International Journal of Engineering & Technology, 7 (3.33) (2018) 174-178

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Design of Prevention Method Against Infectious Diseases based on Mobile Big Data and Rule to Select Subjects Using Artificial Intelligence Concept

Sunghyun KIM[1], Hyunseok HWANG[2], Jaeho LEE[3], Joonki CHOI[4], Jangmook KANG[5], Sangwon LEE[6]*

[1]K-ICT Big Data, National Information Soceity Agency, Seongnam, Korea
[2]Department of Business Management, Hallym University, Chuncheon, Korea
[3]Big Data, Korea Telecom, Seoul, Korea
[4]Big Data, Korea Telecom, Seoul, Korea
[5]Department of Big Data & Industry Security, Namseoul University, Cheonan, Korea
[6]Department of Computer & Software Engineering (Institute of Convergence Creativity), Wonkwang University, Iksan, Korea
**Corresponding author E-mail: sangwonlee@wku.ac.kr*

Abstract

The rapid evolution of transportation has enabled us to travel to any part of the world. At the same time, infectious diseases are now able to reach anywhere in the world via this same ease of travel. In 2015, a MERS epidemic broke out in Korea. The MERS virus infected 186 people with 38 dead and caused economic damage worth 6 billion dollars. In this paper, we investigate a joint project of KT and KCDC (Korea Centers for Disease Control and Prevention) to prevent infectious diseases through the use of mobile roaming Big Data. KT developed a monitoring system that uses telephone roaming data to identify subscribers who traveled to a country affected with an infectious disease. Upon returning to Korea, a subscriber receives a notice that he/she is required to report for potential quarantine in accordance with the regulations and gets guided on the measures to take in case symptoms of the infectious disease occur. The travel information is also provided to healthcare facilities throughout the country for reference when symptomatic individuals visit them. Laws and regulations are enacted to allow personal information to be used to prevent and control infectious diseases. By providing a solution for monitoring individuals with potential risk of having been infected, loss of life and financial loss caused by the spread of panic of infectious disease have been minimized. This study is an example of the significant social contribution of Big Data. The global propagation of this system can reduce the threat of the spread of infection significantly.

Keywords: *AI; Big Data; Disease Management; Infectious Diseases; Prevention Method; Telecommunication*

1. Introduction

Big Data refers to data sets that are much larger than the conventionally managed data size, and they are basically characterized as 3V: Volume, Variety, and Velocity [1]. To that, IBM added Veracity to emphasize data reliability; SAS added Variability to emphasize complexity due to the diversity of data source, and Oracle added Value to emphasize the value obtained from the analysis of a large volume of data [2]. The era of Big Data has been brought about by the rapid decrease of cost of using network and storing data, distribution of smartphones and dispersion of social network service, Internet of Things (IoT), and distribution of machine to machine (M2M) [3]. The concept of Big Data has evolved to encompass data management tool, platform, and analysis technique with focus on value creation and utilization [4]. The development of software technology needed for analysis with open source such as R and Python and the decreasing cost of general-purpose hardware have enabled the quick analysis of a large volume of data at low cost.

The 4th industrial revolution has recently become the key agenda to prepare for Korea's future. It implies the fundamental change of industrial structure with high increase of productivity through intelligent machines, and "data" has emerged as the basic material and tool for intelligence. Big Data led to the emergence of data convergence service -- which had been impossible before -- and the creation of new values. Comparable to petroleum, "data" is a key resource for the future. Like the steel and coal in the industrialization age and the information system in the information age, Big Data will be the key resources for social development in the upcoming intelligent information society [5]. In the era of business intelligence and analytics 3.0 that analyzes the huge volume of data coming from IoT devices, Big Data can greatly improve the public and health sectors [6]. In the health care area where unformatted data is the norm, Big Data can be applied to image analysis and new drug development [7, 8]. With the growing importance of Big Data, the development of data analysis tools and evaluation system has been emphasized as well [9, 10].

Big Data has been recognized as an essential asset for strengthening the competition ability in all areas since the governments and industries in many countries have paid attention to the potential of Big Data. The trend is clear as market research firm IDC reported that the size of the world Big Data market has grown to USD 32.4 billion in 2017, for an annual average growth of 27% [11]. Aided by higher awareness of Big Data by business, more investment in Big Data by large companies, and strong government commitment

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

to cultivate the Big Data industry, the size of domestic Big Data market has reached KRW 344.0 billion (USD 303 million) in 2017, for a 31.1% annual growth (Figure 1). Although the early Big Data market grew with government support, the market growth has recently been led by private investment due to the following: increased value of data to business; expansion of IT business such as open source SW, Big Data analysis, IoT, cloud, and security; emergence of AI mega trend; wider application of de-identified data, and; increased unformatted data. The growth of private investment is expected to continue for some time [12].

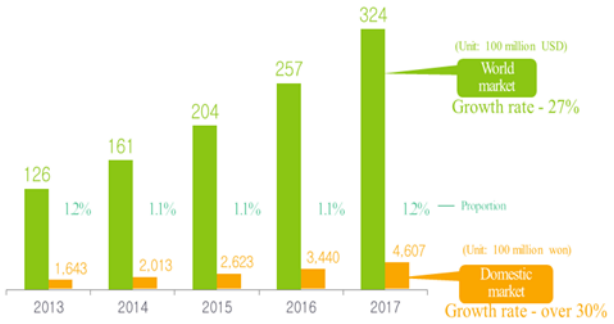

Fig. 1: Size of the Global and Domestic Big Data Markets

The McKinsey Global Institute selected health and medicine, public sector, and communication as emerging areas for Big Data application. According to McKinsey, Big Data can provide value of USD 300.0 billion and annual productivity increase of 0.7% in health and medicine in the US. In Europe, Big Data will provide value of EUR 250.0 billion and annual productivity increase of 0.5% in the public sector. The communication data that can provide location data will provide additional gain of USD 100.0 and value of USD 700.0 billion [3]. Since Big Data can provide such productivity increase and high value-added service, it has become a subject of interest for application in public services to address social problems in not only the US but also EU, China, Japan, and other leading countries.

2. Big Data and Infectious Disease Prevention

2.1. Prevention of Infectious Disease

Since the development of transportation has enabled movement to every corner of the world, infectious diseases can spread to anywhere in the world with traveling people as carriers. As such, quarantine authorities worldwide have established -- and are implementing -- quarantine measures for entrants [13]. Despite that, infectious diseases that began in a specific country have been continuously spread to other countries by people who visit those countries. The cholera that hit Peru in 1991 killed more than 3,000 people in the country, led to the total import ban of fishery products, and dealt a heavy blow to the trade and tourism industries in Peru because of decreased visitors. In 2012 ~ 2013, SARS broke out in the Gwangdong region in Southeastern China and killed 774 of a total of 8,096 people who contracted SARS, for a 9.6% death rate [14]. In 2014, the Ebola virus broke out in the Western Africa region, infecting 27,741 people and killing more than 11,000 [15]. As such, the Commission on Global Health Risk Framework (GHRF Commission) published a report in January 2016 to warn that there is more than 20% probability of one or more global-scale infectious disease occurring within the next 100 years, and that the annual economic loss will exceed USD 60.0 billion.

An infectious disease is spread in a region by infected people visiting the region. In particular, it is characterized by the fast progression of secondary and tertiary infections, and super-spreaders who spread the disease to many people emerge in the process. In Korea, the MERS virus first broke out in 2015 through a 68-year-old male who returned from a travel to the Middle East. It eventually led to 186 confirmed carriers and 38 deaths as well as economic loss of around KRW 6 trillion due to decreased visitors through secondary, tertiary, and quaternary infections [16, 17]. MERS spread without control in Korea because the first MERS carrier could not be monitored for infection since he returned to Korea via a clean country after visiting a country affected by MERS. The import and spread of infectious disease are characterized by first spreading through a person who visited the region affected by the infectious disease, and then spreading fast through several super-spreaders who had contact with the first spreader. As such, if the initial spread of infectious disease is not blocked quickly, it will spread fast throughout the local community, and it will soon be very difficult to block.

Communication data was utilized to contain the infectious disease by KT's proposal when MERS occurred. We Identified the subscriber who seems to have come in close contact with the confirmed MERS patient. And we then used as ancillary data of the epidemiological investigation, such as traffic line of the confirmed patient during the latent period. We also used real-time location tracking of the self-isolator when he/she is reported to leave the designated place. Article 58 of the Personal Information Protection Act was invoked for the MERS crisis, considering the fact that it was a national disaster situation (Figure 2).

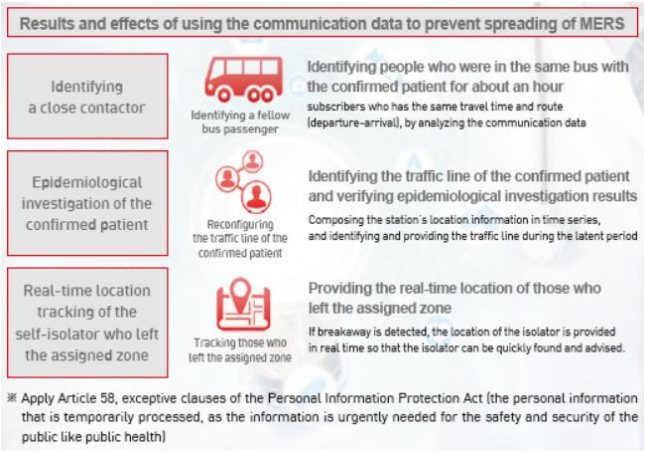

Fig. 2: Using the communication data to contain an infectious disease using the location information TF inside the Central MERS Countermeasure Headquarters

Follow-up measures were prepared to prevent recurrence when the MERS crisis was over. We decided to find a solution while the infectious diseases flowed into Korea first, because there is no way of containing it once it arrives and spreads in Korea. We checked the problem while tracing back how the No. 1 MERS patient was confirmed. (1) The patient has successfully passed the airport quarantine because no infectious symptom associated with MERS was found in the latent period. (2) The infected patient was not aware of MERS, and he did not associate his sickness with travel in the Middle East; neither did he disclose his travel during the medical diagnosis. (3) The doctor who was supposed to diagnose the patient with MERS diagnosed him with pneumonia instead because he did not know of the patient's travel to the Middle East. (4) The quarantine authorities also had difficulty confirming MERS in the early phases because they did not check the patient's visit to Saudi Arabia, a MERS-affected country. And, The Korea Centers for Disease Control and Prevention planned and executed the quarantine service for those who came from the contaminated country to solve the problem. (1) Those who visited the contaminated country: Providing information on the infectious disease that occurred in that area when visiting the contaminated country (2) Medical institution: Providing the history information on the pa-

tient's visit to the contaminated country, so that the information can be retrieved for treatment.

There was a problem, i.e., the quarantine service cannot be provided if a person arrives from a third country that is not affected by the infectious disease even though the person had visited the contaminated country before. In particular, if an emergency occurs due to the emergence of a particularly infectious disease, there is no means of contacting all the people who visited the pertinent contaminated area. Therefore, a countermeasure needs to be prepared. The mobile phone number should be updated when renewing the passport. Note, however, that the information is not updated for all passport holders.

2.2. Big Data Cases Related to Infectious Disease

Most countries including Korea provide services that predict which regions will be affected by an infectious disease in the future by analyzing the moving pattern of people who visited the region affected by the infectious disease. The joint project team of US NIH and University of Arizona used the mathematical model "SEIR" and divided the subjects into people who may be infected, people who are infected, people who are recovering from infection, and people who are in a latency stage for analysis and announced the correlation between the infection rate in a medical institute and the number of those infected when MERS broke out in the Middle East in 2013. They then considered the variable representing the possibility of infectious disease being spread worldwide by a passenger in an aircraft in addition to the latent period and predicted that the possibility of MERS from the Middle East being spread through passenger plane was 66% in Asia, 21% in Europe, and 12% in Africa.

In the US, an EpiSim simulation system was developed to predict where an infectious disease will spread for up to 3 years using the data related to the pathogenic organism, infection path, and population movement through the transport network centered on the main point of infection. As a unique feature of the system, it processed the movement pattern data using transport means into a network form, divided the individual moving pattern or contact with others into 14 levels, and used 4,000 airports, more than 40,000 flights, and moving people as variables. The system was used to prevent the further spread of MERS when two MERS carriers were found in the US.

The UK and Singapore are conducting horizon scanning to prevent the spread of infectious disease by predicting the potential threat factors or opportunity factors using the data. Its purpose is to prepare for uncertainty and carry out systematic response by exploring the environment based on national-level data. In particular, the UK opened the Foresight Horizon Scanning Center in 2005 to establish a future strategy based on data, whereas Singapore is operating the Risk Assessment & Horizon Scanning Center to categorize the threat factors so that responsive measures can be established in the medicine and safety areas. In Korea, the National Health Service combined the citizens' health data generated by National Health Insurance Corporation and social media data to provide the risk trend and alarm data of main diseases in the country. It analyzes the monthly average fluctuation rate and frequency in SNS of frequent infectious diseases using the treatment data for 5 years and provides information on selected diseases such as cold and food poisoning.

2.3. Purpose and Significance

The purpose of this study is to design and build a system for blocking infectious diseases imported from other countries. As discovered in the case of MERS in 2015, the efficacy of preventing infectious disease can be realized only when all cases of staying in countries suspected to be affected - including the final origin point and the stopovers - are searched. In a pilot project implemented in 2015, KT confirmed the possibility of searching all countries where a person stayed using the user's roaming data; the company then built a system for notifying the individual and authority of the data so that suitable measures can be taken. Since the passenger list of the flight that arrives in Korea after departing an affected country is already provided to the authority by the airline, the added roaming data provided by the system will significantly increase the efficacy of the quarantine program (Figure 3).

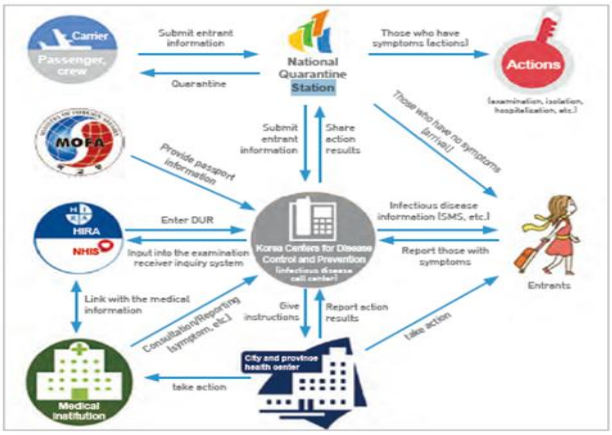

Fig. 3: The purpose of KT's Big Data Project

Since the country is surrounded by sea on three sides, and it faces North Korea in the north, most Koreans return to Korea through airplanes. An infectious disease that broke out in another country is imported by the person who visited the affected region, and it is more likely to have broken out in a country with poor sanitary conditions. The air routes are generally related to economic conditions; thus, there are few direct air routes between the countries affected by infectious disease and Korea. As such, since people staying in a specific country affected by infectious disease are highly likely to have stopped over a clean country, this system is considered very important for the quarantine program.

3. Big Data Implementation

The purpose of this study is to identify the people who have visited the contaminated country and provide proper quarantine service to the pertinent people until the latent period is over. KT cooperated with the Korea Centers for Disease Control & Prevention (KCDC) so that its Big Data system can send the list of visitors to the countries affected by infectious disease upon request by KCDC, which then sends a text message to inform the potentially infected person and also transmits the information to the DUR (Drug Utilization Review) system of the Health Insurance Review & Assessment Service (Figure 4).

In the first-step data collection, the roaming registration request signal generated when a KT subscriber begins roaming from a foreign country is received from the roaming system and collected.

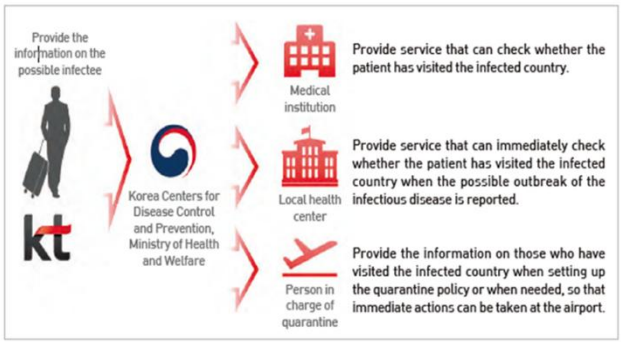

Fig. 4: Conceptual Diagram of KT's Big Data System for Quarantine Support

Table 1: Process of Using Big Data

Step	Details
Data Collection	All data on roaming users and countries affected by infectious disease are collected.
Data Storage	The data on countries affected by infectious disease and people who visited those countries are stored.
Data Analysis	The subjects to be managed are selected according to the data processing rule.
Use of Analysis Result	The information on infectious disease is sent to the selected subjects and to KCDC for policy utilization.

The roaming registration signal and the signal indicating return to Korea are collected from KT's roaming system, and the subscriber information is collected with an API function interfaced to KT's subscriber information system. When a KCDC official in charge of quarantine inputs a country affected by infectious disease, KT manages the country in its system. In the second-step data storage, the data of subscribers who visited the country affected by infectious disease are received from KT's roaming system and stored. The disease data are also collected from KCDC and stored. In the third-step analysis, potentially infected people are selected using the overseas location registration signal in the real-time roaming data as shown in Figure 5.

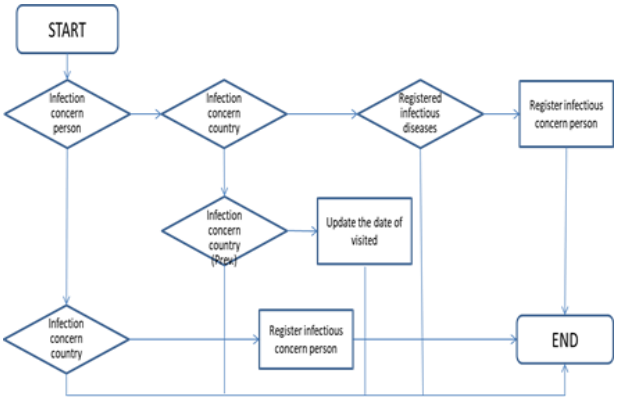

Fig. 5: Rule for Selecting Subjects for Infectious Disease Management

In the fourth-step utilization of analysis result, the information on an infectious disease is sent to those visiting the country affected by infectious disease and those who have returned after visiting the country. The visiting history of the country affected by infectious disease is provided to the medical institute when a selected subject visits the medical institute within the latency period.
KCDC aggregates the records of visiting the affected countries and uses them for the quarantine policy. The changed process is described in Figure 6.

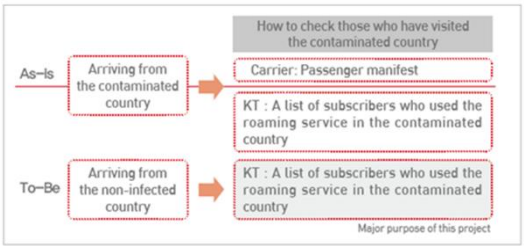

Fig. 6: Changed Process

4. Results

The Big Data system is used to survey the visitors of countries suspected to be affected by infectious disease. During the testing period, approximately 5,000 people who visited or returned from countries affected by infectious disease were confirmed each day. In particular, around 200 people who visited the affected countries and returned through stopover in a clean country were found, meeting the objective of the pilot project; quarantine service was then additionally provided to them.

Table 2: Suspected Patients of Infection Disease Checked through Roaming Data

Type	SMS Sent by		People who returned after a stopover in a clean country	
Description	People who sent SMS upon arriving in a foreign country	People who sent SMS upon arriving in Korea	MERS	Zika Virus
Nov. 22	5,045	3,458	8	144
Nov. 23	4,623	3,363	19	178
Nov. 24	5,129	3,336	16	188
Nov. 25	5,011	4,312	11	209
Nov. 26	4,489	4,326	24	255
Nov. 27	4,636	4,927	29	293

5. Conclusion

This study described KCDC's quarantine management system using Big Data generated by KT, which collected the roaming data, selected the visitors of countries affected by infectious disease, and developed the infrastructure to interface with KCDC and send an SMS message while KCDC sent an SMS message informing the visitors of affected countries of the infectious disease. The Big Data system of KCDC can query the existing DB of visitors of countries affected by infectious disease and the visitor data sent by KT together and send the visitor data to the Health Insurance Review & Assessment Service so that the hospitals can query the information through DUR. The summary of project is described in Figure 6.



Fig. 6: The Summary of Big Data Project

Although the Big Data are widely used in private and public sectors, there are many limitations in using Big Data related to personal information because of the Personal Information Protection Act. Nonetheless, this project showed that Big Data containing some personal information can be used for the public interest of protecting constitutional rights such as right to life and right to property. Increasing the utility of Big Data for public interest like the late night bus system by Seoul Metropolitan City greatly helps increase awareness of Big Data. This pilot test is significant since useful cases of Big Data can be created through social consensus for using personal information for public and private interests. This system is expected to contribute to national development by preventing the recurrence of MERS crisis, which caused economic loss of around KRW 6 trillion, and avoiding loss caused by the import of infectious disease.

The cooperation between KT and KCDC extended to LGU+ and SKT, with all domestic telecommunication operators providing the roaming data in the same form as of April 2017. Moreover, KCDC and KT are discussing with foreign telecommunication operators to provide the same quarantine service to foreigners visiting Korea,

and they have proposed international cooperation among the telecommunication operators worldwide to the UN.

Although the system implemented in this study is a simple one that merely provides the alarm function to subjects with relatively minor Big Data analysis, it is significant as the pioneering application of Big Data for public interest. If the system can be grown into the national risk prediction and monitoring of infectious disease system and expanded globally, it can greatly help improve the health level of Korea as well as the whole world.

There would be two major expected effects in this study. Eliminating the possibility of losing the initial golden time because the infectious disease is misjudged as another disease, by securing information on those who visited a contaminated country until the latent period is over and providing the information to the quarantine authorities and medical institution. Also, changes in people's reaction to symptom development are expected by providing the information related to the infectious disease in a timely manner to those who visited a contaminated country.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP). (No.2018-0-00705, Algorithm design and software modeling for judge fake news based on artificial intelligence)

This work was Supported by a Korea University Grant in 2015. (1.Mar.2015-28.Feb.2016)

The authors would like to thank project members and agency officers for providing project materials and interviews. This paper was supported by National Information Society Agency, Ministry of Science, ICT and Future Planning, Korea Telecom, and Korea Centers for Disease Control & Prevention.

This paper was supported by Wonkwang University in 2018.

References

[1] Laney D. 2001. 3D Data Management: Controlling Data Volume, Velocity and Variety, Gartner.
[2] Gandomi, A. and Haider, M. 2015. Beyond the Hype: Big Data Concepts, Methods, and Analytics. International Journal of Information Management. 35, 2 (2015), 137-144. DOI=http://dx.doi.org/10.1016/j.ijinfomgt.2014.10.007
[3] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A. H. 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. (2011) McKinsey Global Institute
[4] NRI, "The Advent of Big Data Era", Nomura Research Institute, 2012.
[5] MSIP, "Big Data Industry Development Strategy", The Republic of Korea Government, 2013.
[6] Chen, H., Chiang, R. H. L. and Storey, V. C. 2012. Business Intelligence and Analytics: From Big Data to Big Impact. MIS Q. 36, 4 (2012), 1165-1188.
[7] Syed, A., Gillela, K. and Venugopal, C. 2013. The Future Revolution on Big Data. International Journal of Advanced Research in Computer and Communication Engineering. 2, 6 (2013), 2446-2451.
[8] Lee, S. W. and Kim, S. H. 2016. Finding Industries for Big Data Usage on the Basis of Ahp. Journal of Digital Convergence. 14, 7 (2016), 21-27.
[9] Shin, K. S., Chai, S. M., Park, H. J., Jo, N. O., Shin, S. A. and Kim, S. H. 2016. Development of a Big Data Capability Assessment Model. Journal of Information Technology and Architecture. 13, 2 (2016), 271-280.
[10] Kim, S. H., Park, S. B. and Lee, Y. G. 2015. A Development of a Evaluation Framework for Public Sector Ict Adoption: Focused on Big Data, Cloud, Internet of Things. Journal of Information Technology and Architecture. 12, 3 (2015), 419-428.
[11] N. Ashish, V. Dan, "Worldwide Big Data Technology and Services Forecast 2015-2019", IDC, 2015.
[12] NIA, 2016 Korean Big Data Market Review, National Information Society Agency, 2017.
[13] Braden, C. R., Dowell, S. F., Jernigan, D. B. and Hughes, J. M. 2013. Progress in Global Surveillance and Response Capacity 10 Years after Severe Acute Respiratory Syndrome. Emerging infectious diseases. 19, 6 (2013), 864.
[14] Riley, S., Fraser, C., Donnelly, C. A., Ghani, A. C., Abu-Raddad, L. J., Hedley, A. J., Leung, G. M., Ho, L.-M., Lam, T.-H. and Thach, T. Q. 2003. Transmission Dynamics of the Etiological Agent of Sars in Hong Kong: Impact of Public Health Interventions. Science. 300, 5627 (2003), 1961-1966.
[15] Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N. F., Soropogui, B., Sow, M. S., Keïta, S. and De Clerck, H. 2014. Emergence of Zaire Ebola Virus Disease in Guinea. New England Journal of Medicine. 371, 15 (2014), 1418-1425.
[16] WHO, 2015a. Middle East respiratory syndrome coronavirus (MERS-CoV)-Republic of Korea. Disease Outbreak News. 24 May 2015.
[17] Cowling, B. J., Park, M., Fang, V. J., Wu, P., Leung, G. M. and Wu, J. T. 2015. Preliminary Epidemiologic Assessment of Mers-Cov Outbreak in South Korea, May–June 2015. Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin. 20, 25 (2015)