소셜 빅데이터를 활용한 원산지 위변조 불량식품에 대한 국내 소비자 인식 분석

Korean consumers' perceptions of food that was intentionally falsified the country of origin using social big data

장현후[1], 우소영[2], 허성은[2], 정소영[2], 김주란[1], 권기성[3], 전향숙[2*]
Hyunhoo Jang[1], So Young Woo[2], Seong Eun Heo[2], So-young Jeong[2], Jooran Kim[1], 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

원산지(country of origin)란 특정 물품이 생산된 국가 또는 지역을 말한다. 원산지 표시는 국가브랜드가 높은 나라의 수출경쟁력 강화, 수출입 상거래질서 확립 및 생산자와 소비자의 권익보호 등의 원산지표시의 기능과 역할을 하는 바, 이 표시규정을 위반하게 될 경우 그 영향은 크고 중대하게 나타난다(황 및 이, 2012; Johnson 2014). 식품에 있어서도 식품의 안전성과 품질에 대한 소비자의 관심이 증가하면서 원산지는 소비자들이 식품을 선택할 때 매우 중요한 요소로 작용한다(이, 2010; 이 등, 2011).

현재 우리나라에서는 "농수산물의 원산지 표시에 관한 법률"에 따라 국내에서 유통되는 농축수산물, 가공품 및 영업소에서는 반드시 원산지를 표시하도록 의무화 하고 있다. 원산지표시제는 수입개방화 추세에 따라 값싼 외국산 농산물이 국산으로 둔갑 판매되는 등 부정 유통사례가 늘어감에 따라, 소비자의 알권리를 보장하고 공정한 거래 질서를 확립하기 위한 방안으로 도입되었다. 국내 농수산물의 원산지표시제는 '93년 6월부터 도입되었고, 음식점에서 원산지표시제는 2008년부터 시행되었다(최 등, 2009).

이와 같이 원산지표시 제도는 정부의 지속적인 지도단속으로 정착단계에 있으나 그럼에도 불구하고 매년 4천건 이상 부정유통이 적발되고 있다. 이에 이를 방지하기 위해 2017년부터 재범자 형량 하한제, 위반자 의무교육 및 과징금 부과와 같은 위반자에 대한 처벌 강화제도를 도입하는 등 다양한 안전관리 정책을 시행하고 있다(해양수산부, 농림축산식품부, 2017). 뿐만 아니라 의도적인 원산지 위반 행위를 근절하기 위해서는 어떠한 식품이 어떤 범위로 원산지 위반이 일어나고 있는지를 명확하게 파악하고 식품 공급망 내에서 신뢰할 만한 원산지 판별법 등 적절한 근절정책이 이루어져야 한다.

효율적인 식품의 원산지 위반 근절 정책은 정책입안자들이 전적으로 결정해야 할 사안이 아니라, 국민 즉 소비

* Corresponding author: 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

자 개개인의 의식과 인식의 변화가 반영될 때에 불신과 불안을 유발하지 않는 실효성을 나타낼 수 있다. 그러나 현대사회는 사회 모든 분야에서 소통(communication)이 강조되고 있지만 식품안전 분야는 타 분야와 달리 전문가는 과학적 사실에 기반하여 식품안전을 판단하는 반면에 소비자는 경험적, 감성적인 측면에 의존하여 평가하는 간극(gap)이 나타나고 있다(정 등, 2013). 식품의 원산지 판별 개발이나 관련 안전관리 정책 수립 및 집행에 있어서도 식품의 소비자 개개인의 인식이 중요한 영향을 미치기 때문에 우선 소비자의 인식 파악이 무엇보다도 중요하다고 할 수 있겠다.

식품의 원산지에 대한 소비자 인식 연구는 원산지표시제도 전반에 대한 것(한국소비자원, 2006; 최 등, 2009; 이 등, 2011)과 육류(이 및 진, 2009; 전, 2009), 와인(이 및 이, 2011), 커피(황 등, 2011; 황 및 이, 2012; 박 등, 2015) 등 특정 식품에 대해 이루어졌다. 한국소비자원(2006)에서 실시한 조사결과를 보면 농산물가공품의 원산지표시제도에 대해 알고 있는 경우는 47.0%로 그리 높지 않은 것으로 나타났고 원산지표시의 확인정도는 5점 만점 중 3.65점 이었으며 원산지표시제도에 대한 신뢰도는 3.56점으로 그리 높지 않았다. 이후 이제홍(2010)의 연구에서는 소비자의 외국산 제품 구매시 원산지 확인여부를 살펴본 결과 대체적으로 원산지를 확인하는 것으로 나타났다. 이 등(2011)의 연구에서 실시한 조사결과에 따르면 소비자들은 농축산물 구입시 원산지와 신선도를 가장 중요시하며 높은 관심을 갖는 반면 가공식품 구입시에는 원산지가 크게 고려되지 않았으며 낮은 관심을 나타내었다. 또한 원산지표시 이용에서도 농축산물에 비해 가공식품의 경우 이용도가 높지 않았다. 그러나 소셜 빅데이터를 이용한 원산지 변위조 식품에 대한 소비자 인식도 조사는 이루어진 바 없다.

최근 트위터, 블로그, 카페, 게시판, 뉴스채널 등 소셜 네트워크 서비스(SNS)의 규모와 영향력이 점차 증대됨에 따라, 개인의 소통 수단으로서의 SNS는 긴급 상황과 위기 대응에 결정적인 요소로 진화하고 있다. 사람들 간의 인적 네트워크를 기반으로 발생한 비정형 빅데이터인 소셜 빅데이터를 분석함으로써 다양한 정책수요나 사회위험에 대한 예측에 활용되고 있다(송, 2016). 특히 소셜 빅데이터를 소비자 인식 조사에 활용할 경우 다양한 참여자의 생각과 의견을 확인할 수 있기 때문에 기존 설문조사나 Focus Group Interview(FGI) 조사 방식이 가지는 단점을 보완할 수 있다(배 및 조, 2016).

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

II. 소셜 빅데이터 분석 방법

데이터 수집 채널 및 키워드

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. 식약처의 17개 불량식품 유형 가운데 '원산지를 속인 식품유형'의 수집 키워드는 전문가 FGI를 통해 '원산지 허위', '원산지 속인', '원산지 미표시', '원산지 미표기'로 확정하였다. 그런 다음 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, 데이

터의 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 분절하여 키워드 기반의 의미분석 자료로 유형을 변환하여 분석에 사용하였다.

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 의미가 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

원산지를 속인 식품의 유형별 버즈 발생 추이

'원산지를 속인 식품' 유형의 버즈 발생량은 1,642건으로 전체 중 1.6%를 차지하였다. 이 중 가장 버즈량이 높았던 시점은 매년 명절(설/추석)과 김장철(11월)에 사용하는 식품 구입시기와 관련된 것으로 보였다(Fig. 1).

Table 1. 불량식품 17개 유형별 버즈 발생량

불량식품 유형	최종 데이터(건)	최종 데이터 비중(%)
병원성 미생물 등에 오염된 위해우려식품	42,825	42.4
부패/변질된 위해 우려 식품	23,277	23.0
유독/유해한 물질이 함유된 위해우려식품	22,763	22.5
사용이 금지된 물질함유 식품	10,873	10.8
유독/유해물질이 함유된 식품용기/포장	9,807	9.7
유해물질 기준, 규격 부적합 식품	8,279	8.2
제품의 성분(함량), 품질, 가격을 속인 식품	7,075	7.0
불법도축, 병든 고기나 그 원료로 만든 식품	5,432	5.4
유통기한 위/변조식품	5,063	5.0
비위생적으로 제조/조리/재사용한 식품	3,646	3.6
질병치료나 의약품으로 오인/혼동 광고식품	2,545	2.5
성분, 영양가, 신고사항 등 허위표시 식품	2,260	2.2
원산지를 속인 식품	1,642	1.6
어린이현혹저가/저품질 정서저해 식품	1,486	1.5
수입 신고를 하지 않고 반입된 식품	1,149	1.1
무허가/무신고 식품	668	0.7
성분, 규격에 맞지 않는 식품	363	0.4

Fig. 1. 원산지를 속인 불량식품 유형 버즈량 추이.

원산지를 속인 식품의 품목별 언급량 및 관련 국가

'원산지를 속인 식품'에 대하여 축산물, 농산물, 가공식품, 수산물, 임산물 순으로 언급이 되었으며(Fig. 2), 관련 국가는 중국(44.6%), 미국(21.1%), 일본(20.0%), 베트남(3.2%) 등으로 나타났다(Fig. 3). 또한 음식점/식당(37.7%), 재래시장(29.6%), 대형마트(10.6%) 등의 순으로 장소가 언급되었다(Fig. 4).

원산지를 속인 식품 유형 중 발현 키워드

원산지를 속인 불량식품 유형에서 나타난 고빈도 불량식품을 키워드분석으로 살펴본 결과는 Fig. 5와 같다. 해당 유형 가운데 '쌀'의 키워드 발현량이 가장 높았으며, 다음이 '소고기', '돼지고기', '김치' 순으로 나타났다.

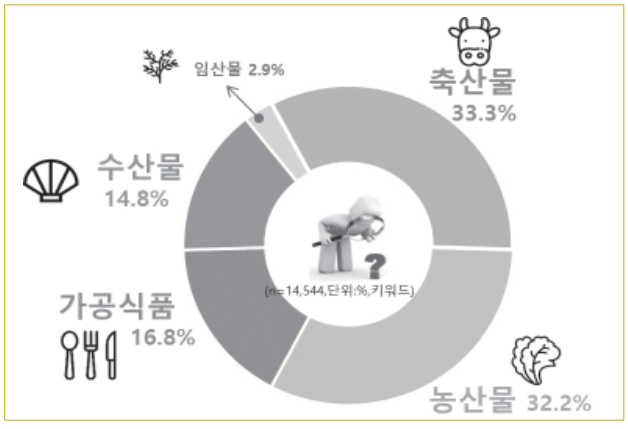

Fig. 2. 원산지를 속인 불량식품의 품목별 언급량.

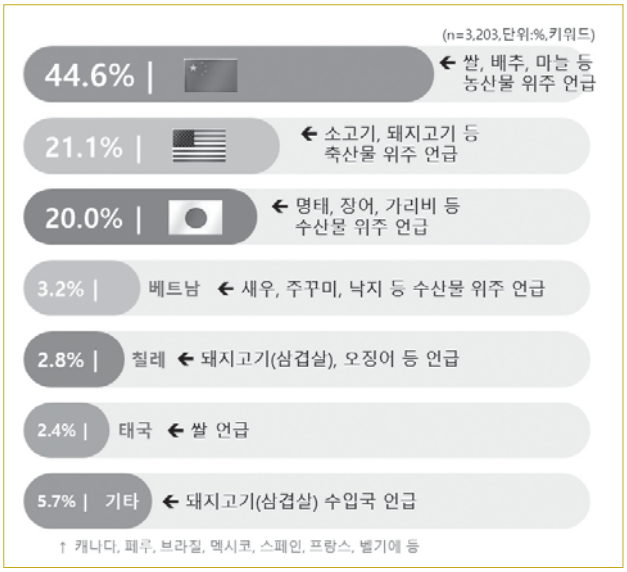

Fig. 3. 원산지 관련 언급 국가 및 식품.

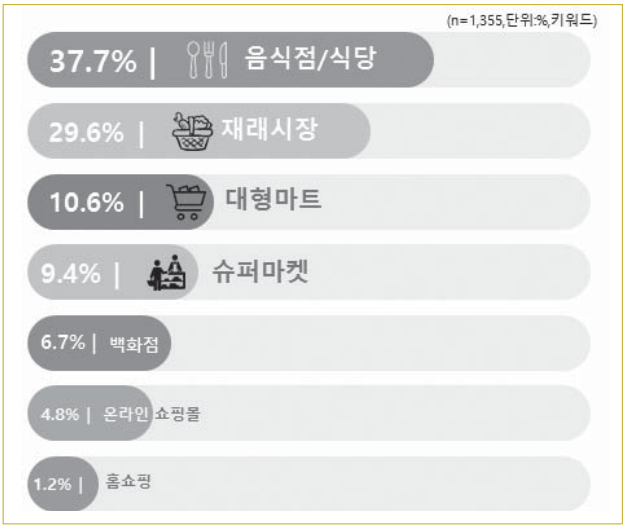

Fig. 4. 원산지 관련 언급 장소.

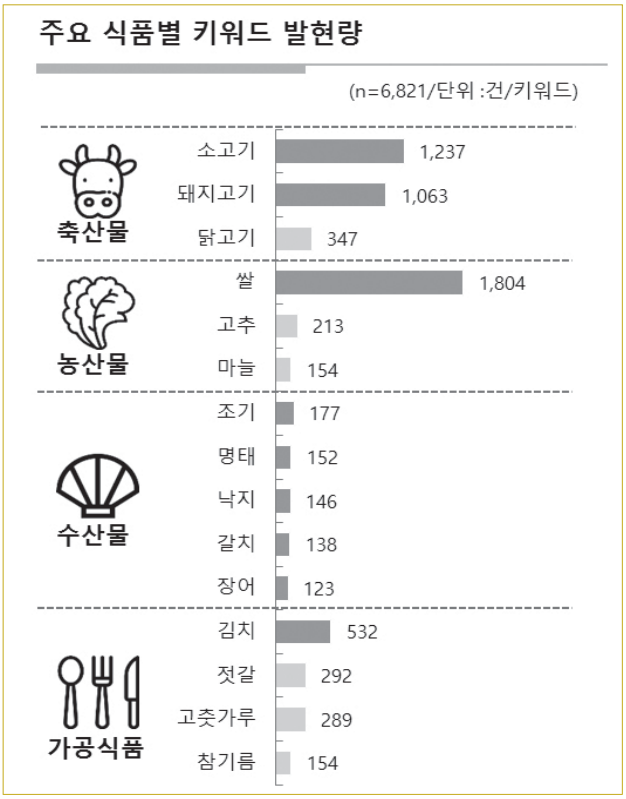

Fig. 5. 원산지를 속인 불량식품 키워드 발생 순위.

2014-2016년 동안 '원산지를 속인 불량식품 유형'에서 나타난 상위 15개 식품군을 대상으로 매년 키워드량 증감율과 버즈 비중을 고려하여 4사분면으로 구분한 2x2 Matrix 분석을 실시하였다. 이때 각 4사분면 영역은 Table 2과 같이 정의하였으며, 이를 시각화하여 나타낸 것은 Fig. 6과 같다. 원산지를 속인 불량식품 유형 가운데, 2014년에는 많이 언급되었으나 이후 점차 감소하는 '안정화 영역' 식품군은 참기름, 장어, 낙지, 고추였고, 2014년 대비 2015년에 언급량이 증가되다가 2016년에 감소되는 '2015년 이슈 식품군'은 소고기였다. 2014 및 2015년 대비 2016년에 언급량이 증가된 '2016년 이슈 식품군'은 돼지고기, 닭고기, 조기, 명태, 갈치, 조기, 고춧가루, 마늘이었고, 2014년부터 2016년까지 꾸준하게 언급되는 '지속 증가 식품군'은 쌀, 김치, 젓갈이었다.

Table 2. 2014-2016년 '원산지를 속인 불량식품' 키워드의 Matrix분석 결과

영역	영역 정의	도출식품
지속 증가 식품군	2014년부터 2016년까지 꾸준하게 불량식품으로 언급되는 식품	쌀, 김치, 젓갈
2016년 이슈 식품군	2016년에 언급량이 증가된 식품	돼지고기, 닭고기, 명태, 갈치, 조기, 고춧가루, 마늘
2015년 이슈 식품군	2015년에 언급량이 증가되었으나 2016년에는 감소된 식품	소고기
안정화 식품군	2016년까지 꾸준하게 언급량이 감소하는 식품	참기름, 장어, 낙지, 고추

Fig. 6. 원산지를 속인 유형의 2x2 Matrix분석 결과. (※원의 크기는 키워드의 양을 나타냄)

원산지를 속인 식품별 상세 분석 결과

원산지를 속인 불량식품 유형에서 나타난 상위 15개 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. 7 및 8과 같다.

원산지를 속인 불량식품 유형 가운데 도출된 '지속 증가 식품군'은 쌀, 김치, 젓갈이었다. 먼저 쌀은 원산지 미표시, 국산과 수입산의 혼합 판매에 대해 우려하고 있는 것으로 나타났다. 김치는 배추, 고춧가루, 소금 등 값싼 수입산 원료로 만들어 국내산 유명 브랜드로 포장해 판매하는 것이 아닐까 하는 의구심을 보였다. 젓갈은 원산지가 표시되지 않은 상태로 판매되는 재래시장의 새우젓, 액젓 등에 대한 우려가 높았다.

원산지를 속인 불량식품 유형 가운데 도출된 '2016년 이슈 식품군'은 축산물(돼지고기, 닭고기), 수산물(갈치, 명태, 조기), 고춧가루, 마늘이었다. 이를 세부적으로 살펴보면, 축산물의 경우 등급표시제를 시행하고 있는 한우보다 값싸게 식탁에 많이 오르는 돼지고기와 닭고기에 대해 정육점이나 음식점

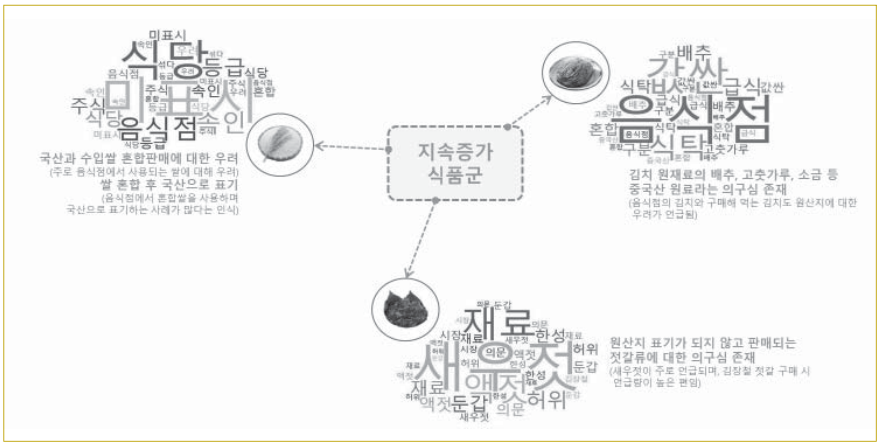

Fig. 7. 지속 증가 식품군의 워드클라우드 분석 결과.

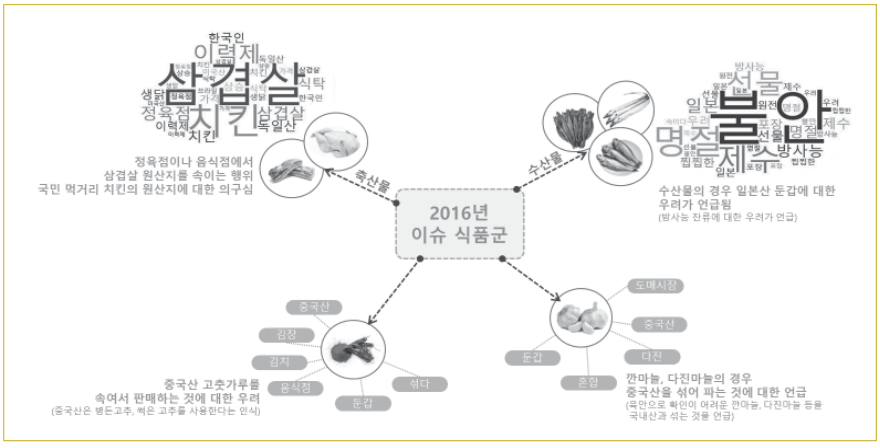

Fig. 8. 2016년 이슈 식품군 분석 결과.

10만 건(5.6%)의 불량식품 관련 데이터를 대상으로 '원산지를 속인 유형'에 대해 분석하였다. 그 결과, 전체 17개 불량식품 유형 가운데 '원산지를 속인 유형'은 1.6%를 차지하였다. '원산지를 속인 유형'에 대해 추가적으로 키워드분석을 실시한 결과, 가장 많이 언급된 식품은 쌀이었고, 소고기, 돼지고기, 김치의 순으로 나타났다. 매년 키워드량 증감율과 버즈 비중을 고려하여 4사분면으로 구분하여 식품을 나타낸 결과, 2014년부터 지속적 이슈 식품으로 '쌀', '김치', '젓갈'이, 2016년 이슈 식품으로 '돼지고기', '닭고기', '명태', '갈치', '조기', '고춧가루', '마늘'이 도출되었다. 이와 같은 결과를 살펴볼 때 그 동안 원산지 표시제도를 지속적으로 발전시켜 왔음에도 불구하고 일상생활에서 주로 섭취하고 있는 식품들에 대해서는 원산지 위반에 대한 우려를 나타내고 있었다. '원산지를 속인 유형' 관련 소셜 빅데이터를 분석할 경우 온라인상에서 다양하게 언급된 소비자들의 무의식적인 의견까지 취합할 수 있으므로 정부의 원산지 위반 식품의 안전관리 정책에 대한 예측 및 평가에 활용될 수 있을 것으로 생각된다.

에서 원산지를 속여 파는 것에 대한 우려가 많았다. 수산물은 특히 명절기간 원산지 단속에 대한 언급을 보였으며, 일본의 방사능 잔류와 관련된 많은 우려점이 현재도 존재하는 것으로 나타났다. 고춧가루와 마늘의 경우 중국산 재료에 대한 불신과 더불어 가루, 다져지는 형태로 국산과 섞어 판매하는 것을 우려하였으며, 특히 김장철에 많이 언급되었다.

IV. 요약 및 결론

2014-2016년 동안 수집된 총 180만 건의 소셜 빅데이터에 대해 기계적, 인적 필터링을 실시한 다음 선별된 약

참고 문헌

1. 황은애, 이승신. 원산지판정기준에 대한 소비자인식 연구-커피제품을 중심으로. 소비자문제연구, 2012, 2, 191-218.

2. Johnson R. (2014) Food Fraud and "Economically Motivated Adulteration" of Food and Food Ingredients. CRS report 7-5700.

3. 이제홍. 한국 소비자의 수입원산지 제품의 국가별 반응도에 관한 연구. 통상정보연구, 2010, 12(1), 353-373.

4. 이계임, 조소현, 전상곤, 김성훈, 송양훈. 농식품 원산지표시의 효과분석과 활용도 제고방안. 2011, 한국농촌경제연구원연구보고서.

5. 최지현, 이계임, 황윤재, 박영진. 가공식품 원료등 원산지표시 방법개선에 관한 연구. 2009, 한국농촌경제연구원.

6. 해양수산부, 농림축산식품부. 농수산물 원산시표시 위반 재범자 처벌 대폭 강화 보도자료. 2017. 5. 29.

7. 정기혜, 김동우, 조주연, 송다영, 김동영. 2013년도 식품안전체감도 조사, 한국보건사회연구원 정책보고서.

8. 한국소비자원. 농산물의 원산지표시제도 개선방안에 관한 연구. 2006, 정책연구보고서.

9. 이승미, 진양호. 일반음식점의 쇠고기 원산지표시에 대한 소비자인식 연구. 한국조리학회지 2009, 15(3), 112-127.

10. 이효진, 이상미. 와인의 원산지가 제품평가 및 구매의도에 미치는 영향. 한국콘텐츠학회, 2011, 11(1), 446-456.

11. 전영호. 육류 원산지의 국가이미지가 품질 지각 및 구매의도에 미치는 영향 연구. 관광연구, 2009, 23(3), 361-382.

12. 황은애, 정용수, 김인숙. 원두커피의 원산지 표시제도 개선방안-원산지 및 원재료 등의 표시를 중심으로. 2011. 한국소비자원 정책연구보고서

13. 박종명, 김종문, 유승민, 유영현, 김혜원, 이민석, 김용우, 윤아리. 식품접객업의 농수산물 원산지 표시관리 실태 조사. 2015, 농업생명과학연구, 49(5), 343-353.

14. 송태민, 소셜 빅데이터를 활용한 식품안전에 관한 감성분석. 보건 · 복지 Issue & Focus, 2016, 4, 1-4.

15. 배지영, 조미숙. 소셜빅데이터 분석을 위한 식품안전 온톨로지 구축. J Health Info Stat 2016, 41(3), 278-289.