식품 안전 분야에 적용되는 빅데이터의 이용 현황

Big Data for Food Safety

황미녕, 정한민*

Mi-Nyeong Hwang and Hanmin Jung*

한국과학기술정보연구원 융합기술연구본부

Convergence Technology Research Division,
Korea Institute of Science and Technology Information, Korea

1. 들어가며

2002년-2014년 사이 식중독 발생 현황을 분석한 결과 학교 및 기업체 등의 단체 급식에 의한 식중독 환자는 연간 3천-8천명이며 식중독 발병률이 전체의 35.5%~77.5%(평균 58.8%)로 가장 큰 비중을 차지하고 있다(1)(그림 1,2). 단체 급식 식중독은 일시에 많은 환자가 발생하고 질병의 특성상 전염의 위험성이 높은 특징이 있어서 식품 안전에서 중요한 이슈가 되고 있다.

국민들의 식품 안전에 대한 의식이 증가함에도 식중독 사고는 매년 감소하지 않고 오히려 집단화, 대량화되고 있는 실정이며 식품산업의 발달과 물류 증가로 인한 식품/잠재 위해인자가 전 지구적인 문제로 대두되고 있다. 유기농업체의 새싹 중 병원대장균 오염으로 인해 유럽 전 지역에 대규모 식중독이 발생한 바가 있고, 미국에서는 2011년 콜로라도주 농장에서 재배된 칸탈루프(cantaloupe) 멜론이 리스테리아균(Listeria monocytogenes)에 의해 오

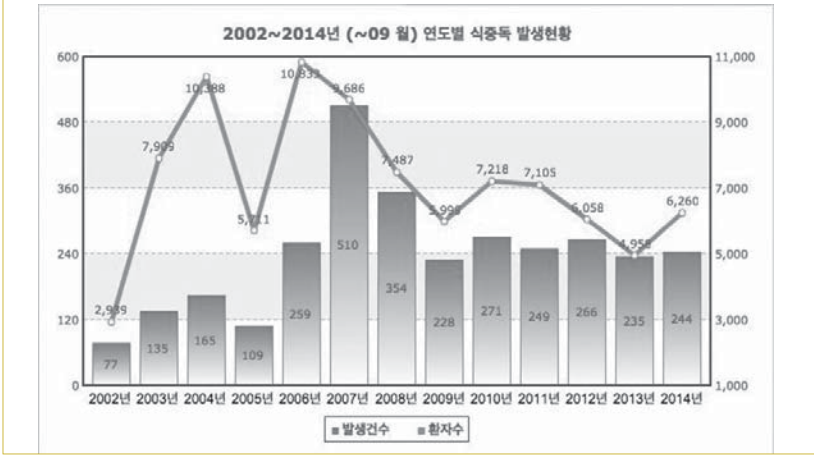

그림 1. 2002~2014년 연도별 식중독 발행현황, 식중독통계시스템, 식품의약품안전처

* Correspondence to: Hanmin Jung
Convergence Technology Research Division, Korea Institute of Science and Technology Information,
Daejeon, 305-806, Korea
Tel: +82-42-869-1772 E-mail: jhm@kisti.re.kr

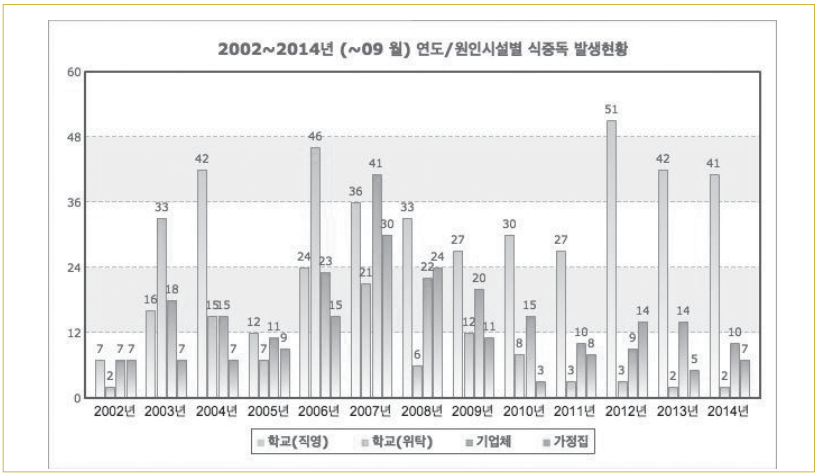

그림 2. 2002~2014년 연도/원인시설별 식중독 발생현황, 식중독통계시스템, 식품의약품안전처

염되어 이로 인해 28개 주에서 147명이 감염되고, 33명이 사망하는 사건이 발생하였다(2).

이런 비고의적 생물학적, 화학적, 물리적 식품오염물질의 발생은 국민건강 및 식품산업의 새로운 위협요인으로 대두되고 있으며 식품 안전사고의 동시화 추세로 빅데이터 등을 이용한 식품 안전 정보 공유서비스 개발이 시급한 실정이다.

본고에서는 식중독 위험을 탐지하고 예측하기 위한 문헌, 웹 뉴스, 정부 공개 데이터, 소셜 네트워크 서비스 데이터 등의 데이터 현황을 살펴보고, 이들 데이터 간의 연계 가능성에 대해 기술하고자 한다.

2. 국내외 빅데이터 이용 동향

빅데이터 수집 및 이용은 의료, 공공행정, 연구개발, 소매, 제조, 개인정보 부문에서 생산성 향상의 핵심요소로써 강조되어서 사용되고 있다. 미 대통령 과학기술자문위원회는 보고서에서 '모든 미국 연방정부 기관은 빅데이터의 전략이 필요하다'고 강조했고, Makinsy, Economist, Gartner 등은 원가 절감 및 제품 차별화, 기업 조직 내부 투명성 강화를 통한 경쟁력 향상의 초석으로 빅데이터 대한 가치 창출 사례 및 효과를 제시하였다(3,4). 모바일 인터넷과 소셜 미디어의 등장으로 다양한 분석 · 예측 · 활용을 가능케 하는 빅데이터에 대한 효율적인 저장, 관리, 연계 기술이 필요하며 이를 통해 융합 지식의 창출이 가능하다.

미국의 Google은 이미 2009년에 구글 검색사이트에 입력하는 검색어의 빈도수를 분석하여 CDC(Center for Disease Control and Prevention)가 발표하는 독감 보고서보다 2주 먼저 독감 바이러스의 확산을 예측하였다(5). 이는 실제 임상 데이터를 수집하여 통계를 낸 데이터가 나오기 전에, 전세계 구글 사용자의 검색어 입력과 위치 정보만을 이용하여 바이러스의 확산을 예측하였다. 또한, 미국 트위터 지진감지기 TED(Twitter Earthquake Detector)는 트위터 글 중에서 지진 낱말을 토대로 지진의 위치정보를 파악하여 지진발생을 실시간으로 전파하는 재난 대응하는 ICT(Information and Communications Technologies) 기술의 실제 사례이다(6). 국내에서도 국민건강보험공단에서 2008년부터 2012년까지의 5년간의 진료데이터를 분석하여 다빈도 상병에 대한 월평균 등락률과 ㈜다음소프트의 트위터 빈도수 분석 결과를 근거로 인플루엔자, 눈병, 식중독, 알레르기성 피부염 등의 4종의 질병을 선정하고, 건강보험 공단으로 집계되는 진료 데이터와 4종 질병 관련어를 포함하는 트위터 데이터의 빈도수를 집계하여 지역별, 연령별로 생활 속에서의 위험도를 알려주는 '국민건강 주의 알람서비스'를 2014년도에 개시하였다(7) (그림 3).

식품의약품안전처에서는 학교 식중독이 발생시 확산을 차단하는 목적을 가지고 식중독 조기경보 시스템을 2008년부터 운영해오고 있다(8). 기존에는 각 학교의 영양교사들이 해당 학교에서 식재료를 구매하는 업체들의 정보를 수기 입력하는 방식으로 시스템을 운영해오다가 2014년부터는 학교급식조달시스템과 나라장터의 유통업체 데이

그림 3. 국민건강 주의 알람 서비스화면

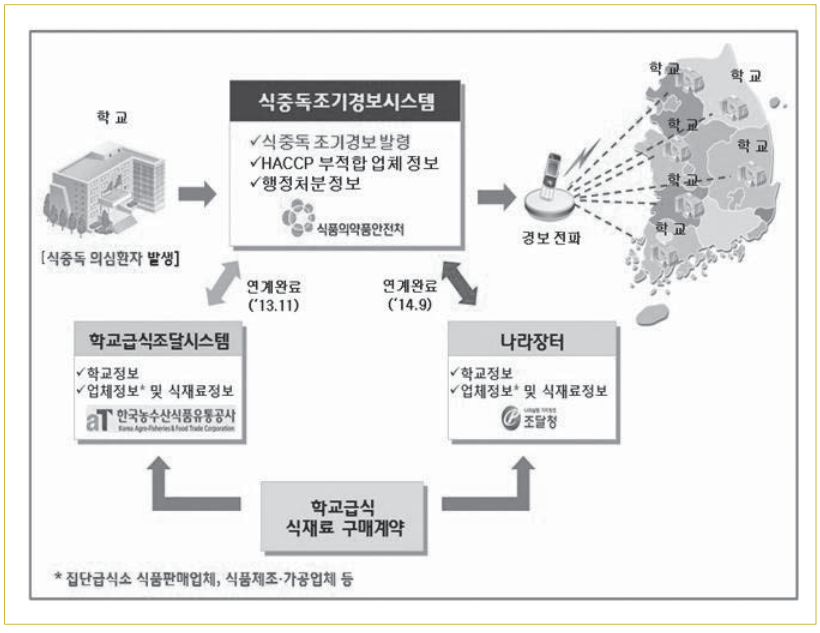

그림 4. 학교 식중독 조기경보 시스템

터 연계를 통해서 학교에서 식중독 의심환자 발생 시 해당학교와 거래하는 업체 및 이 업체와 연계되는 다른 학교를 확인하여 경보를 발령, 해당 학교의 영양사에게 SMS를 보내주는 체계를 가지고 있다(9,10). 이 시스템은 식약처의 식중독 보고관리시스템에서 취합된 데이터를 통해 식중독의 발생이 공식화되면 관계된 업체정보를 추적하는 과정으로 구체화되어있다.

이와 같이, 공공정보를 공개하고 공공데이터를 연계하고자하는 정부 3.0의 사례인 식중독 조기경보 시스템의 경우는 기존에 사람이 데이터를 입력하는 체계에서 각 기관의 데이터를 연계하는 체계로 전환함에 따라 유지비용을 절감하게 되었다. 본고에서는 더 다양하고 이질적인 데이터의 연계를 통해 식중독의 발생 전조를 컴퓨터 시스템이 감지하고, 식중독의 전파를 막기 위해 다종의 데이터들을 어떻게 이용할 수 있을지를 기술하고자 한다.

3. 식품 안전 분석과 연계 가능한 데이터 현황

빅데이터 분석 기술을 적용하여 식중독 질병의 확산 방향을 예측하기 위해서는 분석의 재료로 사용할 수 있는 식품 안전 분야와 관련된 실험 데이터 외에 학술문헌, 웹 뉴스, SNS, 기상 · 해수 데이터와 식중독 발생, 침수, 수해 등의 재난 · 재해 데이터 등의 다종의 자원들에 대한 이해가 필요하며 이를 융합할 수 있는 플랫폼을 구성해야 한다(그림 5).

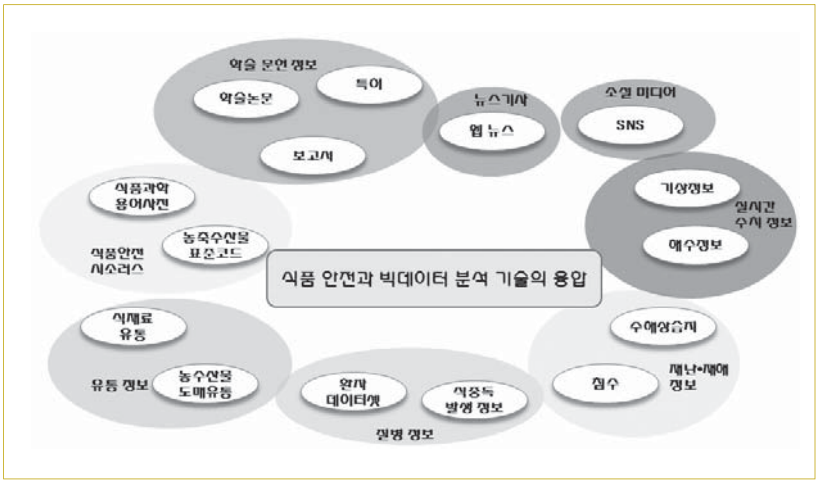

그림 5. 식품 안전 분석과 연계 가능한 데이터 현황

서에서는 식중독 질병 및 식품에 관한 학술적 지식을 추출해 낼 수 있다. 논문은 국내의 논문뿐만 아니라, 국제적으로 공신력이 높은 NCBI(National Center for Biotechnology Information)의 PubMed에 공개된 논문도 대상으로 삼아서 분석에 이용할 수 있다(11)(표 1). 하나의 논문 내에서의 질병에 대한 연관 정보를 추출해내고, 이들 정보들 간의 연관성을 확장하는 텍스트 마이닝 기법을 적용하면 복합된 지식 추출이 가능해진다(12). 논문, 특허, 보고서와 같은 학술 문헌에서 텍스트를 추출해내기 위해서는 발간처에 따른 상이한 구조를 고려한 텍스트 추출기가 필요하며, 각 문헌들의 지적

1) 학술 문헌 데이터

논문, 특허, 보고서 등의 전문적인 지식을 담고 있는 문

표 1. 해외 논문 데이터의 예제

일련번호	필드명	내용	예시
1	Art_ID	논문관리번호	2912873
2	PubDate	게재날짜	2014 Oct
3	Art_Name	논문제목	Stable silver nanoparticles synthesis by citrus sinensis (orange) and assessing activity against food poisoning microbes
4	Page	페이지 번호	815-818
5	Abstract	초록	Silver nanoparticles are considered as good antimicrobial agent. Ag-NPs were synthesized by mixing silver nitrate solution with citrus sinesis …
6	Author_Name	저자	Naila Arooj, Nadia Dar, and Zahoor Qadir Samra
7	Author_Affiliation	저자소속	Institute of Biochemistry and Biotechnology
8	DOI	DOI	10.3967/bes2014.118
9	ISSN	ISSN (p)	0895-3988
10	Name	저널 제목	Biomedical and Environmental Sciences

식중독 우려 도시락 고시원 정기 납품…도시락 납품 충격실태 - Chrome

www.112news.co.kr/news/news_preview.html?news_idx=201411102246072167

THE LAW & POLICE NEWS 법무경찰신문

인쇄하기 닫기

식중독 우려 도시락 고시원 정기 납품…도시락 납품 충격실태

전태영 기자 | 입력 2014-11-10 22:46:08 | 수정 2014-11-10 23:31:06

대전시특별사법경찰(이하 특사경)이 제조년월일을 거짓으로 표시한 무등록 도시락 제조 업체 등 식품위생법 위반업소 5곳을 적발했다.

10일 특사경은 관내 도시락 제조판매업체 중 위법행위가 의심되는 42개 업체를 대상으로 2개월간 수사를 진행했다. 그 결과 무등록식품 제조 1건, 유통기한 경과제품 보관 2건, 제조년월일 허위표시 1건, 식품표시기준 위반 1건 등 총 5건이 적발됐다.

한 업소의 경우 식중독 발생이 우려되는 도시락을 고시학원 등에 정기 납품하면서도 관할구청에 등록조차 하지 않은 것으로 드러났다. 또 다른 3개 업소는 삼각김밥의 제조년월일을 거짓 표기해 대학교 매점 등에 납품했다.

또한 지난 국군의 날 행사 때 도시락을 납품하면서 제품명, 포장재질, 포장단위, 성분 및 함량 등을 미표기한 업체도 적발됐다.

한편 특사경은 적발된 5개 업체 관계자를 검찰에 송치하고, 4개 업체는 관할구청에 행정처분을 의뢰할 방침이다.

[© 법무경찰신문(www.112news.co.kr), 무단 전재 및 재배포 금지]

그림 6. 식중독 관련 뉴스 예제(14)

재산권 문제 때문에 전체 문서를 수집하여 연구용으로 분석하기에 한계점이 있다.

2) 웹 뉴스 데이터

식품 안전에 관한 학술적인 정보는 논문, 특허, 보고서 등의 문헌을 통해서 분석 가능하지만, 뉴스 데이터에는 시회 현상 및 이슈들을 날짜, 시간, 지역 등의 정보가 구체적으로 포함되어 있어서 특정 분야의 국내 및 국제 현안을 시기적절하게 파악하기 위한 중요한 자원으로 이용할 수 있다. 이러한 뉴스들은 과거 데이터부터 현재 데이터까지 수집가능하며 입수된 HTML 문서에서 제목, 본문, 날짜, 키워드, 작성자 등의 유용한 정보를 추출하여 데이터베이스에 저장한 후, 분석에 사용한다(13)(그림 6). 웹 뉴스의 데이터는 인쇄된 기사에 비해 입력과 수정이 빈번하게 이루어지기 때문에, 웹 뉴스 수집기는 동일한 기사가 업데이트가 되는 것이 새로운 기사가 쓰인 것인지의 중복성을 체크해야하는 기능을 가져야 한다.

3) 소셜 네트워크 서비스

스마트폰의 보급으로 개개인도 실시간으로 문헌 및 사진과 같은 이미지 데이터를 실시간으로 생성하는 정보 제공처의 역할을 하게 되었다. 이 중에서 텍스트 정보를 주로 담고 있는 트위터, 페이스북 등의 소셜 네트워크 서비스(Social Network Service)에서도 식품 안전에 관한 정보들을 선별하여 수집할 수 있다.

– 트위터

트위터(Twitter)는 최대 140자의 제한된 글자 수로 언제 어디서나 정보를 실시간으로 교류하는 '빠른 소통'이 가장 큰 특징으로, 뉴스 보다 신속한 '정보 유통망'으로 주목을 받고 있다(15). 트위터 데이터의 전량을 입수하는 것은 제한되어 있지만 OpenAPI를 사용하여 특정 키워드가 포함된 트위터 데이터는 수집이 가능하다(16). 식중독과 관련된 주요한 키워드를 선정하여 이들 키워드가 포함된 트윗을 실시간으로 수집할 수 있다. '식중독' 용어가 포함된 실제 트윗을 수집한 예를 아래 표에서 확인할 수 있다(표 2). 각각의 트윗들은 하나의 고유한 아이디를 가지며, 트윗이 생성된 시간 정보와 사용자 아이디, 내용이 포함된다. 또한, 트윗을 날린 사용자가 위치 정보를 공개하는 경우에는 데이터가 생성된 위치 정보의 수집도 가능하다. 트위터 데이터는 기존의 논문이나 웹 뉴스와 같은 문헌을 분석하기 위한 텍스트 마이닝 기술 외에 추가적인 고려사항이 필요하다. 140자라는 제한된 글자수 때문에 이용자들이 줄임말을 사용하거나 띄어쓰기를 생략하는 경향을 보이고, 스마트폰을 사용하여 입력하기에 오타도 빈번히 나타나며, 감정을 표현하는 기호 문자를 사용하여 감성 표출하기에 기호 분석 기법도 요구된다.

표 2. 수집된 트위터의 실제 예

tweet_id	tweet_dt	user_name	tweet	location
522595488730607616	1410161250	wvxjnqgm348	청주의 한 고교 집단 식중독 증세..일부 감염 확인	–
522615811773460480	1410161411	tbczkim	뭐야 킹크랩 먹으러 가기로 했는데 웬 식중독 ㅠㅁㅠ	–
522668131378929666	1410161739	hyangsoolee	가을 식중독 주의, " 김밥은 식힌 후에 뚜껑 덮으세요 " http://t.co/W0kHi5alv8 출처 : 네이버 뉴스	Gwangju, Korea
529076247008010241	1411031003	nhis_kr	나들이 하기 좋은 계절인 가을, 선선한 날씨에도 식중독은 꾸준하게 발생한다는 사실! 가을에도 식중독 조심하세요~ http://t.co/fvrl4sSRhZ	서울시 마포구 독막길24

4) 수치 데이터

장염 비브리오균, 노로 바이러스와 같은 식중독 질병을 일으키는 원인균들은 기온, 해수온도 등의 외부 환경이 활성화의 주요한 요인이 되기에 이들 수치 데이터들을 수집하여 분석에 이용할 필요가 있다.

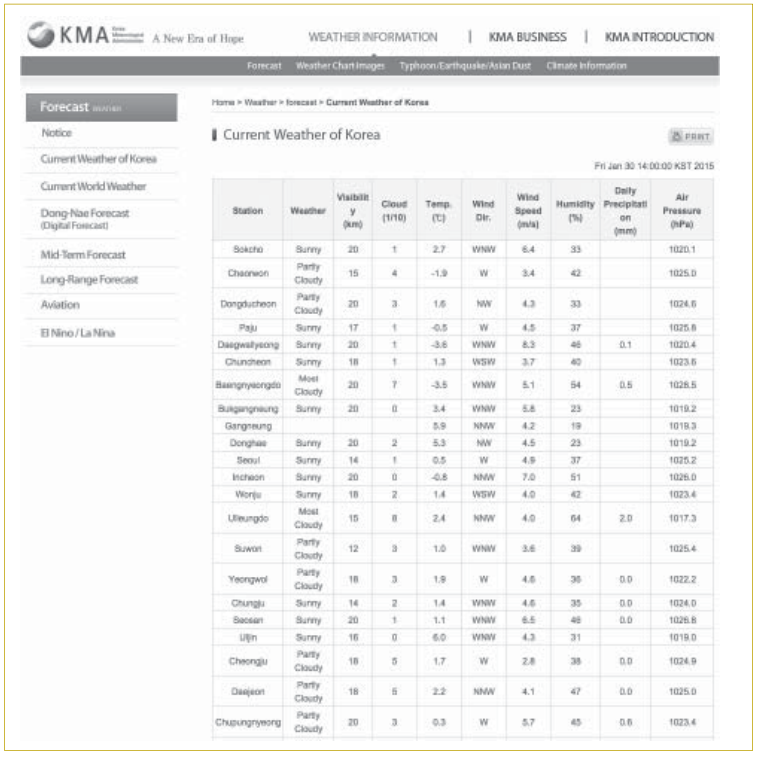

그림 7. 기상청에서 공개하는 기상 관측, 예측 데이터

– 기상청 기상 데이터

기상청에서는 기상 관측 및 예측 데이터를 주기적으로 공개하고 있다. 관측 데이터는 1시간 단위로 발표하고 예측 데이터는 3시간 간격으로 48시간을 예보하며, 10일까지의 중기 예보도 제공한다. 그 데이터는 매 시간마다, 관측지역의 현재 일기와 시정, 운량, 중하운량, 현재기온, 이슬점온도, 체감온도, 일강수, 적설, 습도, 풍향, 풍속, 해면기압 등의 세부 정보를 얻을 수 있다(17)(그림 7).

– 해양조사원 해수 데이터

국내의 연안 바다의 온도 정보는 국립해양조사원에서 해양 GIS 정보를 기반으로 조석, 조류, 해류 예측 정보와 실시간 관측정보, 위성 온도 등의 정보를 서비스하고 있다(18)(그림 8).

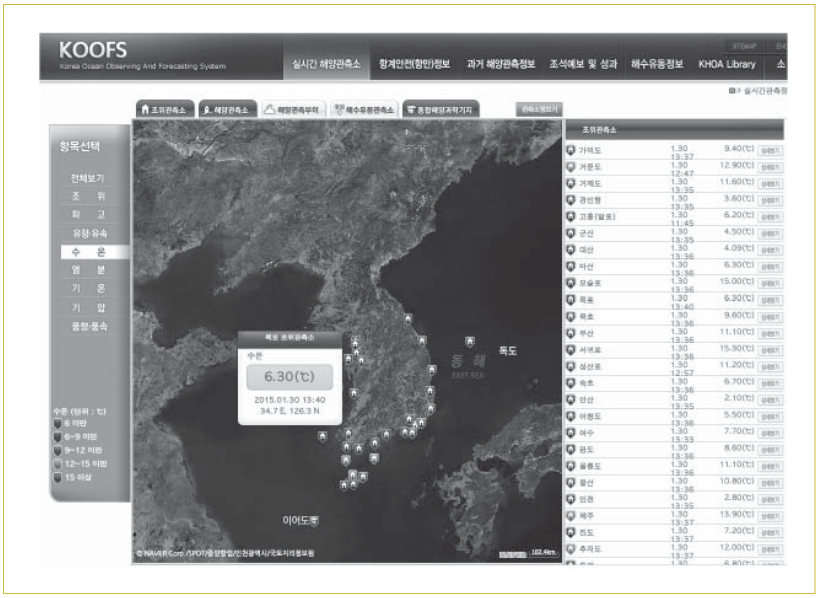

그림 8. 국가해양정보시스템에서 시간별 공개하고 있는 해수 정보 화면

이러한 지역별로 발표되는 수치 데이터를 수집하여 이용하기 위해서는 기관 별로 발표하는 데이터의 연계가 가능하도록 지역명을 표준화시켜서 관리해야한다. 실제로 기상청에서는 '대전시 유성구 어은동'에 대한 기상 관측 데이터를 법정동을 기준으로 '대전시 유성구 온천2동'으로 공개하기 때문에, 기상청지역코드와 법정동 코드 등의 매핑이 필요하다.

5) 정부기관 공개 데이터

– 식중독 발생 정보

식품의약품안전처에서는 국내에서 발생했던 집단 식중독 통계 정보를 연도별, 지역별, 월별, 원인시설별, 원인균별 등으로 구분해서 식중독통계시스템 사이트를 통해서 제공하고 있다(1). 기존에 발생했던 식중독 사고를 비교, 분석해 효율적인 사전 예방 대책을 수립하기 위하여 식중독 발생 이력 자료로 이용할 수 있다.

- 월 별 : 1월~12월까지 각 월별 식중독 발생 건수, 환자수 및 건당 환자수 제공
- 연도별: 2002년부터 기준시점 이전년도까지 1년 단위의 식중독 발생 건수, 환자수 및 건당 환자수를 제공
- 지역별 : 16개 시 • 도별 식중독 발생 건수, 환자수 및 건당 환자수를 제공
- 원인시설별 : 집단급식소(학교직영, 학교위탁, 기업체 등), 음식점, 가정집 등 시설에서 식중독 발생 건수, 환자수 및 건강 환자수를 제공
- 원인균별 : 세균(살모넬라, 황색포도상구균, 장염비브리오균, 바실러스세레우스, 클로스트리움퍼프리젠스, 클로스트리움보툴리눔, 캠필록박터제주니, 병원성대장균, 기타세균), 바이러스(노로바이러스, 기타 바이러스), 화학물질, 자연독 등에 의한 식중독 발생 건수, 환자수 및 건당 환자수를 제공

– 환자 데이터셋

건강보험심사평가원에서는 건강보험 청구자료를 기초로 하여 진료개시일 기준 1년간 진료 받은 환자를 대상으로 표본 추출하여 연구를 목적으로 하는 경우에 개인정보를 제외한 가공된 데이터를 유료로 제공하고 있다(19)(표 3). 데이터 필드 중에서 환자 정보와 치료받은 병원의 위

표 3. 심평원에서 제공하는 데이터셋 종류

데이터셋	내용
IRA-NIS (입원환자데이터셋)	2009년, 2010년, 2011년 데이터 제공 입원환자 추출비율 13% (약 70만 명) 외래환자 추출비율 1% (약 40만 명)
HIRA-NPS (전체환자데이터셋)	2010년, 2011년 데이터 제공 전체환자 추출비율 3% (약 140만 명)
HIRA-APS (고령환자데이터셋)	2010년, 2011년 데이터 제공 고령환자(65세이상) 추출비율 20%(약 100만명)
HIRA-PPS (소아청소년 환자데이터셋)	2010년, 2011년 데이터 제공 소아청소년환자(20세 미만) 추출비율 10%(약 110만명)

치 정보와 질병 정보를 연계하여 분석에 이용할 수 있다.

6) 식품과학 용어사전 및 시소러스

시소러스(thesaurus)란 용어의 사용법과 용어들 사이의 관계에 대한 정보를 제공하는 어휘 도구를 말한다. 용어의 관계성은 일반적으로 상위 개념, 하위 개념, 용례 혹은 동의어, 관계어, 대체어 등으로 분류되는데, 시소러스는 이러한 관계성을 이용, 탐색기 질의에 포함된 용어의 의미를 확대하기 위해 주로 사용된다. 수집된 문헌 등에서 식품 안전에 관련된 개체명을 인식하고, 이들 간의 유의미한 관계를 추출해내기 위해서는 개체명을 인식하기 위한 용어사전이 잘 구축되어 있어야 좋은 성능을 얻을 수 있다. 식품과학에 사용되는 기본적인 한 · 영 쌍 구축 및 연관어 정보는 식품안전정보서비스 식품나라의 식품과학용어사전을 기본으로 사용하고, 식품의 상 · 하위 정보는 농산물품질관리원에서 제정한 농축수산물 표준코드의 대중소분류를 적용하여 식품 분야의 시소러스를 구축하는 것이 가능하다(표 4).

4. 이질 데이터 연계 및 정규화

이와 같이 식품 안전에 관련된 빅데이터를 분석하기 위해서는 형식 및 구성, 표현이 다른 다종의 산재된 데이터 간의 연계 · 공유하는 프레임워크의 개발이 선재되어야 한다. 예를 들어, 국문 뉴스에서의 '홍합'이 영문 뉴스의 'mussel'과 동일한 의미를 가진다는 것은 용어사전을 잘 구축하는 것으로 해결이 가능하지만, 뉴스 속에서 '전라남도 목포시', '전남 목포', '목포'가 같다는 것은 정규화 과정을 거쳐만 컴퓨터가 인지할 수 있다. 또한, 기상청 수집 정보에서의 '46110'의 지역코드와 농수산물 산지 정보에서 '530-000'의 산지코드가 '목포'와 동일하다는 것은 지역명을 통제하는 통제표를 구축하고 지속적으로 관리를 해야만 알 수 있다. 이러한 이유로 기상청, 행정법정동코드, 농림축산식품부의 산지코드 등의 지역명을 통제하는 테이블을 구축하고 관리해야하며, 이는 지역명뿐만이 아니라, 개체명 타입 마다 필요하다.

표 4. 식품과학용어사전의 예제

일련번호	필드명	내용	예시1	예시2
1(PK)	ID	고유번호	00000001	00000020
4	Name_K	한글 용어명	홍합	새우
5	Name_E	영문 용어명	mussel	shrimp
6	Related_Word	연관어	쌍각 조개, 조가비	갑각류
7	Content	설명	쌍각 조개의 하나. 조가비는 쐐기 모양이고 살은 붉은색을 띤 등색이다. 살은 먹는다. 말려 먹기도 한다.	갑각류에 속하는 동물의 총칭. 머리, 배, 고리 세부분으로 이루어져 있고, 수염과 꼬리가 긴 것이 특징이다. 민물이나 바닷물에 산다. 크기에 따라 대하, 중하, 젓용의 작은 새우(젓 새우)로 나눈다.

5. 결론

지금까지 식품 안전에 연관된 데이터들의 현황에 대해 기술하였다. 이들 데이터들은 수치 데이터들처럼 형태가 고정된 정형 데이터들도 있고, 문헌 데이터와 같은 비정형 데이터로 일정한 규칙이 있지 않다. 또한 기상, 해수정보, 웹 뉴스, SNS 데이터들처럼 실시간으로 생성되는 데이터도 있고, 학술 논문처럼 주기적으로 발간되거나, 환자기록 데이터처럼 일괄적이면서 간헐적으로 수집 및 정제된 기록 데이터들이 쏟아져 나오기도 한다. 이러한 다양한 이질적인 데이터들을 수집하고, 전처리 과정을 거친 후 정제된 데이터를 텍스트 마이닝 분석 기술을 이용하여 텍스트 속에서 식중독 위험 정보, 식중독 발생 정보, 식품 안전 연관 정보 등을 추출해 내야하고, 이들 데이터와 수치 데이터와의 연계, SNS 데이터 속에서 실시간으로 발생하는 사건과의 연결점을 찾아낼 수 있다. 이를 통해 식중독 집단 발생 전조 및 위험도를 분석하여 단체급식 제공처, 지방자치단체 및 식품안전관리부처에 상황을 전파하여 신속 · 정확한 의사결정을 가능하게 해주어 식중독 사고로 인한 피해를 최소화할 수 있을 것이다.

감사의 글

본 연구는 한국식품연구원의 연구비 지원(E0142014297)에 의해 수행되었으며 이에 감사드립니다.

참고 문헌

1. 식품의약품안전처 식중독통계시스템, http://www.mfds.go.kr/e-stat/index.do

2. Multistate Outbreak of Listeriosis Linked to Whole Cantaloupes from Jensen Farms, Colorado.http://www.cdc.gov/listeria/outbreaks/cantaloupes-jensen-farms/index.html

3. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH. 2011. Big Data: The next frontier for innovation, competition, and productivity.

4. American Association for the Advancement of Science (AAAS). 2010. Designing a Digital Future: Federally Funded Research and Development Networking and Information Technology.

5. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. 2009. Detecting infulenza epidemics using search engine query data. Nature 457:1012-1014.

6. Earle PS, Bowden DC, Guy M. 2011. Twitter earthquake detection: earthquake monitoring in a social world. Annals of Geophysics 54(6):708-715.

7. 국민건강 주의 알람 서비스, http://forecast.nhis.or.kr/

8. 식중독 조기경보 시스템, http://www.foodnara.go.kr/ews/user/loginForm.do, http://www.mfds.go.kr/fm/content/view.do?contentKey=16&menuKey=128

9. 학교급식 전자조달시스템, http://school.eat.co.kr/atn/etnBidList.jsp

10. 나라장터 국가종합전자조달, http://www.g2b.go.kr

11. PubMed.gov, http://www.ncbi.nlm.nih.gov/pubmed

12. Ohta T, Pyysalo S, Miwa M, Tsujii J. 2011. Event extraction for DNA methylation, J. Biomedical Semantics 2:1-15.

13. Hwang MN, Hwang M, Yeom HN, Kim KY, Shin SM, Kim T, Jung H. 2014. Development of Framework System for Managing the Big Data from Scientific and Technological Text Archives. JIST:33-38.

14. 법무경찰신문 - 식중독 우려 도시락 고시원 정기 납품, 도시락 납품 충격 실태http://www.112news.co.kr/news/content.asp?fs=2&ss=31&news_idx=201411102246072167

15. 트위터, http://twitter.com

16. 트위터 개발자 페이지, http://dev.twitter.com

17. 기상청 동네예보, http://www.kma.go.kr/weather/forecast/timeseries.jsp

18. 국립해양조사원 국가해양정보시스템, http://sms.khoa.go.kr/koofs/kor/observation/obs_real_map.asp

19. 건강보험심사평가원, http://www.hira.or.kr