

# 빅데이터 기반의 식품 정보 분석 시스템의 설계 및 구현

## Design and Implementation of a Food Information Analysis System Based on Big Data

임종태, 류은경, 김천중, 박재열, 이수지, 최기태, 북경수, 유재수\*  
 충북대학교 정보통신공학과

Jongtae Lim, Eunkyung Ryu, Cheonjung Kim, Jaeyeol Park, Suji Lee, Kitae Choi, Kyoungsoo Bok, Jaesoo Yoo\*  
 Chungbuk National University

### 요약

정부 3.0이 출범하면서 공공 데이터들의 개방이 활성화되기 시작했다. 공공 빅데이터로부터 다양한 분석을 수행하고 의미 있는 정보들을 도출하기 위해서는 빅데이터 분석 도구를 이용한 복잡한 분석들을 수행하여야한다. 본 논문에서는 빅데이터 분석을 이용한 식품 정보 분석 시스템을 설계하고 구현한다. 설계한 시스템은 식품에 대한 정보를 분석하여 특징에 따라 분류하고, 빅데이터 분석을 통해 실생활에서 구매자에게 유용한 서비스를 제공하는 것을 목적으로 하고 있다.

## I. 서론

정부 3.0이 출범하면서 공공 데이터들의 개방이 활성화되기 시작했다. 공공 데이터는 다양하게 생산되는 빅데이터 중에서도 가장 가치 있는 데이터 중 하나로 여겨진다. 연구자들은 안전행정부가 운영하는 공공데이터포털[1]이나 서울특별시가 운영하는 서울열린데이터광장[2] 등을 통하여 공공 데이터를 쉽게 접할 수 있게 되었으며, 제공되는 공공 데이터를 활용하여 다양한 분야의 응용 서비스를 개발할 수 있게 되었다. 과거부터 현재까지 데이터로부터 의미 있는 새로운 정보를 도출하는 연구들이 진행됐다. 하지만 시맨틱 웹과 같은 인터넷 기술과 스마트폰, 태블릿PC와 같은 모바일 기기의 발달과 함께 빅데이터[3]라는 개념이 정립되었다. 이러한 빅데이터로부터 기존의 데이터 분석 도구들을 이용하여 의미 있는 정보를 도출하는데 한계가 발생했다. 빅데이터로부터 다양한 분석을 수행하고, 그 결과 단편적인 정보뿐만 아니라 기존에 알지 못했던 의미 있는 정보들을 도출하기 위해서는 기존 데이터 분석과는 다른 새로운 접근 방법이 필요하다. 특히, 사용자 또는 연구자들이 전혀 알지 못했거나 실험적으로 증명되지 못한 가설들의 증명을 위해서는 빅데이터 분석 도구를 이용한 보다 복잡한 분석들을 수행하여야한다.

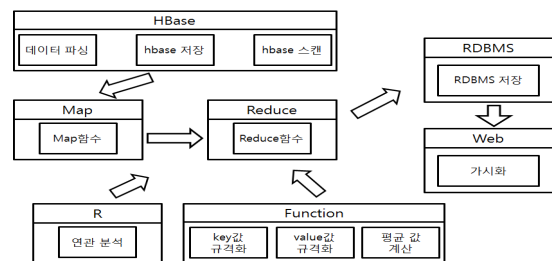
본 논문에서는 빅데이터 분석을 이용한 식품 정보 분

석 시스템을 제안한다. 제안하는 시스템은 식품에 대한 정보를 분석하여 특징에 따라 분류한다. 또한 제안하는 시스템은 빅데이터 분석을 통해 식품의 가격 추이 및 영향을 미치는 요인을 파악함으로써 실생활에서 구매자에게 유용한 서비스를 제공하는 것을 목적으로 하고 있다.

## II. 식품 정보 분석 시스템 설계

### 1. 전체 시스템 구조

그림 1은 제안하는 시스템의 전체적인 프로세스 흐름도를 보여준다. 제안하는 시스템은 하둡(Hadoop)과 HBase를 기반으로 구성했다. 다양한 출처로부터 실시간으로 수집되는 데이터는 파서(Parser)를 통해 제안하는 시스템에서 사용할 수 있는 데이터로 가공된다. 가공된 데이터는 HBase에 저장된다. HBase에 저장된 데이터는 Mahout 또는 R과 같은 빅데이터 분석 도구와 연동된다. 빅데이터 분석을 통해 나온 의미 있는 결과들은 웹서비스를 통해 사용자에게 제공된다. 사용자는 그래프 또는 표와 같이 직관적인 인터페이스를 통해 빅데이터 분석 결과들을 확인할 수 있다.



▶▶ 그림 1. 제안하는 시스템의 프로세스 흐름도

\* 교신저자 : yjs@chungbuk.ac.kr

본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학IT 연구센터육성 지원사업(NIPA-2014-H0301-14-1022), 교육부와 한국연구재단의 지역혁신인력양성사업(No.2013H1B8A2032298), 그리고 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원(No. 2013R1A2A2A01015710)을 받아 수행된 연구임.

## 2. 데이터베이스 스키마

그림 2와 3는 제안하는 식품 정보 분석 시스템을 위한 데이터베이스 스키마를 나타낸다. 제안하는 식품 정보 분석 시스템에서 사용되는 데이터는 공공데이터포털과 공공기관 홈페이지 등에서 직접 수집한 정보를 사용한다. 제안하는 시스템은 국내 농산물과 해외 농산물 정보를 활용하여 품목별 월별 평균 가격 분석, 국내 품목 가격에 영향을 미치는 요소, 그리고 국내 농산물과 해외 농산물의 가격 연관성 등을 분석한다. 국내 농산물은 총 11개 컬럼으로 구성된다. 본 논문에서는 국내 농산물 컬럼을 6개의 컬럼 패밀리로 구분한다.

	price			date	class		maket		grade	tradeamt	
	max	min	avg		mclass	sclass	coname	maketname		grade	amt
rowkey	280	110	195	11-30	bean	bean	법인	시장	1	3	4kg

▶▶ 그림 2. 농산물 가격 데이터베이스 스키마

해외 농산물은 총 11개의 컬럼으로 구성된다. 본 논문에서는 해외 농산물 컬럼을 5개의 컬럼 패밀리로 구분한다. 국내 농산물과 해외 농산물은 rowkey를 통해 연결된다.

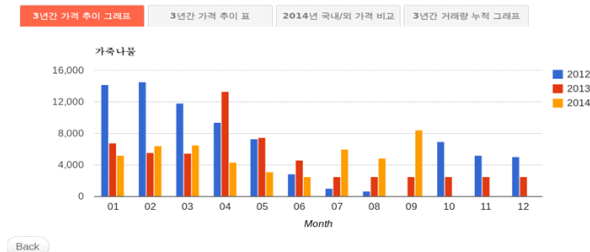
	price			date	class			grade	unit
	max	min	avg		5d-avg	year-avg	lclass		
rowkey	280	110	195	11-30	fruit	banana	banan	1	13kg

▶▶ 그림 3. 해외 농산물 가격 데이터베이스 스키마

## III. 식품 정보 분석 시스템 구현

그림 4는 제안하는 식품 정보 분석 시스템의 가격 추이 분석 그래프를 보여준다. 예제는 가축나물에 대한 월별 가격 변화를 구글 그래프 API[3]의 막대 그래프를 이용하여 표현했다.

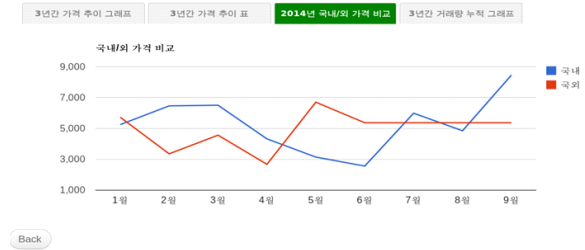
품목 : 가축나물



▶▶ 그림 4. 가격 추이 분석 그래프

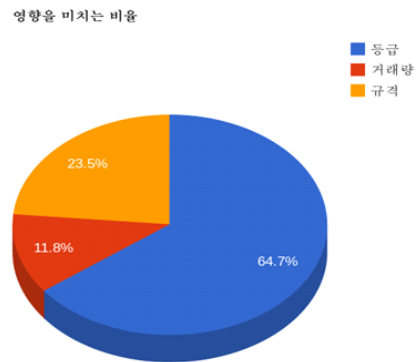
그림 5는 제안하는 식품 정보 분석 시스템의 국내외 가격 비교 분석 그래프를 보여준다. 예제는 가축나물에 대한 국내외의 가격 변화를 구글 그래프 API의 꺾은선 그래프를 이용하여 표현했다.

품목 : 가축나물



▶▶ 그림 5. 국내외 가격 비교 분석

그림 6은 제안하는 식품 정보 분석 시스템의 가격에 영향을 미치는 비율 분석 그래프를 보여준다. 예제는 가격과 등급, 거래량, 규격을 연관 분석하여 가격에 영향을 미치는 비율을 구글 그래프 API의 원통 그래프를 이용하여 표현했다.



▶▶ 그림 6. 가격에 영향을 미치는 비율 분석

## IV. 결론

본 논문에서는 빅데이터 분석을 이용한 식품 정보 분석 시스템을 설계하고 구현하였다. 설계한 시스템은 식품에 대한 정보를 분석하여 특징에 따라 분류한다. 또한 제안하는 시스템은 빅데이터 분석을 통해 식품의 가격 추이 및 영향을 미치는 요인을 파악함으로써 실생활에서 구매자에게 유용한 서비스를 제공하는 것을 목적으로 하고 있다. 향후 연구로는 식품의 가격에 영향을 미칠 수 있는 날씨, 주요이슈 등 보다 다양한 연관 분석을 수행할 것이다.

### ■ 참고 문헌 ■

[1] <http://www.data.go.kr/>  
 [2] <http://data.seoul.go.kr/>  
 [3] P. Zikopoulos, C. Eaton, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data", McGraw-Hill Osborne Media, 2011.  
 [4] <https://developers.google.com/chart/>