

빅데이터 분석을 활용한 콜라겐 키워드에 대한 패턴

유옥경* · 진찬용** · 남수태**

*전북대학교 · **원광대학교

A Pattern Study on Keyword of the Collagen through Utilizing Big Data Analysis

Ok-Kyeong Yu* · Chan-Yong Jin** · Soo-Tai Nam**

*Chonbuk National University · **Wonkwang University

E-mail : okokyu@jbnu.ac.kr

요 약

빅데이터 분석은 기존 데이터베이스 관리 도구로부터 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 말한다. 또한 대량의 정형 또는 비정형 데이터 집합으로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다. 대부분의 빅데이터 분석 기술 방법들은 기존 통계학과 전산학에서 사용되던 데이터 마이닝, 기계 학습, 자연 언어 처리, 패턴 인식 등이 해당된다. 글로벌 리서치 기관들은 빅데이터를 2011년 이래로 최근 가장 주목받는 신기술로 지목해오고 있다. 따라서 대부분의 산업에서 기업들은 빅데이터의 적용을 통해 가치 창출을 위한 노력을 기울이고 있다. 본 연구에서는 다음 커뮤니케이션의 빅데이터 분석도구인 소셜 매트릭스를 활용하여 키워드 분석을 통해 콜라겐 키워드에 대한 의미를 분석하고자 한다. 또한 분석결과를 바탕으로 실무적 시사점을 제시하고자 한다.

키워드

Big data, Social network service, Collagen, 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 산업혁명기라고 할 수 있는 시대 고갈되지 않는 자원으로 불리는 빅데이터(big data)의 시대가 시작된 것이다. 정보통신기술이 빠른 속도로 발전

하고 있으며 이와 더불어 유무선 통신이 가능한 높은 성능을 지니고 있는 다양한 스마트 기기들이 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 기대된다[2].

II. 연구방법

본 연구는 빅데이터를 분석하여 대중들이 연상하는 창조경제 관한 연관 이미지 인물 그리고 여론 성향에 대해 비교 분석하고자 한다. 이를 위하여 연구자가 연구하는 현 시점인 2016년 10월 1개월의 특정 시점을 설정하였다. 빅데이터의 개념은 학자들마다 다양하게 정의하고 있다. 삼성경제

연구소는 과거의 관리 및 분석 시스템으로는 처리할 수 없는 방대한 데이터의 집합으로 데이터의 수집, 저장, 검색, 공유, 분석, 시각화 등도 빅데이터의 범주에 포함시키고 있다[3]. 빅데이터 중에서도 SNS는 대중들의 생각을 잘 나타내는 매체이다. SNS로 대표되는 사이트는 blog, Facebook, twitter, kakaotalk, 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 네이버와 다음에서 제공하는 블로그와 트위터를 분석대상으로 하였다.

III. 연구결과

본 연구에서는 소셜 매트릭스 프로그램을 이용하여 콜라겐에 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. 소셜 매트릭스 프로그램은 검색어에 대한 연관 탐색어 100위까지 제시하여 분석할 수 있도록 제공한다. 콜라겐은 피부, 혈관, 뼈, 치아, 근육 등 모든 결합조직의 주된 단백질로서 세포 간 매트릭스로서 존재한다. 포유동물의 경우, 전체 단백질의 1/3이 콜라겐이라고 한다. 콜라겐은 피부나 연골의 구성성분이기 때문에 먹거나 바르면 피부 탄력 및 뼈에 좋다는 이야기가 널리 퍼져 있다.

표 1. 콜라겐 탐색 연관어 빈도(n: 19,296)

No	연관어	분류	탐색건수
1	피부	속성	3,279
2	화장품	상품	2,313
3	효과	속성	1,890
4	에뛰드	브랜드	1,760
5	아리따움	브랜드	1,690
6	이니스프리	브랜드	1,679
7	로드샵	장소	1,679
8	파데	상품	1,671
9	색조	속성	1,668
10	비용	속성	1,667

본 연구에서는 검색어에 대한 탐색어들을 선택에서 인물, 단체, 장소, 상품, 속성 등으로 분석하였으며 표 1은 2016년 9월 2일에서 10월 2일까지 1개월 동안의 탐색어에 대한 상위 10위까지 제시하였다. 따라서 본 연구가 특정한 기간과 시점의 연구인만큼 다른 기간과 시점에 대한 적용은 불

가능함을 고한다. 먼저 콜라겐 탐색어에 대한 연관어 빈도수를 살펴보면 2016년 6월 시점에 17,787건이었으나 4개월이 지난 2016년 10월 빈도수는 19,296건으로 조금 증가하였다는 것을 알 수 있다. 특이한 것은 1개도 아닌 3개의 브랜드명이 진입한 것을 알 수 있다. 이러한 변화는 관련 산업 군에서 시장을 선점하기 위해 콜라겐 성분이 함유한 상품을 시장에 많이 출시하고 있다는 의미로 해석된다. 뿐만 아니라 콜라겐 성분이 함유된 건강기능식품을 이용하는 주 고객층인 여성 소비자들이 콜라겐 섭취 및 이용에 대한 인식변화가 이루어지고 있음으로 풀이된다.

다음으로 탐색어에 대한 여론은 콜라겐이 함유된 제품을 어떻게 생각하는지 여론/인식을 알 수 있다. 긍정적 여론 키워드 6개, 중립 1개, 부정적 3개인 것으로 나타났다. 부정적 여론도 일부 인지하고 있지만 긍정적 여론이 다수를 차지하고 있다. 콜라겐 연관어에서 살펴본바와 같이 주요 고객층인 여성들이 제품에 대한 관심은 많으나 부정적인 의견도 여전히 공존하고 있음을 확인시켜 주었다. 따라서 새로운 제품이나 서비스가 출시되는 경우에 잠재적인 구매자들이 제품 및 서비스가 어떠한 속성에 영향을 받아 그 제품이나 서비스를 선택하게 되는지에 대한 연구는 연구자에게 있어 매우 필수적이고 흥미로운 주제이다. 왜냐하면, 기업이나 서비스 제공 업체는 제품이나 서비스를 제공할 때 중요하게 생각되는 속성을 찾아 부각시키거나 또는 우려되는 점을 파악해 제거함으로써 제품 및 서비스의 소비자 수용과 확산 속도를 높일 수 있고 현재 고객들의 인식에 대한 분석은 제품 및 서비스의 발전에 따른 미래 수요 예측에 도움이 되기 때문이다.

표 2. 콜라겐 탐색 여론 빈도(n: 5,091)

No	연관어	분류	탐색건수
1	적발	부정	990
2	위반	부정	990
3	도움	긍정	629
4	효과적	긍정	557
5	좋은	긍정	406
6	탱탱하다	긍정	401
7	좋다	긍정	340
8	도움주다	긍정	265
9	고민	부정	261
10	다양함	중립	252

아직도 초기 단계에 머물고 있는 빅데이터에 대한 연구는 다양한 방법으로 진행될 것으로 판단된다.

*지면 관계상 참고문헌 생략함을 고합니다.