소셜 빅데이터를 활용한 불량식품에 대한 국내 소비자 인식 분석
–성분, 품질, 가격을 속인 식품유형을 중심으로–

Korean consumers' perceptions of food fraud using social big data
-Focusing on the food type that deceive ingredients, quality, and prices-

장현후[1], 장수경[2], 조건희[2], 김주란[1], 조천호[3], 권기성[3], 전향숙[2*]
Hyunhoo Jang[1], Su Kyung Jang[2], GunHee Cho[2], Jooran Kim[1],
Cheon Ho Jo[3], Kisung Kwon[3], Hyang Sook Chun[2*]
[1](주)버즈메트릭스, [2]중앙대학교 식품공학부, [3]식품의약품안전평가원
[1]Buzz Metrix,
[2]School of Food Science and Technology, Chung-Ang University
[3]National Institute of Food and Drug Safety Evaluation

I. 서 론

오늘날 식품 공급망이 세계화되고 복잡해짐에 따라 식품 생산자, 제조사, 가공업체, 유통업체, 판매자들이 경제적 이익을 취하기 위해 의도적으로 원료를 바꾸거나 부정물질 등을 식품에 첨가하여 제조한 불량식품[1)]이 증가하고 있다(전 등, 2014). 미국 가공식품제조협회(Grocery Manufacturers Association)의 추정에 따르면 불량식품으로 인해 매년 전 세계 식품업계에 100억–150억 달러의 비용이 발생하고 있고, 상업적으로 판매되는 전체식품의 약 10% 정도가 영향을 받고 있다고 한다(Johnson, 2014). 최근 한국 정부에서는 식품의 위변조 뿐만 아니라 식품의 생산 · 제조 · 유통 · 판매 등 전(全) 단계에서 발생할 수 있는 모든 법을 위반한 제품들을 모두 '불량식품'으로 정의하고 이를 근절하기 위한 여러 가지 대응을 하고 있다(정 등, 2016).

불량식품을 근절하기 위해서는 어떠한 유형의 불량식품이 어떤 범위로 확산되고 있는지를 명확하게 파악해야 한다. 그런 다음 식품 공급망 내에서 파악된 식품 원료나 중간 또는 최종 식품의 품질 및/또는 안전성을 보증할 수 있는 신뢰할 만한 판별법의 확보가 우선되어야 한다(Hong 등, 2017). 기존의 판별법 연구는 개별 연구자 관점에서 판별식품을 선정하고 연구되는 방식을 사용하고 있으나 실제 식품을 섭취하는 대상자는 소비자이기 때문에 소비자 관점에서 불량식품을 파악하고 이에 대한 판별법 개발 등 적절한 대응책을 모색할 필요가 있다.

식품과 관련된 소비자 인식 정도를 파악하기 위한 선행연구는 설문조사에 의존한 경우가 대부분이었다. 설문조사 방식은 소비자의 직접적 생각을 파악할 수 있다는 장점이 있지만 조사 항목에 대한 응답에 의식적인 판단이 개입

1) 식품 위변조(food adulteration), 경제적위변조(economically motivated adulteration), 식품 사기(food fraud) 등 유사한 용어가 사용되고 있으며 용어에 따라 다소 차이가 있으나 국내에서 사용하고 있는 불량식품으로 용어를 통일하여 사용

* Correspondence to: Hyang Sook Chun, Ph.D.
School of Food Science and Technology, Chung-Ang University
#4726 Seodong-daero, Ansung, Kyonggi 456-756, South Korea
Tel: 82-31-670-3290 (office) Fax: 82-31-675-3108 E-mail: hschun@cau.ac.kr

되며, 무의식적인 마음에 대해서는 파악하기 어려운 면이 있다. 또한 Focus Group Interview(FGI) 조사 방식 또한 응답이 사회자에 의해 유도될 수 있다는 한계점이 지적되고 있다(배 및 조, 2016). 뿐만 아니라 불량식품과 관련된 소비자 인식 정도를 조사한 연구는 매우 제한적이다.

수년 전부터 스마트폰, 스마트TV, RFID, 센서 등의 급속한 보급과 모바일 인터넷과 소셜미디어의 확산으로 데이터량이 기하급수적으로 증가하고 데이터의 생산, 유통, 소비 체계에 큰 변화를 주면서 데이터가 경제적 자산이 될 수 있는 빅데이터 시대를 맞이하게 되었다(이 등, 2013; 송 2016). 빅데이터란 기존 데이터베이스 관리도구의 능력을 넘어서는 대량(수십 테라바이트)의 정형 또는 비정형의 데이터를 말하는 것으로 이로부터 가치를 추출하고 결과를 분석하여 여러 분야에 활용하고 있다. McKinsey(2011)는 의료, 공공행정, 소매, 제조, 개인정보 부문에 빅데이터 기술을 적용할 경우 최고 7000억 달러(약 80조원)의 경제적 효과가 있을 것으로 전망한 바 있다. 특히 최근 소셜 네트워크 서비스(SNS)의 규모와 영향력이 점차 증대됨에 따라, 개인의 커뮤니케이션 수단으로서의 SNS는 긴급 상황과 위기대응에 결정적인 요소로 진화하고 있다. 이와 같이 사람들 간의 인적 네트워크를 기반으로 발생한 비정형 빅데이터를 소셜 빅데이터라 부른다(이 등, 2013). 트위터, 블로그, 카페, 게시판, 뉴스채널 등 온라인 매체에 남긴 정치, 경제, 사회, 문화에 대한 메시지가 그 시대의 감성과 정서를 파악할 수 있는 원천으로 등장함에 따라, 소셜 빅데이터 분석을 통해 정책수요나 사회위험에 대한 예측이 가능하다(송, 2016). 소셜 빅데이터의 분석은 방대한 양의 데이터를 활용하여 다양한 참여자의 생각과 의견을 확인할 수 있기 때문에 사회적 이슈에 대한 소비자들의 인식을 보다 유용하게 파악할 수 있다(배 및 조, 2016).

따라서 본 고에서는 불량식품에 대한 판별법 개발이나 안전관리 우선순위 파악 등 적절한 대응책을 모색하는데 도움이 되게 하고자 불량식품과 관련된 소셜 빅데이터를 분석하여 소비자 인식도를 조사하였다.

II. 소셜 빅데이터 분석 방법

데이터 수집 채널 및 기간

2015년 6월 발표된 Internettrend.co.kr 결과에 기반한 국내 온라인포탈 점유율을 고려하여 네이버(Naver, 83.9%)와 다음(Daum, 13.1%)을 1차 데이터 수집 포탈로 선정하였다(http://www.internettrend.co.kr/trendForward.tsp). 선정된 포탈의 채널 중 소비자들의 의견이 많이 발생되는 블로그와 카페에 게시된 웹문서(버즈)를 분석대상으로 하였다. 불량식품의 관심도와 트렌드를 분석하기 위한 데이터는 2014.1월부터 2016.12월까지 3년간 생성된 문서를 사용하였다.

데이터 수집 키워드

식약처의 17개 불량식품 유형을 토대로 유형별 수집 키워드를 Table 1과 같이 확정하였고, 크롤러(crawler)를 이용해 각 키워드가 포함된 웹문서의 원문을 수집하였다. 이때 수집의 정확성을 높이고자 불용어(불량식품과 연관성 없는 주제로 작성된 문서의 키워드)를 지정하여 진행하였다.

비정형 빅데이터 전처리

데이터 수집 채널을 대상으로 문서를 수집한 다음, 데이터의 정확도 제고와 분석시간 단축을 위해 1차적으로 시스템을 통한 기계적인 필터링을 통해 관련도가 떨어지는 노이즈 데이터를 삭제하였다. 그런 다음 시스템으로 확인이 어려운 부분은 인적 필터링으로 추가적으로 노이즈를 제거한 다음 이를 분석용 데이터로 사용하였다. 이때 수집된 텍스트 형태의 원데이터(raw data)는 비정형 데이터로 연구자가 분석하기에는 어려움이 있어, 자연어처리(natural language processing) 기반의 텍스트마이닝(text mining)을 통해 형태소 단위로 분절한 다음 키워드 기반의 의미분석 자료로 유형을 변환하여 분석에 사용하였다.

Table 1. 불량식품 유형별 수집키워드

No.	불량식품 유형	수집키워드
1	부패/변질된 위해 우려 식품	상한, 변질, 썩은, 곰팡이, 벌레, 구더기, 부패, 불량
2	유독/유해한 물질이 함유된 위해 우려식품	유해, 농약, 자연독, 독성, 항생제, 발암물질, 발암성, 방부제, 보존제, 색소, 동물용의약품
3	사용이 금지된 물질함유 식품	마약류, GMO, 유전자조작, 비식용, 불법첨가제, 멜라민, 색소, 사용금지, 미승인
4	불법도축, 병든 고기나 그 원료로 만든 식품	기생충, 불법도축, 밀도축, 병든, 밀도살, 불법도살, 불량고기
5	유독/유해물질이 함유된 식품용기/포장	포장, 용기, 가소제, 중금속용출, 증발잔류물
6	유해물질 기준, 규격 부적합 식품	기준치 초과, 부적합, 이물혼입, 위반
7	비위생적으로 제조/조리/재사용한 식품	재사용, 기준치, 위생불량, 비위생
8	수입 신고를 하지 않고 반입된 식품	밀수입, 밀반입, 수입신고
9	제품의 성분(함량), 품질, 가격을 속인 식품	가짜, 짝퉁, 중량미달, 불법증량, 함량미달, 거짓, 카라멜, 불량품질, 불량성분
10	성분, 영양가, 신고사항 등 허위표시식품	허위표시, 미표시
11	병원성 미생물 등에 오염된 위해우려 식품	식중독, 대장균, 장염, 바이러스, 살모넬라
12	성분, 규격에 맞지 않는 식품	규격미달, 기준미달, 부적합 성분
13	무허가/무신고 식품	무허가, 무신고
14	원산지를 속인 식품	원산지 허위, 원산지 속인, 원산지 미표시, 원산지 미표기
15	유통기한 위/변조식품	유통기한
16	질병치료나 의약품으로 오인/혼동 광고식품	허위과장, 과대광고, 소비자혼동
17	어린이현혹저가/저품질 정서저해식품	어린이, 정서저해, 미끼

Table 2. 최종 분석 빅데이터

수집채널	수집량(건)	최종 데이터(건)	최종 데이터 비중(%)	적합률(%)
네이버 블로그	586,882건	18,115건	17.9%	3.1%
네이버 카페	293,708건	65,783건	65.1%	22.4%
다음 블로그	232,574건	4,613건	4.6%	2.0%
다음 카페	702,271건	12,598건	12.5%	1.8%
전체 버즈량	1,815,435건	101,109건	100.0%	5.6%

빅데이터 분석방법

본 연구에서 사용한 빅데이터 자료의 분석방법은 크게 세가지를 활용하였다. 첫째, 수집 문서의 건수를 기반으로 문서 발생량의 트렌드와 주요 이슈를 확인할 수 있는 문서(버즈)량 추이 분석으로 불량식품에 대한 관심도와 주요 이슈를 파악하였다. 둘째, 의미망 분석(Semantic network analysis)으로 텍스트 마이닝(text mining)으로 분절된 형태소 단위의 키워드를 분류, 군집화, 요약을 하여 숨겨진 의미 있는 정보를 발견하는 분석 방법을 사용하였다. 본 연구에서는 오피니언 마이닝(opinion mining)을 통해 특정 주제에 대한 긍정/부정 키워드를 기반으로 평판이나 성향을 파악하는 방법인 감성분석(sentiment analysis)은 배제하였다. 이는 불량식품 자체가 부정적인 성향을 내포하고 있기 때문이다.

III. 소셜 빅데이터 분석결과

불량식품 관련 버즈량 추이

본 연구에서는 총 180만 건의 수집된 소셜 빅데이터에 대해 기계적, 인적 필터링에 의해 압축된 약 10만 건(5.6%)의 불량식품 관련 데이터를 분석하였다(Table 2). 그 결과, Fig. 1과 같이 2014-2016년 동안 매년 약 2-3만 건의 불량식품 관련 버즈가 발생하여 전년대비 매년 12~15%씩 꾸준히 증가하는 추이를 보였다. 계절별로 살펴볼 때 여름철은 식중독과 관련한 버즈량이 증가하였고, 겨울철은 바이러스성 장염으로 인한 버즈량이 증가되는 것으로 나타났다. 불량식품 관련 버즈량이 급증했던

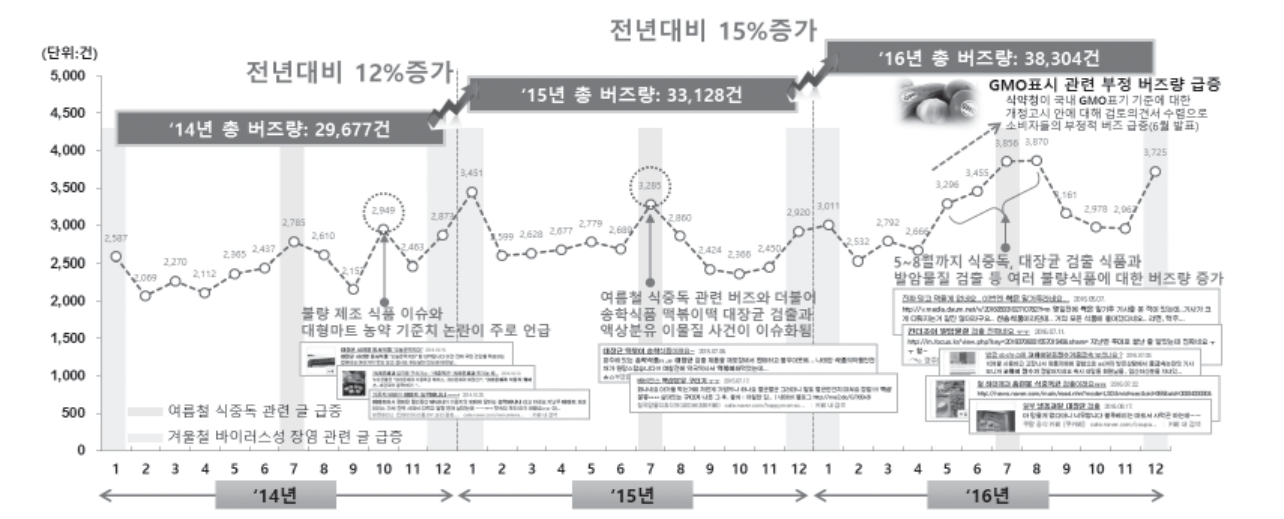

Fig. 1. 전체 불량식품 버즘량 트렌드.

시점은 2014년 시리얼 제품 중 대장균 검출, 수입 바나나 중 농약의 기준치 초과, 2015년 떡볶이 떡 중 대장균 검출, 액상분유 중 이물질 검출, 2016년의 경우 GMO 표시 기준 개정, 초콜릿류 발암물질, 하절기 식품 중 대장균 검출 등과 일치되는 것으로 나타났다. 이와 같이 불량식품 관련 버즘량이 급증하는 시점은 특정 식품사고 발생으로 관련 보도의 증가와 함께 트위터, 카페 게시판에서도 관련 내용의 소통량이 증가한 것으로 생각된다.

불량식품 유형별 버즘 발생량

수집된 불량식품 데이터에 대해 국내에서 분류하고 있는 17개 불량식품 유형으로 구분하여 분석을 실시하였다. 그 결과, 17개 불량식품 유형 중 가장 많이 언급되는 유형은 '병원성 미생물 등에 오염된 위해우려 식품'(42.4%)이었으며, 그 다음으로 '부패/변질된 위해 우려 식품'(23.0%), '유독/유해물질이 함유된 위해우려 식품'(22.5%) 순으로 나타났다(Fig. 2).

전 세계적으로 불량식품은 식품 위변조(food adulteration), 경제적 위변조(economically motivated adulteration, EMA), 식품 사기(food fraud) 등 다양한 용어로 사용되고 있으며, 아직까지 공식적인 명확한 정의는 없다. 2008년 미국 FDA에서는 잠정적으로 '경제적 이익을 위해 식품(제품)의 성분을 의도적으로 대체, 추가 또는 희석하는 행위 또는 그 식품(제품)'으로 정의하였으며, 현재 CODEX에서도 정의를 위한 작업반이 구성되어 진행되고 있다. 이에 본 연구에서는 국내 17개 불량식품 유형 중 미국 FDA의 불량식품 정의에 가장 근접한 유형인 '제품의 성분(함량), 품질, 가격을 속인 식품'에 대해 상

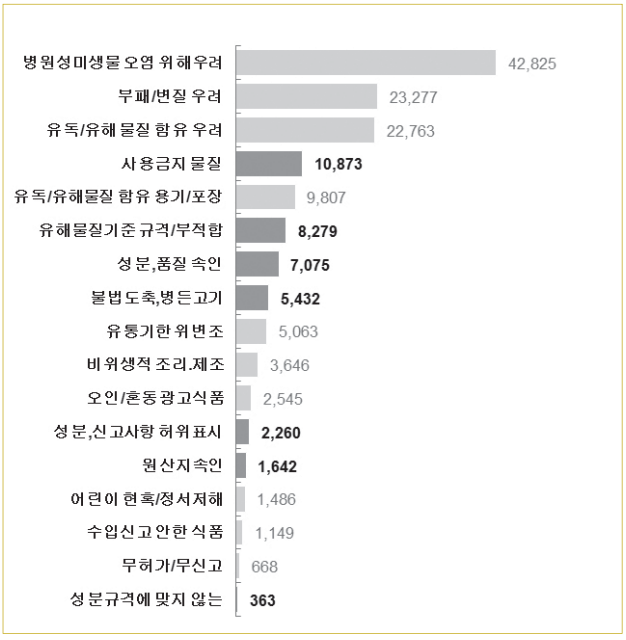

Fig. 2. 불량식품 17개 유형별 버즘 발생량.

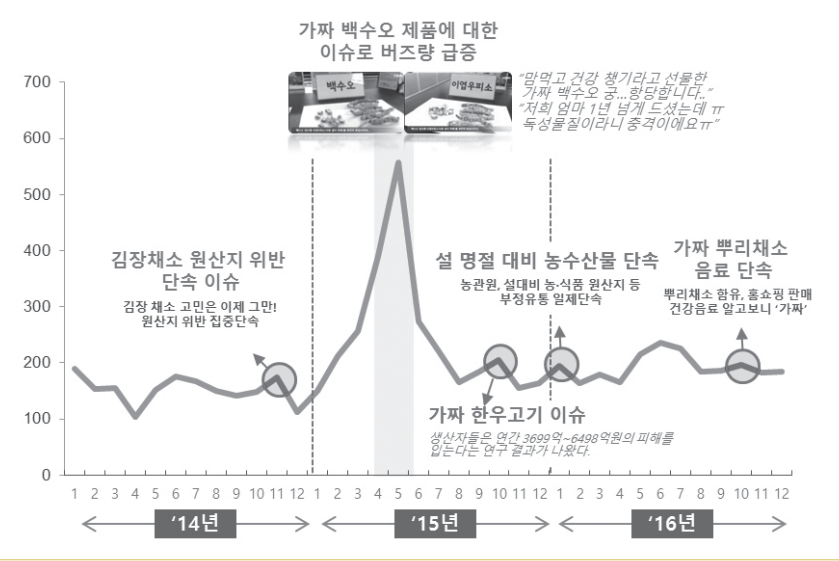

Fig. 3. 제품의 성분, 품질, 가격을 속인 불량식품 유형 버즈량 추이.

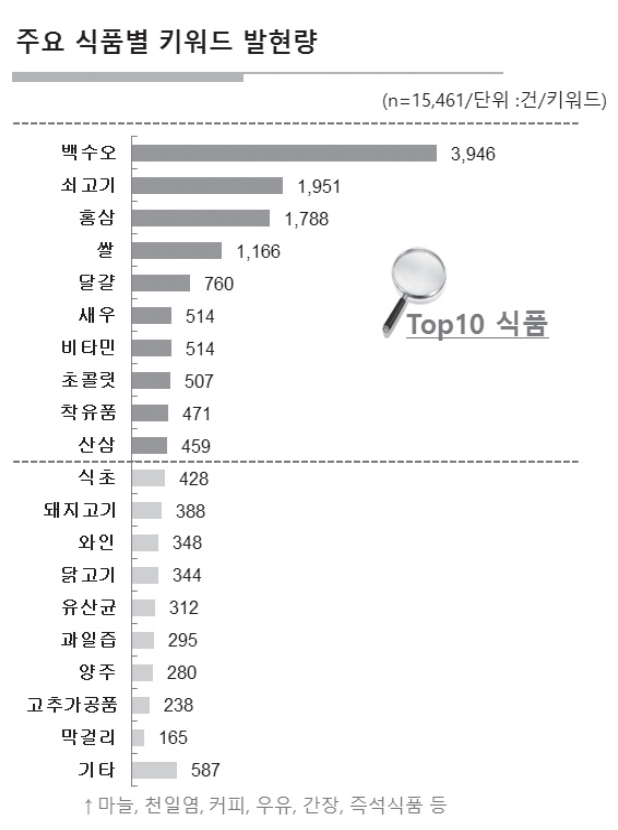

Fig. 4. 제품의 성분, 품질, 가격을 속인 불량식품 키워드 발생 순위.

세 분석을 실시하였다. '제품의 성분(함량), 품질, 가격을 속인 식품' 유형의 버즈 발생량은 7,075건으로 전체 중 7.0%를 차지하였다. 이 유형 중 가장 버즈량이 높았던 시점은 2015년 5월로 가짜 백수오 사건 발생과 관련된 것으로 보였다. 이 외 버즈량은 월별 100~200건으로 꾸준하게 나타났다(Fig. 3).

제품의 성분, 품질, 가격을 속인 식품 유형 중 발현 키워드

제품의 성분, 품질, 가격을 속인 불량식품 유형에서 나타난 고빈도 불량식품을 키워드분석으로 살펴본 결과는 Fig. 4와 같다. 해당 유형 가운데 '백수오'의 키워드 발현량이 가장 높았으며, 다음이 '쇠고기', '홍삼/인삼', '쌀' 순으로 나타났다.

2014-2016년 동안 '성분, 품질, 가격을 속인 불량식품 유형'에서 나타난 상위 19개 식품군을 대상으로 매년 키워드량 증감율과 버즈 비중을 고려하여 4사분면으로 구분한 2x2 Matrix 분석을 실시하였다. 이때 각 4 사분면 영역은 Table 3과 같이 정의하였으며, 이를 시각화하여 나타낸 것은 Fig. 5와 같다. 제품의 성분(함량), 품질, 가격을 속인 불량식품 유형 가운데, 2014년에는 많이 언급되었으나 이후 점차 감소하는 '안정화 영역' 식품군은 참기름, 들기름 등의 유지류이였고, 2014년 대비 2015년에 언급량이 증가되다가 2016년에 감소되는 '2015년 이슈 식품군'은 백수오, 달걀, 새우, 식초, 막걸리, 과일즙, 쇠고기였다. 2014 및 2015년 대비 2016년에 언급량이 증가된 '2016년 이슈 식품군'은 산삼, 돈육, 유산균, 고추가공품, 양주, 와인이었고, 2014년부터 2016년까지 꾸준하게 언급되는 '지속 증가 식품군'은 쌀, 홍삼/인삼, 닭, 초콜릿, 비타민이었다.

Table 3. 2014–2016년 '성분, 품질, 가격을 속인 불량식품' 키워드의 Matrix분석 결과

영역	영역 정의	도출식품
지속 증가 식품군	2014년부터 2016년까지 꾸준하게 불량식품으로 언급되는 식품	쌀, 홍삼/인삼, 비타민, 닭, 초콜릿
2016년 이슈 식품군	2016년에 언급량이 증가된 식품	고추가공품, 와인, 양주, 산삼, 돼지, 유산균
2015년 이슈 식품군	2015년에 언급량이 증가되었으나 2016년에는 감소된 식품	백수오, 달걀, 새우, 식초, 막걸리, 과일즙, 쇠고기
안정화 식품군	2016년까지 꾸준하게 언급량이 감소하는 식품	유지류(참기름/들기름)

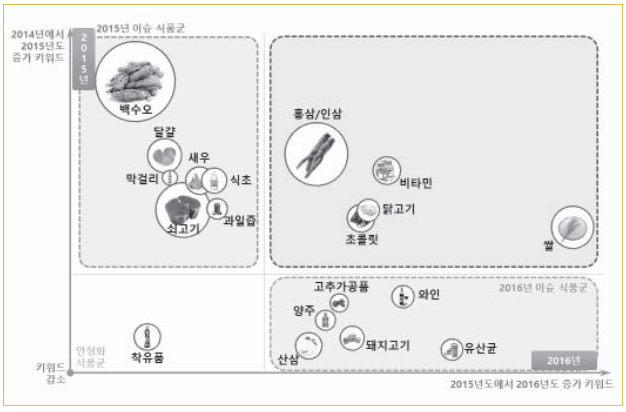

Fig. 5. 제품의 성분, 품질, 가격을 속인 유형의 2x2 Matrix 분석 결과. (※원의 크기는 키워드의 양을 나타냄)

식품별 상세 분석 결과

제품의 성분(함량), 품질, 가격을 속인 불량식품 유형에서 나타난 상위 19개 불량식품을 대상으로 2x2 Matrix 분석을 진행한 결과 도출된 '지속 증가 식품' 및 '2016년 이슈 식품군'을 추가적으로 분석하였다. 그 결과를 워드클라우드 기법으로 시각화한 것은 Fig. 6 및 7과 같다.

제품의 성분(함량), 품질, 가격을 속인 불량식품 유형 가운데 도출된 '지속 증가 식품군'은 쌀, 홍삼/인삼, 닭, 초콜릿, 비타민이었다. 먼저 쌀은 햅쌀의 진위여부, 쌀등급 표시제의 실효성(미검사쌀 다수 유통), 음식점에서 사용되는 쌀 품질(중국산을 국산으로 판매)에 대해 우려하고 있는 것으로 나타났다. 홍삼/인삼은 사포닌 함량이나 연근(age of root) 등을 속인 품질에 대한 우려와 유명 브랜드와 유사하게 만들어진 짝퉁 홍삼 제품에 대한 의견이 언급되었다. 닭은 일반 소비자가 생닭과 냉동 후 해동시킨 닭을 구별하기 어렵기 때문에 냉해동 닭을 생닭으로 속여 판매하지 않을까 하는 의구심과, 폐닭이나 산란닭의 유통, 닭가공 냉동식품으로 분쇄육에 다른 저가, 저품질 육류사용에 대한 우려가 나타났다. 초콜릿은 동남아산 제품이 고급 유럽 제품으로 둔갑판매 될 수 있다는 우려와, 국내 초콜릿의 성분이 카카오 함량이 낮은 저급원료로 만들어지는 것에 대한 부정적 의견이 언급되었다. 비타민은 합성비타민에 대한 우

Fig. 6. 지속 증가 식품군의 워드클라우드 분석 결과.

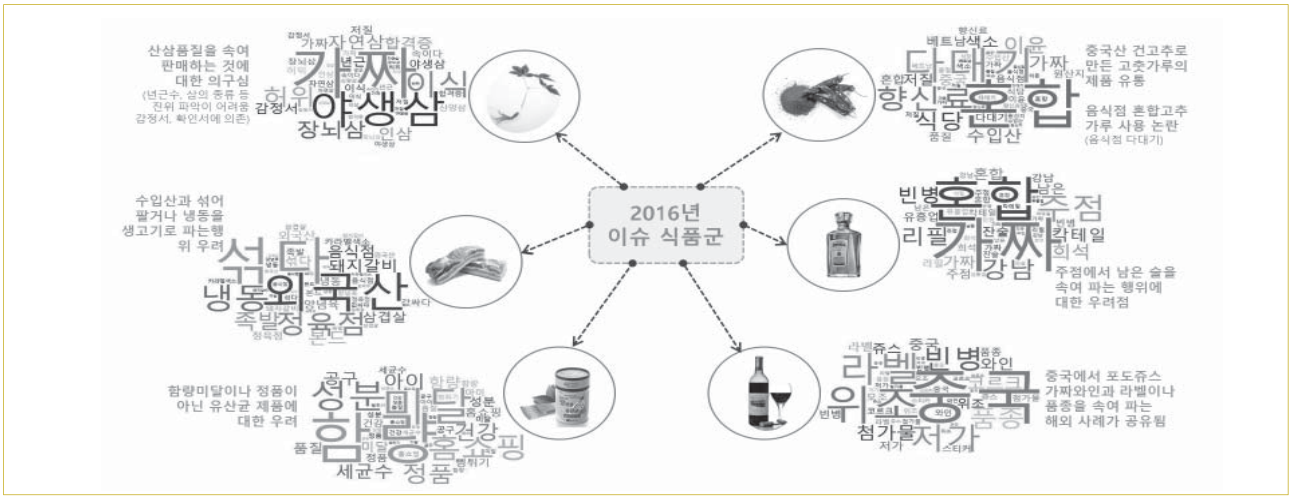

Fig. 7. 2016년 이슈 식품군 분석 결과.

려로 천연과 합성비타민을 구별하지 못하는 점과 비타민 추출원료로 GMO작물을 사용한다는 우려가 나타났다.

제품의 성분(함량), 품질, 가격을 속인 불량식품 유형 가운데 도출된 '2016년 이슈 식품군'은 산삼, 돈육, 유산균, 고추가공품, 양주, 와인이었다. 이를 세부적으로 살펴보면, 산삼의 경우 장뇌삼을 야생삼으로 속여 팔거나 연근을 속이는 것들에 대한 우려가 많았다. 돈육은 수입산과 국내산을 섞어 팔거나 냉동을 생고기로 파는 품질적인 측면의 우려가 주로 언급되었다. 유산균은 균수 함량미달이나 정품이 아닌 유산균 제품들에 대한 우려가 있었다. 고추가공품은 중국산 건고추를 섞어 만든 고춧가루에 대한 우려가 주로 언급되었으며, 음식점에서 사용하는 혼합고추가루(다대기)의 경우 저품질 원료가 사용되거나 다른 저가 원료로 대체될 가능성 등이 언급되었다. 양주의 경우 먹다 남은 잔주를 섞어 파는 리필양주에 대한 의견이 가장 많았다. 와인의 경우 국내보다 중국의 사례가 주로 언급되었으며, 포도주스를 와인으로 속여 팔거나 라벨에 품종을 속이는 행위가 언급되었다.

IV. 요약 및 결론

2014-2016년 동안 수집된 총 180만 건의 소셜 빅데이터에 대해 기계적, 인적 필터링을 실시한 다음 선별된 약 10만 건(5.6%)의 불량식품 관련 데이터를 분석하였다. 그 결과 매년 약 2-3만 건의 불량식품 관련 버즈가 발생하여 전년대비 매년 12~15%씩 꾸준히 증가하는 경향을 보였다. 불량식품 관련 버즈를 17개 유형별로 구분하여 분석한 결과 '성분, 품질, 가격을 속인 유형'은 전체 중 7.0%를 차지하였다. '성분, 품질, 가격을 속인 유형'에 대해 키워드분석을 실시한 결과, 가장 많이 언급된 식품은 2015년 이슈가 된 백수오였으며, 쇠고기, 홍삼, 쌀, 달걀, 새우, 비타민, 초코렛, 다류, 산삼의 순으로 나타났다. 매년 키워드량 증감율과 버즈 비중을 고려해서 4사분면으로 구분하여 식품을 나타낸 결과, 2014년부터 지속적 이슈 식품으로 '쌀', '홍삼/인삼류', '비타민', '닭고기', '초콜릿'이 2016년 이슈 식품으로 '산삼', '양주', '돼지고기', '고추가공품', '와인', '유산균'이 도출되었다. '성분, 품질, 가격을 속인 유형' 키워드분석 분석결과를 살펴볼 때, 도출된 식품의 대부분이 선행 또는 현재 연구되고 있는 불량식품이었다. 그러나 일부 새로이 나타나는 품목도 있었으며, 과거에 연구된 동일 품목이라도 새로운 유형의 불량 이슈가 도출되는 것으로 보인다. 따라서 불량식품 관련 소셜 빅데이터를 분석할 경우 설문조사처럼 국한된 양식의 의견이 아니라 온라인상의 다양하게 언급된 소비자의 의견을 취

합하므로 다각적인 VOC[2] 탐색이 가능하며 연구자 중심이 아닌 소비자 중심의 불량식품 연구에 활용 가능할 것으로 생각된다. 뿐만 아니라 기존 연구방식과 불량식품 관련 빅데이터 분석을 병행 활용할 경우, 정부의 불량식품 안전관리 정책에 대한 예측 및 평가에 대한 신뢰성이 더욱 높아질 것으로 기대된다.

참고 문헌

1. 전향숙, 하상도, 이헌주, 조승룡, 김건희, 강건욱. 국 · 내외 불량식품의 발생 및 대응 현황. 식품과학과 산업, 2014, 6, 3-12.
2. Johnson R. (2014) Food Fraud and "Economically Motivated Adulteration" of Food and Food Ingredients. CRS report 7-5700.
3. 정해랑, 하상도, 이헌주, 권기성, 조천호, 전향숙. 전 세계 주요 불량식품과 사회경제적 영향 평가 사례. Safe Food, 2016, 10(2), 3-12.
4. Eunyoung Hong, Sang Yoo Lee, Jae Yun Jeong, JungMin Park, Byung Hee Kim, Kisung Kwon, Hyang Sook Chun. Modern analytical methods for the detection of food fraud and adulteration by food category. J Sci Food Agric, 2017, 97(12), 3877-3896.
5. 배지영, 조미숙. 소셜빅데이터 분석을 위한 식품안전 온톨로지 구축. J Health Info Stat 2016, 41(3), 278-289.
6. 이충희, 허 정, 오효정, 김현진, 류법모, 김현기. 소셜 빅데이터 이슈 탐지 및 예측분석 기술 동향. 전자통신동향분석, 2013, 28(1), 62-71.
7. 송태민, 소셜 빅데이터를 활용한 식품안전에 관한 감성분석. 보건 · 복지 Issue & Focus, 2016, 4, 1-4.
8. McKinsey Global Institute, Big Data: The Next Frontier for Innovation, for Innovation, Competition, and Productivity. McKinsey Inc. 2011.

2) Voice of customer