

# 빅데이터 분석을 활용한 4차 산업혁명 키워드에 대한 통찰

남수태\* · 진찬용\*

\*원광대학교(융복합창의연구소)

A Insight Study on Keyword of 4th Industrial Revolution Utilizing Big Data

Soo-Tai Nam\* · Chan-Yong Jin\*

\*Wonkwang University (Institute of Convergence and Creativity)

E-mail : stnam, jcy85366@wku.ac.kr

## 요 약

빅데이터 분석은 데이터베이스에 잘 정리된 정형 데이터뿐 아니라 인터넷, 소셜 네트워크 서비스, 모바일 환경에서 생성되는 웹 문서, 이메일, 소셜 데이터 등 비정형 데이터를 효과적으로 분석하는 기술을 말한다. 대부분의 빅데이터 분석 기술 방법들은 기존 통계학과 전산학에서 사용되던 데이터 마이닝, 기계 학습, 자연 언어 처리, 패턴 인식 등이 이에 해당된다. 글로벌 리서치 기관들은 빅데이터를 2011년 이래로 최근 가장 주목받는 신기술로 지목해오고 있다. 따라서 대부분의 산업에서 기업들은 빅데이터의 적용을 통해 가치 창출을 위한 노력을 기하고 있다. 본 연구에서는 다음 커뮤니케이션의 빅데이터 분석도구인 소셜 매트릭스를 활용하여 2017년 5월, 1개월 시점을 설정하고 “4차 산업혁명” 키워드에 대한 소비자들의 인식들을 살펴보았다. 빅데이터 분석의 결과는 다음과 같다. 첫째, 4차 산업혁명 키워드에 대한 연관 검색어 1위는 “후보”가 빈도수(7,613)인 것으로 나타났다. 둘째, 연관 검색어 2위는 “안철수”가 빈도수(7,297), 3위는 “문재인”이 빈도수(5,183)로 각각 나타났다. 다음으로 “4차 산업혁명” 키워드에 대한 검색어 긍정적 여론 빈도수 1위는 새로운(895)으로 나타났고, 부정적 여론 빈도수 1위는 위기(516)가 차지하였다. 이러한 결과 분석결과를 바탕으로 연구의 한계와 시사점을 제시하고자 한다.

## 키워드

4차 산업혁명, 키워드, 통찰, 빅데이터, 데이터 마이닝

### 1. 서 론

빅데이터 분석은 기존 데이터 웨어하우스 관리 도구에 저장된 대용량 데이터 집합 속에서 의미 있는 새로운 상관관계, 패턴 그리고 추세를 발견하여 새로운 가치를 창출하는 과정이다. 또한 대량의 정형 및 비정형 데이터 집합으로부터 새로운 가치를 추출하고 결과를 분석하는 기술을 의미한다. 대부분의 빅데이터 분석 기술과 방법 중에는 기존 통계학 및 전산학에서 사용되던 데이터마이닝, 기계학습, 자연언어처리, 패턴인식 등이 이에 해당된다[1]. 글로벌 리서치 기관들은 빅데이터를 2011년 이래로 최근 가장 주목받는 신기술로 지목하고 있다. 따라서 대부분 산업의 기업들은 빅데이터의 적용을 통해 새로운 가치를 창출을 위한 노력을 기울이고 있다. 오늘날 스마트 기기의 대중화는 시간과 공간을 초월한 인터넷 사용의 대중화를 가능하게 하였으며 이를 통해 사람과 사람은 물론 사람과 사회를 연결하는 소

셜 네트워크 서비스를 발달시키는 유인이 되었다. 소셜 네트워크를 통해 사람들은 다른 사람들과 자유롭게 정보를 공유 할 수 있게 되었을 뿐만 아니라 다양한 정보를 언제 어디서나 손쉽게 얻을 수 있게 되었다. 이러한 사회적 현상은 사람들의 삶 속에 없어서는 안 될 가장 중요한 부분으로 자리매김하게 되었고 지금 이 순간에도 수많은 사람들은 다양한 정보를 여러 가지 형태의 데이터로 생성 및 저장하여 공유하고 있다. 이제 더 나아가 대중들이 만들어 놓은 자료와 정보를 필요와 목적에 맞게 다시 데이터화하고 그 데이터를 분석함으로써 사람들의 행동방식과 심리상태를 예측할 수 있는 단계에까지 도달했다고 할 수 있다[2]. 제2의 산업혁명기라고 할 수 있는 시대 고갈되지 않는 자원으로 불리는 빅데이터의 시대가 시작된 것이다. 정보통신기술이 빠른 속도로 발전하고 있으며 이와 더불어 유무선 통신이 가능한 높은 성능을 지니고 있는 다양한 스마트 기기들이 개발되고 실생활에 적용됨에 따라 차세

대 인터넷 핵심기술로서 사물인터넷의 중요성이 강조되고 있다. 기존의 인터넷은 사람이 만들어 낸 데이터를 가공한 정보와 지식을 공유하기 위한 공간으로 활용되어왔다. 하지만 사물인터넷으로 대변되는 미래의 인터넷에서는 유무선 네트워크를 통해 수많은 사물들이 인터넷에 연결되어 정보에 접근하고 새로운 정보를 생산 및 공유함에 따라 이전에 없던 새로운 차원의 경제적 가치를 창출할 것으로 기대된다.

## II. 연구방법

본 연구는 빅데이터를 분석하여 대중들이 연상하는 “4차 산업혁명”에 대한 연관 이미지 인물 그리고 긍정 부정적 여론 성향에 대해 비교 분석하는 것이 목적이다. 이를 위하여 연구자가 연구하는 현 시점인 2017년 5월 1개월의 특정 시점을 설정하였다. 빅데이터의 개념은 학자들마다 다양하게 정의하고 있다. 삼성경제연구소는 과거의 관리 및 분석 시스템으로는 처리할 수 없는 방대한 데이터의 집합으로 데이터의 수집, 저장, 검색, 공유, 분석 그리고 시각화 등도 빅데이터의 범주에 포함시키고 있다. 빅데이터 중에서도 소셜 네트워크 서비스는 대중들의 생각을 잘 나타내는 매체이다. 소셜 네트워크 서비스로 대표되는 사이트는 blog, Facebook, twitter, kakaotalk, Instagram 등이 있다. 그런데 카카오톡은 비공개 소셜 네트워크 서비스로서 분석이 불가능하며 페이스북은 공개적인 것과 비공개적인 것이 분리되어 있지만 대부분 비공개이므로 수집 및 분석이 불가능하다. 따라서 본 연구에서는 정보의 접근이 용이한 네이버와 다음에서 제공하는 블로그와 트위터를 분석대상으로 하였다.

## III. 빅데이터 분석 및 결론

본 연구에서는 소셜 매트릭스 프로그램을 이용하여 “4차 산업혁명”에 대한 대중들의 인식들을 살펴보기 위해 빅데이터 분석을 실시하였다. 이를 위하여 모집단을 기반으로 한 직, 간접적인 설문 조사와 같은 방법과 달리 사람들의 블로그 및 트위터와 같은 소셜 데이터를 기반으로 하는 소셜 매트릭스 프로그램을 이용하고자 하였으며 네트워크 구조 분석을 위하여 다음커뮤니케이션의 소셜 매트릭스를 활용하여 분석하였다. 소셜 매트릭스 프로그램은 검색어에 대한 연관 탐색어 100위까지 제시하여 분석할 수 있도록 제공한다. 본 연구에서는 검색어에 대한 탐색어들을 선택에서 인물, 단체, 장소, 상품, 속성 등으로 분석하였으며 그림 1은 2017년 04월 06일에서 05월 06일까지 1개월 동안의 탐색어에 대한 상위 10위까지를 제시 하였다. 따라서 본 연구가 특정한 기간과 시점의 연구인만큼 다른 기간과 시점에 대한 적용은

불가능하다.



그림 1. “4차 산업혁명” 키워드에 대한 연관어 맵

표 1. “4차 산업혁명” 긍정/부정 여론 빈도수

Ranking	Populace opinion	Category	Frequency
1	새로운	중립	895
2	특템	긍정	669
3	그레이	기타	585
4	다양함	중립	560
5	위기	부정	516
6	잘어울리다	긍정	448
7	적극적	긍정	424
8	필요하다	중립	416
9	강조하다	중립	391
10	좋은	긍정	384

본 연구가 진행되고 있는 시점이 대한민국 19대 대선 선거기간이라는 점을 감안하여 바라봐야 정확한 해석이 가능하다. 본 연구의 취지와 다르게 나타난 결과는 한편으로는 의미를 찾아볼 수도 있다는 연구자의 판단으로 계속 진행하였다. 따라서 그러한 관점으로 접근하여 4차 산업혁명 키워드에 대한 의미를 찾아볼 수 있다. 4차 산업혁명은 일자리 창출에 목적을 두고 있다. 그런데 최근 선거기간에 대선후보 간의 쟁점이 일자리 문제를 어느 후보가 잘 해결해줄 수 있을까?, 이와 같은 쟁점 연관어 인물에서 1위, 2위가 “안철수”, “문제인”이 차지한 것은 당연하다고 말할 수 있겠다. 이러한 빅데이터 분석에 대한 정확함에 아무도 부인할 수 없을 것이다. 다음으로 속성에서도 “후보”인 것으로 나타나 선거기간이라는 것을 감안하면 정확이 설명된다. 4차 산업혁명을 주도하는 4대 기술 분야로 사물인터넷, 클라우드 컴퓨팅, 빅데이터 그리고 모바일이라고 한다. 더불어 인공지능, 생명과학, 무인이동체 그리고 로봇 기술 분야가 산업혁명을 주도하는 산업 군으로 분류하고 있다. 이러한 관점에서 연관 검색어 상

위를 차지한 “전문가”, “경제”, “기술”, “정책”, “교육”이 차지한 것은 나설지 않다. 또한 이러한 연관 검색어 빈도수(2,560-3,488) 범위 내에 속함으로 볼 때 어느 키워드가 중요도가 높고 어느 키워드가 낮음을 가늠하기 힘든 일자리 창출과 관련된 키워드임을 빅데이터 분석을 통해 증명되었다.

#### 참고문헌

- [1] O. K. Yu, D. G. Kim, & S. T. Nam, “A Preference and Insight Study of Populace for Jeonju Moju Utilizing Big Data Analysis,” in Proceeding of Women’s ICT Committee Conference, pp. 33-37, 2016.
- [2] O. K. Yu, C. Y. Jin and S. T. Nam, A Pattern Study on Keyword of the Collagen through Utilizing Big Data Analysis, in Proceeding of the 40th Fall Korea Institute of Information and Communication Engineering, Daejeon, pp. 85-86, Oct. 2016.