기후변화와 식품안전 간 융합 데이터베이스 및 데이터 마이닝 시스템 구축

The Development of Convergence Database and Data-Mining Systems between Climate Change and Food Safety

송보람, 이헌주*

Boram Song, Hunjoo Lee*

켐아이넷㈜

Cheminet Co., Ltd

서 론

IPCC(Intergovernmental Panel on Climate Change) 보고서를 비롯한 국내 · 외 연구들에 따르면, 기후변화(Climate change)는 인류가 가까운 장래 내지 향후 100년 이내에 직면하게 될 가장 복잡한 도전 중에 하나이며 특히 온도, 습도 등 환경에 민감한 생물학적 위해인자(세균 등) 및 생물유래 화학적 위해인자(자연독소 등)의 식품안전 분야에도 큰 영향을 줄 것으로 전망하고 있다. 다양한 식품 매트릭스와 위해인자 간 조합의 형식으로 각 국가의 고유한 기후변화에 따른 부정적 혹은 긍정적 영향에 대하여 예측하고 각각의 영향에 따른 적응 시스템 구축에 박차를 가하고 있다.

기후변화에 따른 식품안전 분야 영향 예측을 위해서는 기반이 되는 광범위한 데이터가 요구된다. 첫째, 기후변화와 관련된 과거 기상 자료 및 미래 기후변화 예측 시나리오가 요구되며 둘째, 식품안전 분야의 관측 자료들의 확보가 요구된다. 그러나 이 두 가지 자료는 매우 이질적인 자료이며 부처별로 산발적으로 생산된 자료이다 보니 체계적인 데이터베이스 모형으로 융합하는 것 자체에도 상당한 난점을 가지고 있다. 또한 대용량 자료(소위, Big-data)이다 보니 예측 모형을 개발하려는 연구자들 입장에서 두 자료원간 통합 · 가공 등 데이터 전처리는 물론 자료의 데이터 분석 · 처리에 상당한 어려움을 갖고 있는 것이 현실이다.

본고에서는 2010년부터 현재까지 식품의약품안전처의 『기후변화 대응 식품안전관리 사업단』 과제의 일부로 추진된 기후변화와 식품안전 간 융합 데이터베이스인 CC-MIMS(Monitoring Information Management System due to Climate-Change)와 데이터마이닝 시스템인 CCFS-DMC(Data-Mining Center for Food Safety against Climate-Change)의 전체적인 프레임워크 및 주요 기능(Features)등 구축 내용에 대하여 살펴보고자 한다.

* Correspondence to: HunJoo Lee
Cheminet Co., Ltd., 43, Mokdongjungang-ro, Yangcheon-gu, Seoul, Korea
Tel: 82-2-2647-4930 Fax: 82-2-2647-4932 E-mail: adstar2000@empas.com

본 론

1. 기후변화와 식품안전 간 융합 데이터베이스 프레임워크

가) 기상 · 기후 데이터 생산 현황

기후변화에 따른 식품안전 분야의 영향을 평가하기 위해서는 필수적으로 한반도의 시공간적인 관점에서 과거의 기상자료 및 미래의 기상 예측 시나리오가 요구된다.

기상자료는 국립기상청에서 1999년 이후 전국의 95개의 유인 종관기상관측장비(ASOS, Automated Surface Observing System)와 477개소의 무인 자동기상관측장비(AWS, Automatic Weather System) 등 총 572개 지점에서 일 단위로 관측을 실시하고 자료를 수집하고 있으며 미래 기후변화는 기후변화 대응정책의 수행 여부에 따른 온실가스 농도를 등급화한 온실가스 배출 시나리오(RCP, Representative Concentration Pathway) 2.6, 4.5, 6.0 및 8.5 시나리오가 제시되고 있으며 남한의 경우 약 1Km단위로 상세화하여 제시하고 있으며 미래 기후변화 관련 예측 연구에 활용되고 있다.

나) 식품안전 관측 자료의 생산현황

특히 기후변화와 연관될 것으로 생각되는 식품안전 관측 자료는 크게 식품에 영향을 주는 생물학적 및 화학적 위해인자는 그림1 과 같다. 해당 위해인자에 대한 분야별

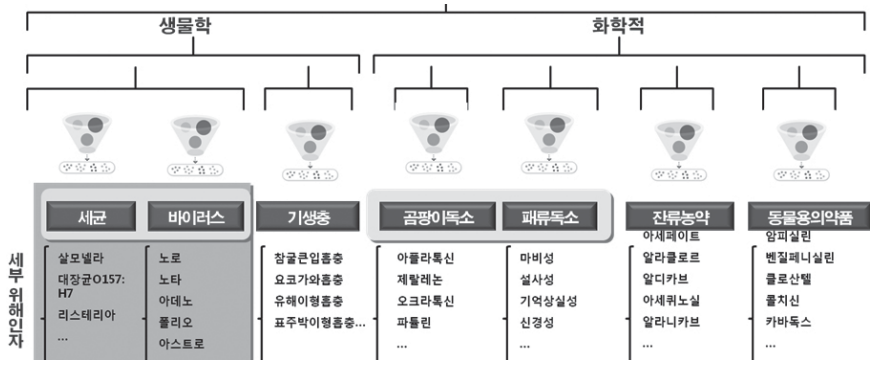


그림 1. 기후환경 위해인자 계층도

그림 2. CC-MIMS 초기 화면

모니터링 자료가 사업단 내지는 국가기관에서 탐색 내지는 조사 모니터링을 통하여 생산된 자료들이 데이터베이스화 되었고 수집된 자료의 출처는 주로 식품의약품안전처를 비롯한 농림축산식품부, 해양수산부 등 식품 유관 부처의 관측 자료이다.

다) 기후변화와 식품안전 간 융합 데이터베이스 CC-MIMS 소개

CC-MIMS는 기후변화에 핵심적인 자료인 기상 · 기후 데이터와 식품안전 관측 자료인 생물학적 · 화학적 모니터링 데이터를 상호 융합하여 기후변화에 따른 식품위해인자의 예측 및 재난 대응 체계 구축에 활용지원을 위한 상호 네트워크 구축 도구이다.

CC-MIMS는 웹을 기반으로 설계된 시스템으로 식품관측 자료에 대하여 생물학적 · 화학적 위해인자별 특성 및 표준화된 식품 코드체계에 따라 수집, 저장, 정제, 가공할 수 있는 기능을 제공하여 충실성 있는 식품 관측 자료를 생산할 수 있도록 설계되어 있을 뿐만 아니라 부처별로 이질적인 자료에 대한 표준화된 연계 인터페이스를 제공하고 있다. 또한 기상 · 기후 데이터의 네트워크 구축을 위한 생산 · 수거 시간 및 지역을 기상 관측 거점

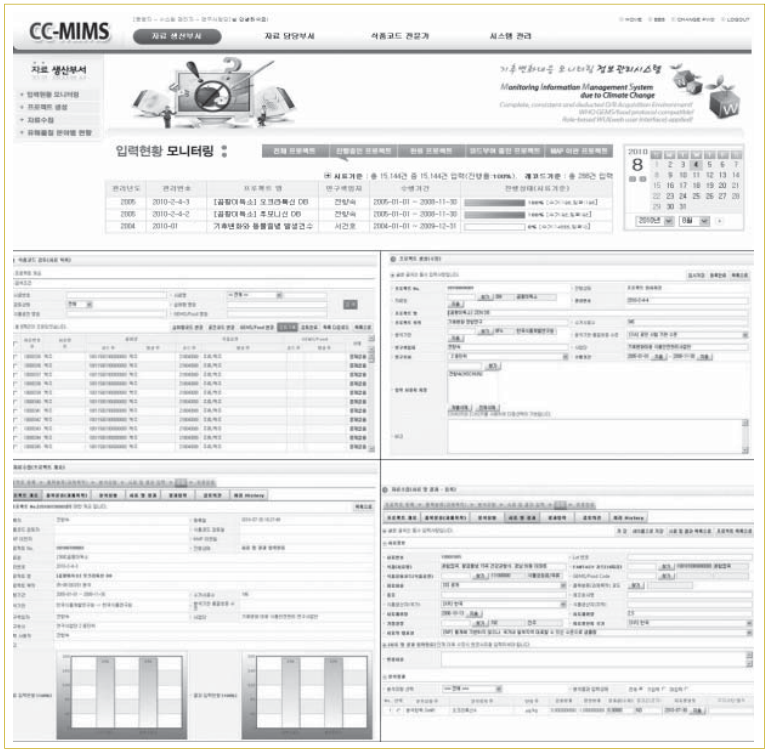


그림 3. CC-MIMS 주요화면

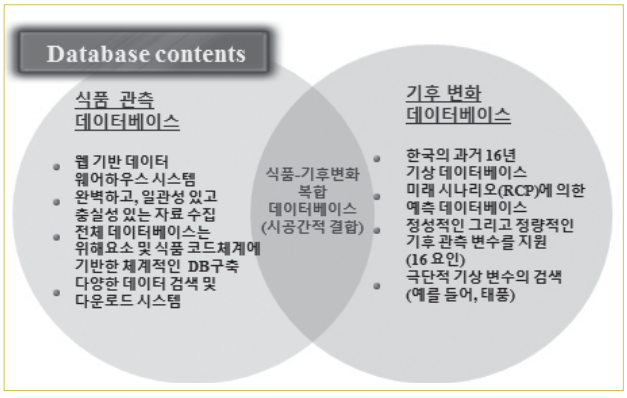


그림 4. CC-MIMS 기후변화-식품안전 융합 개념도

과 매칭함으로써 시공간적 결합을 할 수 있도록 설계되어 있다. 특히 기상 · 기후 데이터의 경우 그림4 와 같이 향후 기후변화에 따른 식품안전 관리의 용이성을 위하여 유사 기온 및 강수량 대역을 가지는 지역을 3개의 권역(제1,2,3)으로 클러스터링 하였다. 그 기준이 되는 기후 특성은 표1 과 같다.

2. 기후변화와 식품안전 데이터 마이닝 시스템(CCFS-DMC) 구축

가) 데이터 마이닝(data-mining) 개요

데이터마이닝은 데이터베이스 속에서의 지식탐색(KDD, Knowledge Discovery in Database)의 핵심과정으로 단어의 의미에 내포되어 있는 바와 같이 데이터부터 유용한 지식이나 통찰력(Insight)을 캐내는 컴퓨터과학 분야의 한 기법이다. 특징은 대량의 자료를 취급한다는 점이고, 인공지능, 통계적 기법 등을 활용하여 새롭고 유용한 패턴을 탐색하거나 미래의 결과를 예측하는 데에도 유용하게 활용되고 있다. 최근에는 해당 기술이 확대되어 빅데이터 분석 기술의 핵심 기술로도 사용되고 있다.

나) CCFS-DMC 개요

CCFS-DMC는 기후변화에 따른 식품위해인자의 변화 패턴 탐색 내지는 기후변화에 따른 위해인자의 영향 예측을 목적으로 만들어진 데이터 마이닝 시스템이다. 본 시스템은 CC-MIMS의 기후변화 식품안전 간 융합 데이터베이스를 그 원천으로 하여 구동되며 기후변화에 따른 식품안전 예측에 필요한 다양한 데이터베이스 및 분석 기능을 포괄적으로 가지고 있다.

다) CCFS-DMC 개발 및 구동환경

CCFS-DMC의 개발은 웹을 기반으로 한 Intranet 환경에서 구동되도록 시스템이 구성되어 있다. 따라서 사

표 1. 유사 권역 구분 기준

권역	기준 기후 특성
1	연평균 기온 11.7℃, 연 강수량 1286.6㎜ 인 지역
2	연평균 기온 11.3℃, 연 강수량 1263.4㎜ 인 지역
3	연평균 기온 13.0℃, 연 강수량 1346.9㎜ 인 지역

그림 5. 유사 기후특성에 따른 제1,2,3권역 클러스터링

용자 클라이언트의 경우 웹 브라우저를 통하여 허가된 사용자에 의해서 접근이 가능하며 관계형 데이터베이스 기반의 분산 서버에서 동작한다. 표2는 세부적인 개발환경 및 개발도구를 명세한 것이다.

그림 6. CCFS-DMC 메인화면

표 2. 개발환경 및 개발도구

항목	사양
개발 플랫폼	JAVA/JSP기반의 J2EE 플랫폼(JDK 1.5)
서버 운영체제	Windows Server 2008이상 혹은 UNIX계열 운영체제
웹 서버	Apache tomcat 1.6
DBMS	Oracle 10.2g
통계도구	R server 엔진
클라이언트 운영체제	IE 8이상(권장) 및 크롬(Chrome) 브라우저
보안체계	공공기관 보안 & 취약점 등 구현반영(정보통신산업진흥원 가이드라인 준수)

라) CCFS-DMC 주요 기능(Features)

표3은 CCFS-DMC의 전체 시스템 구성을 보인 것이다. CCFS-DMC는 총 9개의 메뉴로 구성이 되어 있다. 데이터웨어하우스(DW, Datawarehouse) 관점에서 보면 크게 기상 · 기후 DW, 식품안전 관측 DW 그리고 CC-RPS(Risk Profiling System due to Climate-change) DW 3가지 종류의 데이터베이스를 기초해 구성되어 있다.

첫째, 기상 · 기후 DW의 경우 평균 · 최고 · 최저기온, 강수량, 습도, 신적설, 풍속, 최대풍속, 최대풍향, 황사, 일조, 일사량, 운량, 태풍 및 기상재해 등 기상 변수에 대하여 권역단위, 시간(일, 주 및 월) 단위 병합 분석 기능을 자체적으로 내장하고 있으며 황사나 신적설 등 기상이변의 발생 유무 등 연관 통계 분석 기능을 갖추고 있으며 RCP 시나리오가 기반인 기후 예측 모듈을 내장하고 있다.

둘째, 식품안전 관측 DW의 경우 세균, 바이러스, 기생충, 곰팡이독소, 패독, 동물용의약품, 중금속, 잔류농약 및 해충 등 식품 위해인자에 대하여 위해인자 중심 식품관측 패턴분석 및 통계분석 기능과 식품을 중심으로 위해인자에 대

표 3. CCFS-DMC 주요 기능

대메뉴	중메뉴	소메뉴	주요 기능(Features)
메인화면			시스템 소개 및 메인화면
정의/특성	기상인자		기상인자(ex:강수량, 습도 등) 소개화면
	식품	개요	식품기술 소개화면
		현황	식품등록 현황 조회화면
		조회	식품 분류별 조회 및 관리화면
	위해인자	개요	위해인자 소개화면
		현황	위해인자 현황 조회화면
		탐색	위해인자 분류별 조회 및 관리화면
관측DB 현황	자료원 개요		자료원별 구축현황 조회
	총괄현황		관측DB 총괄 구축현황
	식품군별		식품군별 관측DB 지식탐색
	위해인자별		위해인자별 관측DB 지식탐색
	자료원별		자료원별 관측DB 등록 지식탐색
	현황DB		해충/항생제 데이터의 지식탐색
관측DB 분석	DB검색		관측DB 원시데이터 검색 화면
	특정 위해인자 중심		자료원 · 위해인자별 관측DB의 통계분석
	특정 식품 중심		자료원별 식품별로 관측DB의 통계분석
기상DB 현황	권역안내(2011)		2011년도 권역안내
	권역안내(2012)		2012년도 권역안내
	총괄현황		기상DB 의 총괄 데이터 현황
	기후예측(RCP4.5)		기후예측(RCP4.5)
	기후예측(RCP8.5)		기후예측(RCP8.5)
기상DB분석	원자료(1st)		기상 원시데이터 조회화면
	권역별DB검색(2nd)		권역별 기상데이터 분석(일/주/월)
DM도구	병합분석1		관측DB와 기상DB의 병합분석
	병합분석2		현황DB와 기상DB의 병합분석
CC-RPS	분류검색		위해인자별 Risk Profile 관리화면
	간편검색		CC-RPS의 단순 검색화면
	상세검색		CC-RPS의 항목별 상세검색화면
	출처 관리		CC-RPS의 출처관리화면
	CC-RP관리		CC-RP관리
빅데이터 분석	기상지도		위해인자별 활성화 기상 조건 지역별 분포 조회
	RCP지도		RCP 시나리오별 위해인자별 활성화 기상 조건 지역별 분포 조회
	오염도지도		위해인자별 오염도 지역별 분포 조회
	활성화리스트		위해인자별 활성화 기상조건 입력 및 조회

표 4. CC-RPS 주요 변수

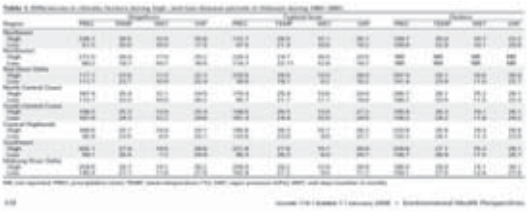
구분(국문)	구분(영문)	컬럼명(국문)	예시
환경변수	Environment	노출국가	일본
		노출도시	후쿠오카
		노출장소	
		노출시작년도	1999
		노출종료년도	2007
		노출식품(군)코드	
원인변수	Cause	기상변수코드	상대습도
		기상주기	일간
		기상환경명칭	대기
		기상환경코드	
전이변수	Transition	**식품위해인자코드**	
		조화코드	식약처 식품공전 및 국민건강영양조사 코드 (관리자수행)
		변화코드	
		변화량 표시용	
		변화량 계산용	
		변화량단위	
		결과영상설명	베트남에서의 세균성 장질환 발병 사례
		결과영상	
결과변수	Result	**질병코드**	전염성위장병
		변화코드	감소
		변화량 표시용	1
		변화량 계산용	
		변화량단위	%
		결과영상설명	
		결과 영상	
출처변수	Reference	참고문헌 요약	
		출처명	[a] D.Onozuka, M. Hashizume and A. Hagihara, (2009), Effects of weather variability on infectious gastroenteritis, Epidemiol. infect., page 1 of 8.
		출처 URL	
		국가코드	일본
		연구기관코드	
		참고문헌 공신력	
		증거문서 전문	

그림 7. 기상정보 심층 분석 화면(예시)

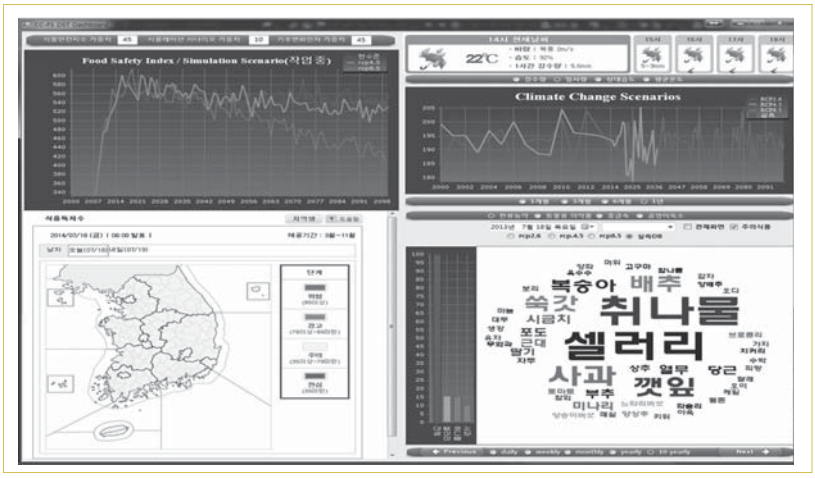


그림 8. 기후변화 예측 모형에 따른 식품위해인자 위해도 예측화면

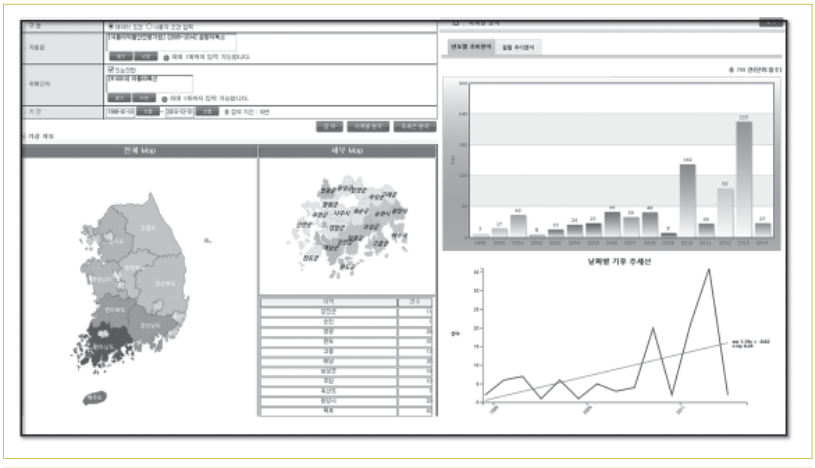


그림 9. 식품위해인자의 생육환경기반 시계열 및 공간 분석 화면

한 다양한 분석 기능을 가지고 있다.

셋째, CC-RPS DW의 경우 문헌상에서 조사된 다양한 기후변화에 따른 위해인자별 위해상황에 관한 지식 DB를 구축할 수 있는 시스템으로 기후환경과의 인과관계에 대한 시각화를 지원하며 입력된 자료는 Text-Mining을 통한 문헌 클러스터링 처리가 된다.

넷째, 기상 · 기후 DW와 식품안전 관측 DW 간의 다양한 추출 및 의사결정 지원을 위한 데이터마이닝 도구를 지원하며 최종적으로 해당 분야 식품안전의 위해요소 전문가들의 영향평가 모델 개발을 지원한다. 특정 위해인자는 지연효과(Lag-effect)로 인하여 기상관측 시간과 위해인자 관측시간 간 교차 병합이 필요하며, 다양한 시간 간격(예: 월, 주, 일) 단위에서 최고, 최저, 평균 등 다양한 형식의 자료 및 패턴분석을 할 수 있는 기능을 가지고 있다. 또한 지리정보시스템 및 빅데이터 기술을 접목하여 위해인자의 기후변화에 따른 위해발생 우려지역에 대한 사각지대를 발굴 지원을 할 수 있는 시스템이 마련되어 있다.

결 론

기후변화 대응을 위한 다양한 식품위해인자의 영향 평가 및 미래 예측 나아가 적응을 위해서는 대용량 자료(소위, Big data)의 처리 및 시뮬레이션 기술(수집, 저장, 분석 및 시각화)이 요구되며 이를 위해서는 ICT(Information and Communication Technologies)

와의 융합이 절대적으로 필요하다.

CC-MIMS 및 CCFS-DMC는 기후변화라는 위해상황에 대한 재난 예측 및 적응을 목적으로 정책입안자 내지는 연구자들의 의사결정을 지원하기 위해 개발되었고 기상·기후와 식품안전 간 융합 데이터베이스 모형 정립 및 대용량 자료 분석을 위한 컴퓨팅 기술 측면에서의 데이터마이닝 시스템의 마련으로 지속가능한 기후변화 대응 및 적응에 활용될 것으로 생각된다.

다만 해당 자료들이 서로 다른 정부부처에서 생산되다 보니 데이터 연계에 따르는 경제적인 비용 문제와 정보 부족(data gap)으로 인한 예측의 한계점이 존재하고, 연계가 가능하다 하더라도 분석 후 발생하는 불확실성 역시 많이 발생하고 있다. 이를 개선하기 위해서는 기존에 우리나라에서 부처별로 진행되고 있는 식품안전 모니터링 프로그램에서 기후변화를 고려한 샘플링 계획의 개선 및 생산부터 소비까지의 식품안전 관측 자료의 추적성을 보증할 수 있는 제도 내지는 표준 연계체계 마련이 선결되어야 할 것으로 생각된다.

기후변화와 식품안전 간 융합 데이터베이스 및 데이터 마이닝 시스템 구축의 최종 목표는 기후 변화(장기적 측면)와 변동(단기적 기상이변)으로 야기되는 미래 식품안전 분야의 영향을 정확하게 예측하고, 악영향을 최소화 및 예방할 수 있도록 식품안전 정책을 마련하는 것이며 이를 실현하는 수단으로 최신 ICT 기술인 빅데이터 기술이 접목된 사전 조기 경보 시스템(Proactive early warning system) 진화가 향후 필수적이라고 사료된다. 기후변화 연구 분야의 경우 단기적인 투자로 효과를 보기 가장 어려운 분야 중 하나인 만큼 장기적 관점에서의 로드맵을 통한 정부의 투자와 관심이 절실하다고 하겠다.

감사의 글

본 연구는 2015년도 식품의약품안전처의 연구개발비(15162식품안063)로 수행되었으며 이에 감사드립니다.

참고 문헌

1. IPCC, Climate change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press

2. 이헌주 외, 켐아이넷(주), 기후변화 대응 식품안전 DB구축 및 지식탐색 연구, 연구결과보고서, 2012-2014

3. 김신호, 한국기상산업진흥원, 빅데이터 관점에서의 기상정보와 타산업간의 융합방안, 2013

4. 이헌주 외, 켐아이넷(주), 기후변화 정보자료 수집·분석 및 통계기반 구축, 연구결과보고서, 2010-2012

5. 이화식, 엔코아정보컨설팅, 새로쓴 대용량 데이터베이스 솔루션 Vol. 1, 2005

6. 강현철 외, 자유아카데미, 데이터마이닝, 2001