# Design of Disaster Collection and Analysis System Using Crowd Sensing and Beacon Based on Hadoop Framework

Eun-Su Mo[1], Jae-Pil Lee[1], Jae-Gwang Lee[1], Jun-Hyeon Lee[1],
Young-Hyuk Kim[2], and Jae-Kwang Lee[1](✉)

[1] Department of Computer Engineering, Hannam University, Daejeon, South Korea
{esmo,jplee,jglee,jhlee,jklee}@netwk.hannam.ac.kr
[2] Public Procurement Service, Information Management Division, Daejeon, Korea
kimyh86@korea.kr

**Abstract.** Currently, disaster data is collected by using site-based, limited regional collection. In this study, a system that collects location information of users that have a mobile device is proposed. The proposed system collects real-time disaster data by using crowd sensing, a user-involved sensing technology. In order to quickly and accurately determine a large amount of unstructured data, among big data frameworks, the Hadoop framework is applied as it efficiently sorts a large amount of data. Also, to enable fast local evacuation alert for users, a beacon-based ad-hoc routing interface was designed As an integrated interface of the proposed systems, a hybrid app based on HTML5, which uses JSON syntax.

**Keywords:** Crowd sensing · Big data · Beacon · Disaster · Calamity · Warning system

## 1 Introduction

In general, disaster refers to a naturally-caused accident, while calamity means artificially-caused one. The former includes typhoon, flood, storm, tsunami, heavy snow, and yellow dust and the latter fire, breakdown, explosion, traffic accident, malfunction of national infrastructure, and infectious disease [1].

Many accidents from disaster and calamity are taking place all around the world. In general, disaster and calamity result in a wide range of damage and loss and spreads quickly. Therefore, it is crucial to secure the “golden time” to reduce danger.

However, the current system is focused on reporting rather than problem-solving, and, therefore, unable to effectively secure the golden time [2].

In Japan, which has the best system related to meteorology and earthquake technology, March 11, 2011, earthquake of 9.0 magnitude hit the Pacific Ocean near Tohoku, Japan, resulting in over-10m tsunami and some of the worst damages in history, including deaths, property damage, and radiation leakage The first early warning for earthquake and tsunami was sent three minutes after strong earthquake [3].


O. Gervasi et al. (Eds.): ICCSA 2015, Part IV, LNCS 9158, pp. 106–116, 2015.
DOI: 10.1007/978-3-319-21410-8_8

South Korea has total 11,542km of coastline, which is relatively long for the land, in comparison to other countries. The coastline length of South Korea compared to the area, in percentage, is 117%, which is much higher than Japan 87%, UK 57%, and New Zealand 56%. For that reason, South Korea suffers more serious damage from disaster and calamity than other countries. However, the country currently has many problems regarding disaster and calamity detection and subsequent evacuation of residents [2] [4].

Information related to all accidents is referred to as safety information or safety service information. Research has been conducted to take measure against increasing safety-related accidents. Most notably, in IoT (Internet of Things), importance, value, and utility of big data are increasing. Recent research is focused on how to integrate and converge big data into a system.

Most cases of integrating big data into safety service are concentrated in crime analysis, such as crime analysis and domain awareness system of New York and crime prediction service of San Francisco and Los Angeles.

New York has built an anti-terrorism detection system named DAS (Domain Awareness System). Over 4,000 surveillance cameras installed in Manhattan are used to analyze information on suspicious people, objects, or cars, which is immediately provided to local police, fire station, or other relevant institutions [5].

San Francisco analyzes past crime data to predict areas with highest risks of new crimes. Among 10 districts that were warned of possible crimes, seven actually saw criminal cases [5].

In South Korea, geographical profiling service of National Police Agency is one of the safety services that integrated big data, which also include SOS Resident Safety Service of Ministry of Security and Public Administration, and iNavi Safe [6].

Safety-related information and service in most of the current services based on big data are only focused on crimes, and irrelevant to natural disaster and calamity.

For that reason, in this study, the following system was designed for the purpose of collecting and analyzing disaster and calamity data and the system is divided into three steps: collection, analysis, and provision.

The main component of this three-step system is crowd sensing, a user-involved sensing technology, for collecting real-time disaster and calamity information.

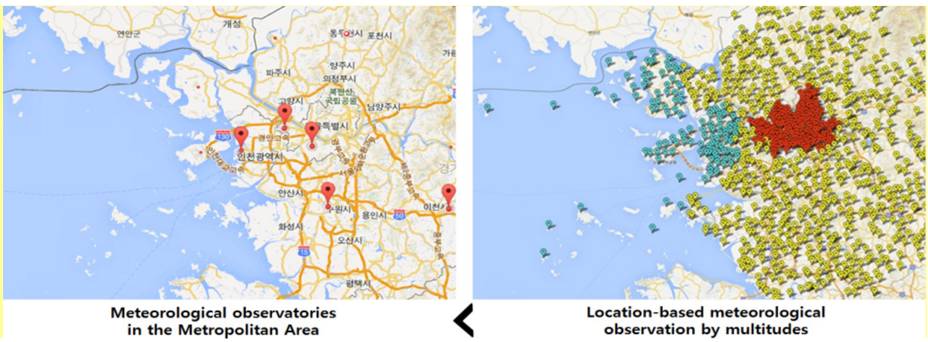


**Fig. 1.** Comparison of data collection between conventional data import and crowd sensing [7]

Crowd sensing is a compound word of crowd and sensing and collects data as users who have sensors and computing devices participate, creating new knowledge by extracting sensor data from a wide area. [Figure 1] shows the core of crowd sensing. It is 'the multitudes in different regions' and not 'a small number of experts' who can find out about rapidly-changing information of an area in the most accurate and fastest way. Their collaboration brings about accurate and fast data.

In Chapter 2, previous research on crowd sensing, big data, and beacon, which are core technologies in the three steps, will be reviewed. And then, in Chapter 3, the front-end, back-end system will be designed to provide the target service. Finally, the authors will discuss advantages and problems of the design.

## 2 Related Studies

### 2.1 Analysis of Conventional Disaster and Calamity System

#### 2.1.1 Forest Fire Prediction Service Based on Big Data

The research in Korea on forest fire prediction service based on big data analyzed forest vulnerability according to climate change, and aimed to improve the national forest fire warning system that was used since 2003 [8]. This study has the following characteristics: First, location information of people who are hiking or climbing the mountain is collected from weather information provided by Korea Meteorological Administration (KMA), forest soil digital mapping, mountain slope, and direction, and data without personal information provided by telecommunication providers. Second, forest disaster is predicted and the information is sent to users based on the collected data. In the forest soil mapping, mountain slope, and direction, trails can change according to the weather and population in motion. These data must be re-established based on different periods and require a lot of time and workforce. In this study, crowd sensing and smartphone sensors will be used to receive transformed data.

#### 2.1.2 Heavy Rain Prediction System of Rio de Janeiro

Brazil's Rio de Janeiro heavy rain forecast system is a large-scale study initiated due to a natural disaster forecast system suffered casualties due to landslides caused by

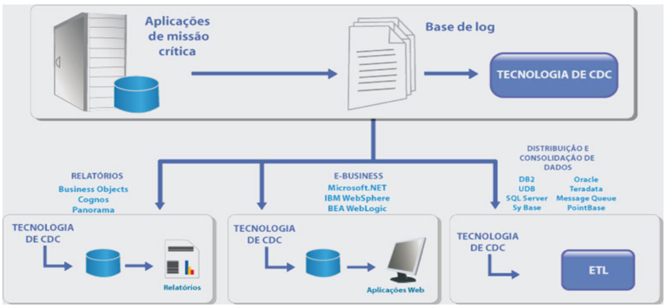


**Fig. 2.** Structure of heavy rain prediction system of Rio de Janeiro [9]

heavy rainfall concentrated in 2011. [Figure 2] shows the system structure. Data and processes of more than 30 institutions in Rio de Janeiro were integrated into an intelligent system, in order to manage and supervise natural disaster, traffic, and power supply. The intelligent operation system enables prediction of heavy rain 48 hours in advance based on urban management and high-resolution weather forecast system and state-of-art modeling system. Unlike the initial design, the system expanded in its scope to all areas, crimes, and accidents that can take place in the city [10].

## 2.2 Crowd Sensing

### 2.2.1 Semantic Map Research

In semantic map research using crowd sensing, data including operation types, customer visit patterns, and sale characteristics of stores near universities was collected by using crowd sensing and implemented in a map for visualization

Stores near Yonsei and Ehwa Womans' Universities show clear characteristics as shown in [Figure 3]. Near Ehwa Woman's University, most of the businesses are stores for shopping. By contrast, near Yonsei University, businesses are focused on night life and entertainment. Behind the universities are crowded by accommodations, and workplaces are distributed close to the universities. Verifying and researching this type of data by an individual or small group requires a substantial amount of time and effort, and this study demonstrated that crowd sensing helps collect valuable and accurate data quickly and even visualize it [11].

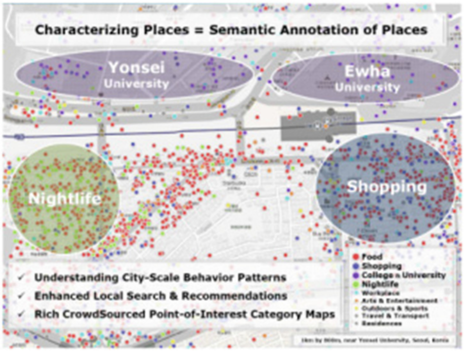


**Fig. 3.** Semantic map visualization using crowd sensing [11]

### 2.2.2 Service for Finding Missing Children

The research designed to find missing children based on crowd sensing is focused on situations in which children who have a smartphone or smart tag are gone missing. When a child is missing from a theme park, department store, or in other large spaces, adjacent users are asked for help and sent the basic information about the child's appearance and message. Information about the route of the child is provided by using sensing data of the users. The server, then, compares data collected from the child's smartphone and data provided by users, and provides matching data so that the child can find the parents. This study helps find missing children by using crowd sensing [11].

## 2.3 Big Data

### 2.3.1 Security Log Analysis System Using Hadoop

In the research on security log analysis system using Hadoop, the security log of security equipment is supplemented based on SIEM (security information & event management) by using the Hadoop framework rather than the conventional ESM (Enterprise Security Management). So far, logs from security equipment were analyzed by individual equipment or analyzed as a whole based on ESM. However, it was not possible to analyze a large amount of data that was accumulated over a long period, due to limited data capacity and data processing of ESM, or search data quickly. These limitations were supplemented by using Hadoop framework, which enables fast collection, storage, analysis, and visualization of data. [Figure 4] shows composition of Big Data Platform Software [12].

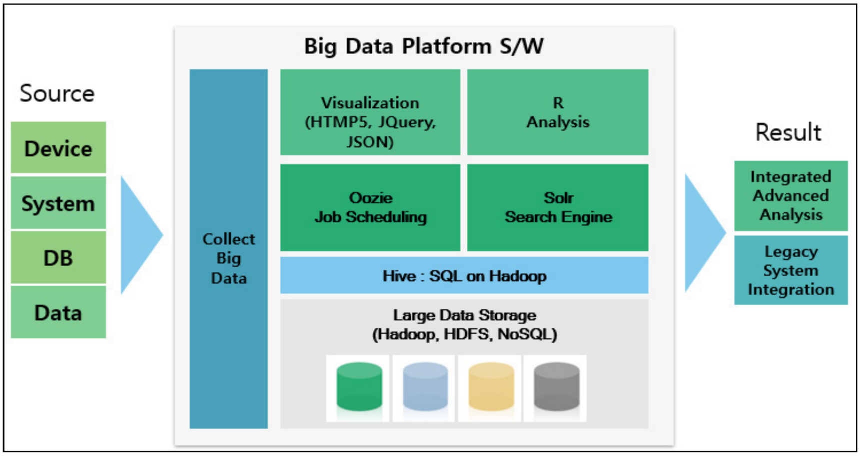


**Fig. 4.** Big Data Platform S/W [12]

### 2.3.2 Reduction of Network Use by Hadoop Cluster

The research on block relocation algorithm was designed to reduce network consumption by reducing data locality, which is weakness of Hadoop cluster.

The job scheduler of map/reduce uses FIFO (First In First Out) to sequentially assign tasks. This method has low processing efficiency when the internal/external data of the cluster is called when there is a large amount of tasks. Because the scheduler operates in a delayed state, that is, when the initially located blocks are not relocated, there is difference in load received by different nodes. The relocation algorithm is used to improve efficiency by redistributing node blocks according the cluster work pattern [13].

### 2.4 Beacon Structure and Service Trend

Beacon is a type of wireless sensor that applies real-time location awareness system.

It is based on the Bluetooth communication protocol and uses BLE (Bluetooth Low Energy). Location is measured by exchanging data with smartphones via RSSI (Received Signal Strength Indicator). Although smartphone LBS (location-based service) using NFS and GPS and real-time data communication technology has been introduced, it has the advantages of location accuracy and scope in comparison to indoor location determination technology. NFC can be used with 1:1 contact within 10cm, Beacon has a long available distance of maximum 50 to 70m. While GPS signals cannot locate a smartphone user, Beacon is capable of accurate location, by 5cm, and can be used both indoors and outdoors. Positioning service can be divided into three methods: checkpoint, zone, and track [14] [15].

Beacon service has been adopted at the main store and Icheon store of Lotte Premium Outlet. The smartphone app is automatically recognized by the sensor, and provides welcome message, store map, and shopping information to the visitors [15].

## 3 Design and Implementation

### 3.1 System Model

[Figure 5] is a block diagram of the system for the proposed service.

Section A collects data from users and receives data provided by the control center.

Section B is a control center that manages I/O of the entire data. All data must pass the control center and, as it is distributed and arranged, the level of load balancing is important.

Section C is a backbone where data is stored, analyzed, and processed. To quickly and accurately process a large amount of data, distributed processing based on Apach Hadoop Framework is performed. Also, Hadoop has high fault-tolerance and, thereby, system stability.

The proposed system is executed as follows:

In the first step, collection, crowd sensing is used for real-time collection of disaster and calamity data. One of the prerequisites for this is that the user has a mobile device. For a wider range of users, an HTML5-based hybrid app is provided to expand applicable mobile devices. Second, the latest BLE is used to limit system resources used for user participation. Third, BLE-based direct communication network is proposed to create a user network within a site, even if there is a problem with connection with the control center.

In the second step, analysis, big data framework is used to quickly and accurately detect and determine a large amount of unstructured data collected from users. The Hadoop framework is applied to the framework applied in this study as it is effective for large data arrangement. Hadoop arranges data in couple based on the key value, showing high speed in crowd sensing and Beacons data arrangement and processing. Particularly, this advantage is useful for larger amounts of data, when there are more users and, thereby, collected data.

In the last step, Beacon is used for sending prompt evacuation warning to local users based on the analyzed data.

BLE-based Beacon uses less battery than the conventional Bluetooth version. The relevant service is provided as it automatically locates the users without users' action. Also, location is determined by using RSSI. In this study, information is provided so that users can prepare for and evacuate from unspecified situations, by using the user network (hereafter, BLE Ad-hoc routing). JSON is used for communication between these user systems. Unlike XML, JSON does not use tags but text-based exchange with key values. It has the advantage of not relying on language.

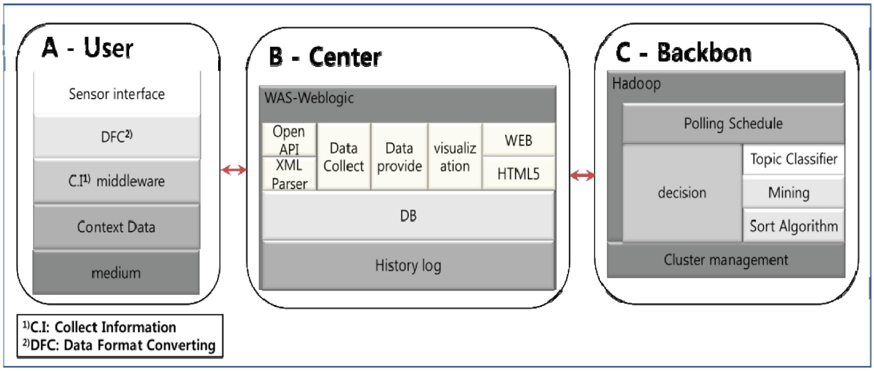


**Fig. 5.** Service Model

### 3.2 Data Collection

[Figure 6] shows the process that collects and stores data by crowd sensing. All sensing data collected from users' mobile devices are converted into JSON by the DFC module. The converted data combines users' unique IDs, and, thereby, verifies data destination. Finally, the finished data is sent to the control center through either BLE or network. Also, users can exchange data and user information within the same site by using BLE Ad-Hoc on Beacon.

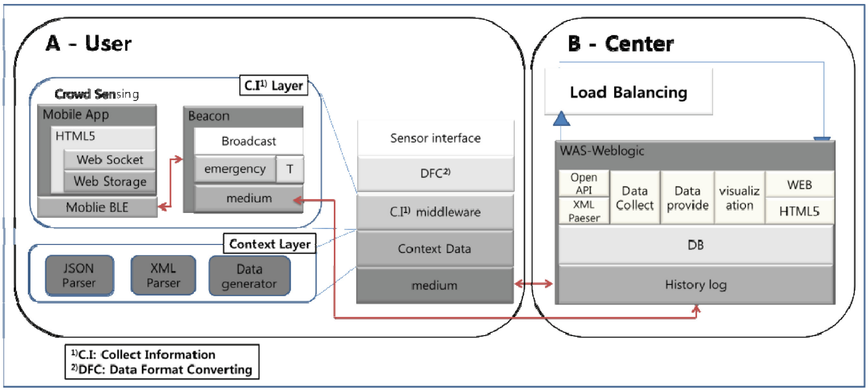


**Fig. 6.** Data collection process

### 3.3 Data Analysis

[Figure 7] shows the flow by which all data is sent through the control center to the backbone, for distributed storage. The control center manages and distributes all data. It is responsible for receiving and sending all data. The control center stores the received data in DB, and this is included in record management in the history log.

In the backbone, Section C, the polling scheduler performs real-time monitoring of DB of the control center. It takes newly stored data for distributed storage in multiple clustered nodes via Map/Reduce.

For storage, the topic classifier performs data mining, and, as main categories, the topics are classified as climate, earthquake, fire, typhoon, and tsunami. After the first classification, specific topics are classified by a top-down method.

Semantic resolution of the data divided into the minimum unit is performed by using data mining. For fast semantic resolution, the pattern algorithm that defined particular patterns is used for filtering.

Finally, data of distributed nodes is stored, and, cluster managing algorithm, which arranges data by grouping them based on nodes, is used for distributed storage.

As a distributed storage method, Hadoop Framework, which is an Apache Open Project, is used. The system was composed of the master server, where the polling scheduler is operated, and slave nodes according to Hadoop standard. Disaster and calamity data are distributed-stored by Map/Reduce, and the structural characteristics allow the system to operate without a problem even if there is a small amount of data node error.

Especially, the Map/Reduce method uses API or its own query language that is different from the conventional SQL query method, because it is processed in a key-value form.

Lastly, nation-level public data is additionally collected by using OpenAPI (OPEN Application Programming Interface) of public data portals relating to disaster and calamity. This supports reliability of the proposed system in that it can freely use highly reliable, large-scale data in real time. Also, Open API is provided as a method that facilitates implementation. Notably, SBIS (System Build & Integration Service) Weather API, a climate/weather distribution service of KMA that was used as reference in this study, has the following characteristics: First, the weather data is made into XML according to the relevant website. Second, to minimize users' burden, the data is taken in real-time XML instead of being stored in the server.

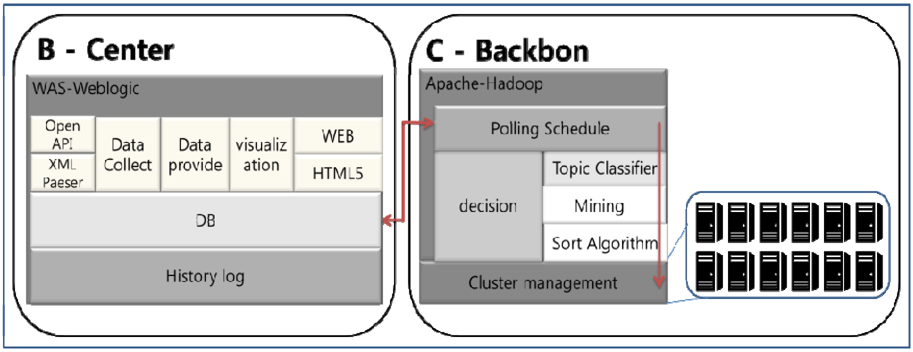


**Fig. 7.** Data processing

### 3.4 Service Scenario

The disaster and calamity warning scenario of the proposed system assumes that the backbone, Section C, determines disaster or calamity, and can be divided into three types as shown in [Figure 8].

In the first scenario, the decision server that exists in Section C of [Figure 5] detects disaster or calamity, and sends the data via the control center in Section B to Beacon and, finally, to users' mobile devices. Beacon devices have their unique IDs and the Beacon of the relevant site sends them via BLE Ad-Hoc to Beacon and users' mobile devices.

In the second scenario, the decision server detects disaster and calamity and directly warns the users through the control center. Users' mobile devices have unique IDs and the warning is sent based on the user mobile device ID and collected location data. This method implies that data can be sent even if the local Beacon does not function normally. Users' mobile devices recognize the special situation in which data was directly sent without Beacon, and continue to the following, third scenario.

The final scenario is activated by the second scenario. Without the decision/control server, users warn disaster and calamity on their own, when the infrastructure does not function normally during disaster and calamity. In this situation, the data is sent directly to the local Beacon and users from the time disaster or calamity was notification was received.

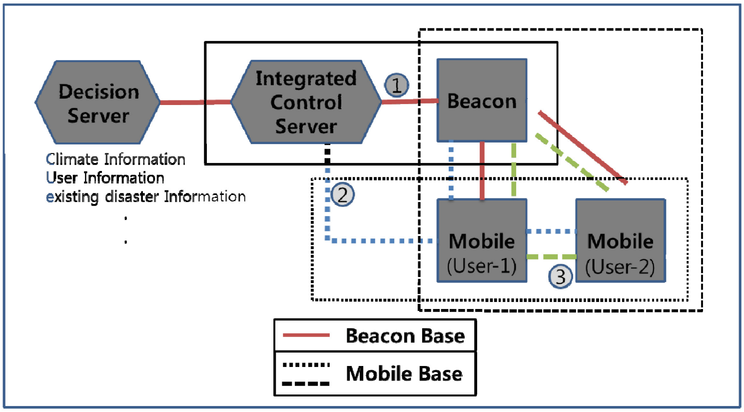


**Fig. 8.** Provision of service

# 4 Conclusion

This study proposed a system that subdivides disaster and calamity by town units, and collects data quickly and accurately. The entire process was defined by dividing it into three steps of collection, analysis, and provision. Golden time is especially important in disaster and calamity. For this, crowd sensing was used to improve reliability by integrating user-based location data and local data using OpenAPI. All data uses the control center, the central server, which links users and large-scale backbone clusters. Therefore, all data passes through the control center for distributed storage by Hadoop. In Hadoop framework, the polling scheduler, which is a master that monitors new data of the control server, is placed above, and, below it, slave nodes process big data.

JSON structure is used for data exchange between systems, and allows users to use their mobile devices. Especially, as the service is provided through a hybrid app. It can be provided to various devices and OS. The system was designed to deliver data from the control center to Beacons that consist of local infrastructure, and Beacons send the data to users via broadcasting. This is part of load balancing of the control center, and a backup plan was made so that, when there is an error, users can form a network and exchange data. As a result, this can radically address problems with conventional disaster and calamity system, and also detects disaster and calamity by small area basis.

However, to improve the service, the data processing algorithm of Hadoop must be researched so as to accommodate it to particular situations of disaster and calamity. Also, the three step scenarios proposed in this study have the disadvantage that they cannot determine users with ill intention. A policy to resolve this problem will be needed.

**Acknowledgment.** This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(No. 2014R1A1A2055522).

## References

1. Framework Act on the Management of Disasters and Safety
2. Hyeon-Cheol, S.: Improvement of Disaster Management System in Korea: The Institute for the Future of State (May 2014)
3. Korea Meteorological Administration: Mid-to-Long-Term Policy Development for Improved Meteorological Service (October 2011)
4. National Emergency Management: Plan for Mid-to-Long-Term Research and Development Against Earthquakes and Tsunamis (April 2013)
5. Jang, K.: Choi: Big Data-based Convergence Service Industry Creation: Science & Technology Policy Institute. Policy Research **2013–20**, 210–214 (2013)
6. Jang, Kim: Main policy and suggestions for Big Data-based Convergence Service Industry Creation; Science & Technology Policy **192**, 4–13 (2013)
7. Wearherplanet. http://www.weatherplanet.co.kr/about/
8. Sang-Woo, B.: Implementation of integrated systems for forest disaster management. The Korea Contents Association: The Journal of the Korea Contents Association **12**(2), 73–77 (2014)
9. MD2. http://www.md2net.com.br/ibm_InfoSphere_CDC.asp
10. BIGDATA Strategy Forum: Data analysis for a better future, National Information Society Agency (2013)
11. Ho-Jung, C.: Mobile CrowdSensing: Korea Internet Conference, TRACK J1-2 (2014)

12. Han, K.-H., Jeong, H.-J., Lee, D.-S., Chae, M.-H., Yoon, C.-H., Noh, K.-S.: A Study on implementation model for security log analysis system using Big Data platform: Korean Society of Computer Information. Journal of Digital Convergence **12**(8), 351–359 (2014)
13. Kim, J.-S., Kim, C.-H., Lee, W.-J., Jeon, C.-H.: A Block Relocation Algorithm for Reducing Network Consumption in Hadoop Cluster: The Society of Digital Policy and Management. Journal of the Korea Society of Computer and Information **19**(11), 9–15 (2014)
14. ITworld. http://www.itworld.co.kr/slideshow/85994
15. KB Financial Group Management Institute, KB Vitamin knowledge, vol. 94 (2014)