

빅 데이터 분석을 활용한 스마트폰 플랫폼 키워드에 대한 패턴

진찬용* · 남수태*

*원광대학교

A Pattern on Keyword of the Android through Utilizing Big Data Analysis

Chan-Yong Jin* · Soo-Tai Nam*

*Wonkwang University

E-mail : jcy85366@wku.ac.kr

요 약

빅 데이터 분석은 기존 데이터베이스 관리 도구로부터 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 말한다. 대부분의 빅 데이터 분석 기술 방법들은 기존 통계학과 전산학에서 사용되던 데이터 마이닝, 기계 학습, 자연 언어 처리, 패턴 인식 등이 해당된다. 최근 스마트 기기의 발달과 정보통신 기술의 발전은 트위터, 페이스북, 인스타그램 등의 소셜 네트워크상에서 유통되는 정보량이 폭발적 증가하고 있다. 이러한 변화는 데이터화가 가속화되고 있는 현대사회에서 데이터의 가치는 점점 높아질 것으로 예상되며, 데이터로부터 가치 있는 정보와 통찰력을 효과적으로 이끌어내는 기업이 경쟁력 확보를 위한 핵심가치가 되었다. 본 연구에서는 다음 커뮤니케이션의 빅 데이터 분석도구인 소셜 매트릭스를 활용하여 키워드 분석을 통해 스마트폰 플랫폼 키워드 의미를 분석하고자 한다.

키워드

Big data, Social network service, Android, Keyword, Pattern

I. 서 론

스마트폰의 대중화는 시간과 공간을 초월한 인터넷 사용의 대중화를 가능하게 하였으며 이를 통해 사람과 사람은 물론 사람과 사회를 연결하는 소셜 네트워크 서비스(social network service)를 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 된 것이다. 이러한 사회적 현상은 사람들의 삶 속에 없어서는 안 될 가장 중요한 부분으로 자리매김하게 되었고 지금이 순간에도 수많은 사람들은 다양한 정보를 여러 가지 형태의 데이터로 생성 및 저장하여 공유하고 있다. 이제는 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 산업 혁명기라고 할 수 있는 시대 고갈되지 않는 자원으로 불리는 빅 데이터(big data)의 시대가 시작된 것이다. 정보통신기술이 빠른 속도로 발전하고 있으며 이와 더불어 유무선 통신이 가능한 높은 성능을 지니고 있는 다양한 스마트 기

기들이 개발되고 실생활에 적용됨에 따라 차세대 인터넷 핵심기술로서 사물인터넷(internet of things)의 중요성이 강조되고 있다. 기존의 인터넷은 사람이 만들어 낸 데이터를 가공한 정보와 지식을 공유하기 위한 공간으로 활용되어왔다. 하지만 IoT로 대변되는 미래의 인터넷에서는 유무선 네트워크를 통해 수많은 사물들이 인터넷에 연결되어 정보에 접근하고 새로운 정보를 생산 및 공유함에 따라 이전에 없던 새로운 차원의 응용 서비스와 경제적 가치를 창출할 것으로 기대되고 있다[2].

II. 연구방법

본 연구는 빅 데이터를 분석하여 대중들이 연상하는 안드로이드에 대한 연관 이미지 인물 그리고 여론 성향에 대해 비교 분석하고자 한다. 이를 위하여 연구자가 연구하는 현 시점인 2015년 8월과 2016년 5월 10개월의 특정 시점을 설정하였다. 빅 데이터의 개념은 학자들마다 다양하게 정의하고 있다. 삼성경제연구소는 과거의 관리 및 분석 시스템으로는 처리할 수 없는 방대한 데이터의

집합으로 데이터의 수집, 저장, 검색, 공유, 분석, 시각화 등도 빅 데이터의 범주에 포함시키고 있다[3]. 빅 데이터 중에서도 SNS는 대중들의 생각을 잘 나타내는 매체이다. SNS로 대표되는 사이트는 블로그(blog), 페이스북(Facebook), 트위터(twitter), 카카오톡(kakaotalk), 인스타그램(Instagram) 등이 있다. 하지만 카카오톡은 비공개 SNS로서 분석이 불가능하며 페이스북은 공개적인 것과 비공개적인 것이 분리되어 있지만 대부분 비공개이므로 수집 및 분석이 불가능하다. 따라서 본 연구에서는 정보의 접근이 용이한 네이버(naver)와 다음(daum)에서 제공하는 블로그와 트위터를 분석대상으로 하였다.

III. 연구결과

본 연구에서는 소셜 매트릭스 프로그램을 이용하여 안드로이드 플랫폼에 대한 소비자들의 인식들을 살펴보고자 하였다. 이를 위하여 직, 간접적인 설문조사와 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.

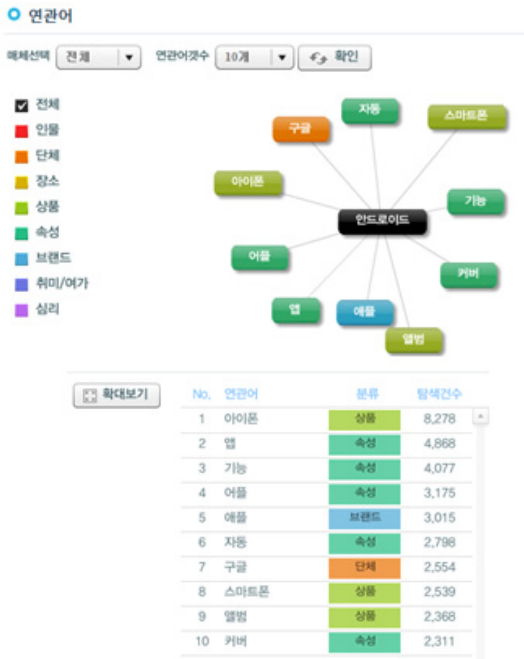


그림 1. 안드로이드 탐색 연관어 빈도(2015년 8월)

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 분

석하였으며 그림 1은 2015년 8월 11일에서 9월 11일까지 1개월 동안의 탐색어에 대한 상위 10위까지 제시 하였다. 따라서 본 연구가 특정한 기간과 시점의 연구인만큼 다른 기간과 시점에 대한 적용은 불가능함을 고한다.



그림 2. 안드로이드 탐색 연관어 빈도(2016년 5월)

2015년 8월 대중들은 안드로이드에 대한 전체 탐색어는 그림 1에서와 같이 아이폰과 앱 그리고 스마트폰의 기능과 관련하여 생각하는 것을 확인할 수 있었다. 2016년 5월 대중들은 안드로이드에 대한 전체 탐색어는 그림 2에서와 같이 아이폰과 앱 그리고 동조림, 소년을 연관하여 생각함을 알 수 있다. 최근 유행하는 게임 동조림소년의 인기를 높다는 것을 반영하고 있다. 다음으로 키워드 아이폰에 대한 빈도수의 변화이다. 8,278건에서 22,690건으로 약 14,000건 증가하였음을 알 수 있다. 다음으로 키워드는 투표이다. 2016년 4월 치른 20대 국회의원 선거 때문인 것이 틀림없다. 이러한 변화는 무엇을 의미하는 것인지 알 수 있다. 최근 빅 데이터의 중요성은 다양한 종류의 데이터를 실시간으로 분석함으로써 객관적이고 올바른 의사결정이 가능하도록 하는데 많은 도움이 될 뿐만 아니라 일정한 흐름을 포착해 미래의 예측이 가능하다는데 있다고 말할 수 있다. 아직도 초기 단계에 머물고 있는 빅 데이터에 대한 연구는 다양한 방법으로 진행될 것으로 판단된다.

*지면 관계상 참고문헌이 생략함을 고합니다.