 조사하였다. 또한 박성건 외(2017)은 2016리우올림픽 축구 관련 이슈 분석을 통해 리우올림픽 개최 전·중·후 축구 관련 키워드 및 한국남자 축구 주요 이슈 등을 연구하면서 이처럼 스포츠 분야에서 빅데이터 분석을 활용한 연구가 증가하고 있다. 다만, ‘제4차 산업혁명의 현재와 미래’라는 주제로 키워드 네트워크 분석을 수행한 연구는 존재한다(양현재, 2017). 여기서는 사회적으로 논의가 증가하고 있고 이슈화 되고 있으나 일반적 담론에서 머무르고 있는 제4차 산업혁명과 관련한 내용들이 인터넷 사용자들 사이에서 어떤 관련성으로 논의 되었는지 키워드 네트워크를 통해 분석하였다. 또한, 조성환(2018)은 키워드 네트워크 분석에 사용되는 텍스트마이닝과 의미연결망 분석 방법을 활용하여 블록체인의 산업활용분야로 언론 및 정부발표에서 언급되고 있는 ‘금융’, ‘에너지’, ‘물류’를 언급한 기사들을 비교 분석하였다.

본 논문에서는 국내 검색점유율 1위를 차지하는 네이버와 전세계 1위의 검색사이트인 구글을 이용하였으며, 두 사이트는 뉴스 데이터마이닝에 가장 적합한 것으로 판단하였다. 분석채널은 “뉴스”로 선정하였으며, 자료검색을 위한 키워드는 ‘Data Technology’로 한정하였다. 전체 네트워크에서 중요한 역할을 할 수 있는 노드만을 사용하여 간략화하기 위해 상위 30개의 키워드를 중심으로 데이터의 노이즈 값을 조정하였다.

〈표 3〉 분석의 틀과 분석 항목

〈분석방법〉	〈분석항목〉	➔	〈분석 결과〉
기술통계분석	클라우드 분석(빈도분석)		Data Technology 관련 핵심키워드 도출
네트워크분석	키워드 네트워크 분석 등		

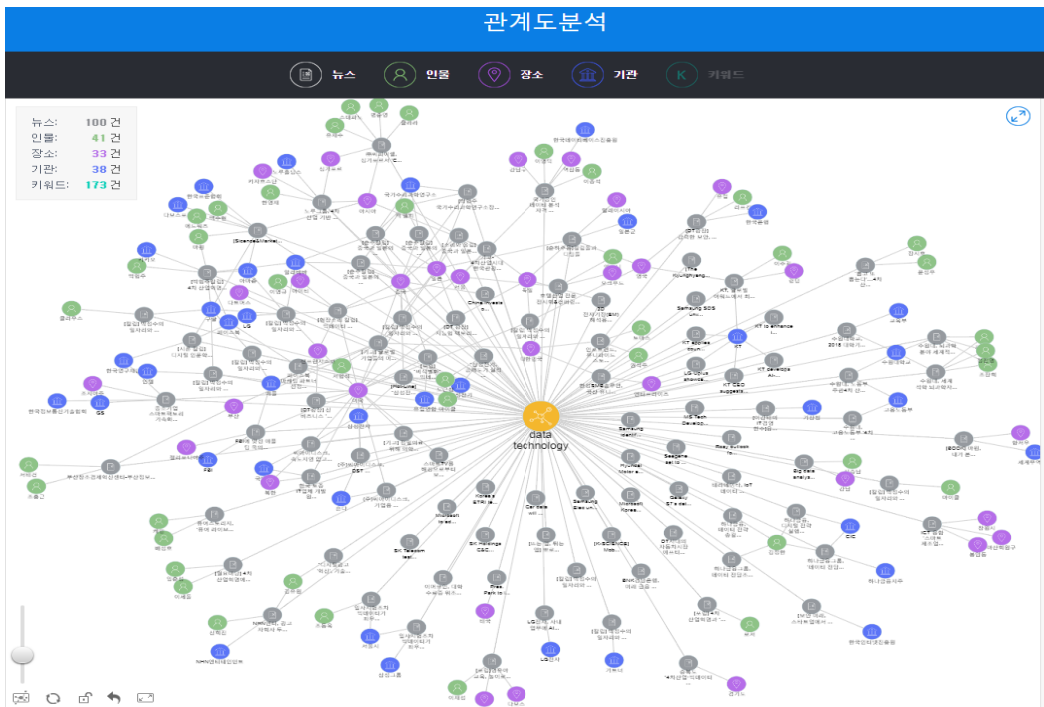
자료의 분석기간은 2015년 11월부터 2018년 11월까지로 설정하였으며, 이는 마윈 회장의 데이터 테크놀로지 발언 이후로 관련 기사 및 뉴스가 대부분 생성된 현실을 감안 하였다.

아울러, 구글 및 네이버의 뉴스 검색에 앞서 국내 주요 언론사를 포함한 방송국의 거시적 트렌드를 검토하기 위해, Data Technology 관련 3개년의 추이(빅카인즈, Bigkinds(2019))를 살펴보았다.

〈표 4〉 Data Technology 관련 3개년의 추이(빅카인즈) 개요

수집년도 : 2015년 11월 ~ 2018년 11월
- 검색 결과 뉴스 : 100 건 / 인물 : 41 건 / 장소 : 33 건 / 기관 : 38 건 / 키워드 : 173 건

관계도 분석의 총괄 (그림 4)를 살펴보면 검색 결과로 Data Technology는 제4차 산업혁명의 키워드를 중심으로 구글과 미국 중심의 키워드 형성이 연결되어 있는 것을 확인할 수 있다. 그림에서 회색은 뉴스, 연두색은 인물, 보라색은 장소, 파란색은 기관, 초록색은 키워드를 의미한다.



(그림 4) Data Technology의 관계도 분석

(그림 4)의 관계도 분석을 좀더 자세히 살펴보면 (그림 5)와 같다. 여기서 워드클라우드 분석은 크게 4가지 카테고리인 키워드, 인물, 기관, 장소로 나누어 설명하고 있다. 키워드는 산업혁명을 필두로 세계경제포럼, 삼성전자와 클라우드 및 알리바바 등이 다빈도 키워드임을 알 수 있다. 그리고 인물로는 잭웰치와 김정환, 마윈등이 다빈도 키워드로 등장하였고, 기관으로는

구글과 애플, KT 등의 빈도값이 높음을 알 수 있다. 마지막으로 장소로는 미국, 중국, 일본, 독일 등이 빈도값이 높음을 확인 할 수 있다.



(그림 5) 관계도 분석의 상세정보

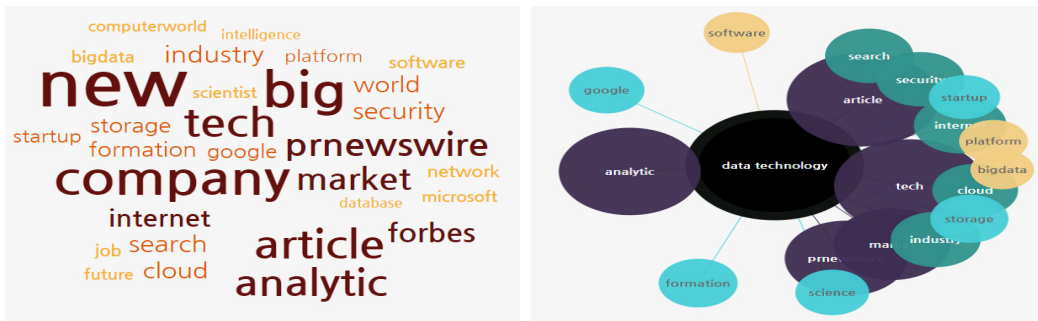
동일한 검색 결과를 활용한 키워드 트렌드를 살펴보면 검색기간인 2015년 11월 ~ 2018년 11월까지의 기사건수로 살펴보면 2017년 10월에 급격히 기사의 양이 증가하고, 언론의 관심이 높아지는 것을 (그림 6)에서 확인할 수 있다.

〈표 5〉 국내/외 뉴스 데이터 분석 개요

수집년도 : 2015년 11월 ~ 2018년 11월
DATA BASE : 구글 기사 / 네이버 기사
분석 TOOL : TEXTOM
분석키워드 : “Data Technology”

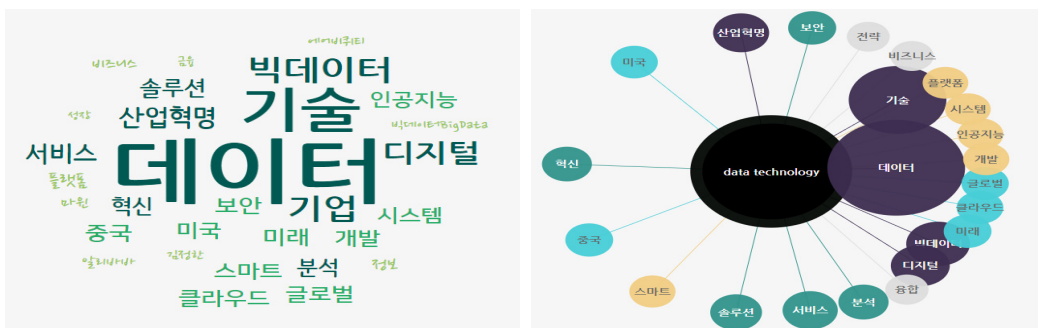
본 논문에서 사용한 최근 3년동안의 구글과 네이버를 통해 수집된 뉴스의 데이터양과 수집량은 구글은 4,989,000건과 1,024건이고, 네이버는 6,022건과 6,142건이다.

영문으로 뉴스를 검색하여 자료를 수집한 구글의 워드 클라우드 분석의 경우 아래의 (그림 8)과 같이 new, big, 기술, 기업, 시장, 논문, 분석, 빅데이터와, 소프트웨어 분석 툴과 새로운 스타트업의 출현 등이 자주 언급되는 단어로 나타났다.



(그림 8) 구글을 통한 워드 클라우드 분석

네이버의 경우 아래 (그림 9)와 같이 데이터, 기술, 빅데이터, 산업혁명, 디지털, 인공지능 등이 자주 언급되는 단어로 나타났다.



(그림 9) 네이버를 통한 워드 클라우드 분석

DT의 시대는 Big Key Word로 Big Data, O2O, AI, IoT, Cloud Computing라는 5가지 핵심 기술과 키워드를 빼놓고는 설명이 불가능하다. 데이터는 기계와 사람 간에만 창조되는 것이 아니라 사물과 사물 간(IoT), 기계와 기계간(M2M)에도 이뤄진다. 그리고 그 데이터를 저장하고 처리하는 공간은 개별 모바일기기에서 이뤄지지 않고 클라우드(Cloud)에서 이뤄지고 있다. 위의 5가지 핵심 키워드에 관한 내용을 정주용(2015)는 다음과 같이 정의하고 있다.

첫째, 빅데이터로 과거 데이터는 미개발의 영역이었다면 인공지능의 발달로 데이터를 효율적으로 가공할 수 있는 기술이 진보해 거대한 규모의 데이터에서 유의미한 정보, 지식, 인사이트를 뽑아낼 수 있게 되었다. 데이터의 효율적 분석은 인간의 행동패턴을 예측하는 것을 넘어 행동패턴을 조정할 수 있는 방법을 알려준다. 데이터를 가장 효율적으로 분석하고 가공하는 방법은 기계에게 인간의 합리성을 장착시키는 것이다. 그것을 머신러닝(Machine Learning)이라 부르며, 기계는 이제 단순한 정보의 연산자가 아닌 판단하는 주체가 되어가는 중이다.

둘째, O2O(Online to Offline)는 오프라인의 모든 서비스가 온라인으로 빨려 들어가는 현상이며 모바일 기기들이 오프라인에 대한 리모컨이 되어가는 현상이다. O2O는 중국에서 우버, 에어비앤비를 지칭하면서 사용한 용어지만, 본질적으로는 온디맨드(On Demand)를 지칭한다고 볼 수 있다. O2O는 기업-고객간 거래(B2C)에 무게중심이 놓여진 트렌드라면, O2O가 기업간 거래(B2B) 영역으로 확산되기 위해서 필수적인 개념이 바로 사물인터넷(Internet of Things)이다.

셋째, 인공지능(Artificial Intelligence, AI)은 데이터-정보-지식-인사이트의 데이터 정제작업을 사람의 손이 필요 없이 기계가 스스로 할 수 있게 되었다. 스스로 판단하고, 스스로 예측하며, 사람의 판단을 유도하는 단계까지 진화해나가고 있다. 인간 의사는 하나의 전공분야를 연구하기 위해 수십 년의 경험이 필요하지만, 인공지능은 수많은 전공분야를 혼자서 마스터하고 전공분야를 통섭하는 솔루션을 도출해 낼 수 있다. 인간 두뇌 한계를 극복하는 기계로 과거 산업혁명 시대에 기계의 놀라운 물리적 능력을 경험했다면, 인공지능이 만들어가는 새로운 세상에서 우리는 기계의 놀라운 지적 능력에 감탄하는 현실을 마주하고 있다.

넷째, 사물인터넷(Internet of Things, IoT)은 모든 것(Things)이 온라인으로 연결(Connected)되는 현상을 말한다. 사람 뿐 아니라 사물들도 모바일환경에 실시간 접속돼 전지구적인 온라인화를 가속화시킨다. O2O와 IoT 모두 핵심에는 데이터가 있다. 사람과 사람, 사람과 사물, 사물과 사물간에 교신되는 데이터 속에서 소비의 패턴이 숨어있고, 가치가 숨겨져 있기 때문이다. O2O IoT의 확산으로 수집되는 무수히 많은 데이터를 효율적으로 분석하고 정제해서 유의미한 지식, 인사이트로 정제하는 기술이 빅데이터라고 볼 수 있다.

마지막으로 클라우드 컴퓨팅(Cloud Computing)은 온라인의 가상공간에서 공유된 서비스,

저장공간을 통해서 컴퓨터의 저장하고 연산하는 기능을 수행하는 것이다. 이제 PC나 스마트폰에 저장하기 보다는 클라우드 공간에 저장하고 실시간으로 내려받는 것이 보편화된다.

수많은 데이터 속에서 패턴을 찾고 유의미한 정보-지식-지혜로의 고도화 가공작업을 기계가 대신해주려면 인공지능이 필요하고 머신러닝이 필요하다. 생산되고 가공된 정보, 지식은 다시 구름 속으로 빨려 들어간다. O2O, IoT, Big Data, AI, Cloud Computing, 5가지 키워드는 상호 연관적으로 얽혀 DT시대를 새롭게 형성해나가고 있다.

V. 결 론

마윈이 예상한 DT 시대가 도래하면 방대한 고객 데이터를 활용해 개별 고객의 요구에 부응할 줄 아는 기업이 미래 30년을 이끌 것이다. 고객 데이터, IT기술이 결합한 고객 중심적인 서비스 분야에서 새로운 사업들이 생겨날 것으로 예상된다.

본 논문은 빅데이터 분석기법인 Textom을 활용하여 구글과 네이버의 2015년 11월부터 2018년 11월까지 신문기사를 데이터 마이닝 및 정제하여 마윈이 주장하고 있는 Data Technology 키워드로 한정하여 관련 핵심 키워드를 도출하였다. 그 결과, 기존 문헌에서 설명하고 있는 것처럼, 빅데이터, O2O, 인공지능, 사물인터넷, 클라우드 컴퓨팅의 핵심 키워드 기술이 Data Technology와 관계가 있음을 알수 있었다.

한국 기업은 미래 DT시대의 중국 신흥시장을 예측하고 DT시대와 관련 각종 신흥 산업분야에서 한국 중소기업들이 적극적으로 협업할 수 있는 틈새시장 등과 전략을 준비해야 한다. DT 시대에는 빅데이터가 미래 30년을 책임지고 있다고 말할 만큼 많은 주목을 받고 있고 향후 시장에 큰 영향을 미칠 것으로 예상되고 있다. 하지만 아직 현실적으로 몇 몇의 대기업을 제외하면 많은 중소기업들이 빅데이터를 활용하는 것은 어려움이 있다. 특히 보안 문제, 예산 문제, 분석 역량 및 전문가 인재의 부족 등은 빅데이터 활용을 더욱 어렵게 만드는 요소이므로 정부 차원의 적극적인 지원이 필요하며, 기업들이 정부 서비스를 적극 활용할 것을 장려하고 있다. 빅데이터 역량은 곧 데이터 분석 과학자의 확보 문제와 직결된다. 빅데이터의 숨겨진 가치를 찾아내고 의미 있는 통찰력을 제시할 수 있는 인재를 육성시켜야 한다. 빅데이터 분석을 잘하고 있는 해외 글로벌 기업들은 빅데이터 분석을 외부에 의뢰하기 보다는 데이터 과학자 조직을 구성하고 있는 추세이다. 따라서 정부의 적극적인 빅데이터 분석 인재 양성을 통해 한국 국내 빅데이터 시장의 활성화뿐만 아니라 중국 등과 같이 DT시대 신흥시장도 미리 선점해야 할 것이다.

본 논문을 통해 향후 Data Technology와 관련하여 주력해야 할 핵심기술 키워드를 도출하

있으며, 향후 데이터 경제에 대한 산업의 방향에 참고자료로 사용될 수 있을 것으로 판단된다. 그러나 본 논문은 소셜 네트워크 분석을 위해 연결정도 중심성, 매개중심성, 군집계수, 포괄성 등 다양한 지표가 있는데 사용하지 않았고, Data Technology 키워드 추출을 구글과 네이버의 기사내용으로만 한정하여 사용한 한계점이 있다. 이에 추후 후속연구에서는 위의 다양한 지표를 논문이나 특허로 확대하여 분석한다면 의미 있는 연구가 될 수 있을 것이라 생각한다.

참고문헌

- 강준영 (2018), “데이터 기술 기반의 새 시장이 열린다”, 「KDI 나라경제」, 336: 10-11.
- 과학기술정보통신부 (2018), “NTIS 국제 ICT 어워드 공공부문 최우수상 수상”.
- 김해원·전채남 (2014), “빅데이터를 활용한 콘텐츠 제작방안에 관한 탐색적 연구”, 「사이버커뮤니케이션 학보」, 31(3): 5-51.
- 마윈 (2017), 「마윈, 내가 본 미래」, 서울 : 김영사.
- 박성건·이수원·황영찬 (2017), “소셜 빅데이터 기반 2016리우올림픽 축구 관련 이슈 및 인물에 대한 연관단어 분석”, 「한국체육학회지」, 56(2): 303-320.
- 박성욱·박지영·이영구 (2018), 「Data Technology와 과학기술정보인프라 정책연구」, 대전 : 한국과학기술정보연구원.
- 양현채 (2017), “4차 산업혁명의 현재와 미래 : 키워드 네트워크 분석을 중심으로”, *STEPI WORKKING PAPER SERIES*, WP 2017-02.
- 장윤중 (2018), 「4차 산업혁명 주도 데이터기술 산업의 변화와 한국의 과제」, 세종 : 산업기술연구원 산업경제.
- 정용찬 (2018), 「4차 산업혁명 시대의 데이터 거버넌스 개선 방향」, 진천 : 정보통신연구원.
- 정주용 (2015), “DT 시대의 5가지 키워드를 연결하자”, MOBIINSIDE, <http://www.mobiinside.com/kr/2015/12/14/dt-5keyword>.
- 조성환 (2018), “키워드 네트워크 분석 방법을 활용한 블록체인 트렌드 분석에 관한 연구”, 「한국정보전자통신기술학회논문지」, 18(10): 550-555.
- 지원철 (2017), 「빅데이터 시대의 데이터 마이닝」, 서울 : 민영사.
- 한국과학기술정보연구원 (2015), 「미래과학기술 정보 정책과 전략 연구 및 KISTI 성과확산 기반 구축」, 대전 : 한국과학기술정보연구원.
- 한국소프트웨어기술인협회 빅데이터전략연구소 (2017), 「빅데이터 개론(Big Data)」, 서울 : 광

문각.

- 한국정보화진흥원 (2017), 「지능정보사회 법제도 정립방향」, 대구 : 한국정보화진흥원.
- 한진욱·안정찬·오승욱·신동일 (2015), “소셜 빅데이터를 통한 한국 프로야구 SNS 밈(meme) 분석”, 한국스포츠산업경영학회지, 20(5): 1-16.
- Barefoot, K., Curtis, D., Jolliff, W. A., Nicholson, J. R. and Omohundro, R. (2018), *Defining and Measuring the Digital Economy*, Bureau of Economic Anlysis.
- BIGKINDS (2019), <https://www.bigkinds.or.kr>.
- Feldman, R. and Dagan, I. (1995), “Knowledge Discovery in Textual Databases (KDT)”, *KDD-95 Proceedings*, AAAI: 112-117.
- Hotho, A., Nürnberger, A. and Paaß, G. (2005), A Brief Survey of Text Mining, LDV FORUM-Band, 20: 19-62.
- IBM 비즈니스가치연구소 (2012), 「분석 : 빅데이터의 현실적인 활용」, IBM.
- Marr, B. (2016), *Big Data in Practice*, Wiley.
- Miller, R. E. and Blair, P. D. (2009), *Input-Output Analysis: Foundations and Extensions*, 2nd ed, Cambridge University Press, New York.
- MSIT WEBZINE (2018), “데이터를 가장 잘 쓴 나라를 만들기 위해”, <https://www.msit.go.kr/webzine/posts.do?postIdx=330>.
- Schwab, K. (2016), *The Fourth Industrial Revolution*, World Economic Forum.
- Sunil Soares (2013), *Big Data Governance*, MC Press.
- Wikipedia (2018), Fourth Industrial Revolution, https://en.wikipedia.org/wiki/Fourth_Industrial_Revolution.
- World Economic Forum (2016), “The Fourth Industrial Revolution: What it means, how to respond”, <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>.
- WIKITREE 비즈 (2016), “[데이터 테크놀로지 시대 1부] 데이터는 홍수를 일으켰다”, http://www.wikitree.co.kr/main/news_view.php?id=251876.

박성욱

전남대학교에서 경제학으로 박사학위를 취득하고 현재 한국과학기술정보연구원 선임연구원으로 재직중이다. 또한, 한국연구재단 기술사업화(기술금융분야) 전문위원으로 위촉되었다, 관심분야는 과학기술정책, 기술경제, 경제성분석, 기술기획/평가, 기술가치평가 등이다.