

A Design of Application through Physical Therapy Big Data Analytics

Woo-Hyeok Choi¹, Jun-Ho Huh^{2,*}

Abstract

According to the National Health Insurance Corporation in 2008, there were 17,764,428 physical therapy patients, exceeding 31 percent for the population covered by health insurance. This means that three out of 10 Koreans received physical therapy. And now, 10 years later, due to the aging population and the increase in the sports population, the number of patients with physical therapy is expected to be much more than a decade ago. Among them, many physical therapy patients were orthopedic and neurologic disorder. However, in the medical field applied to physical therapy, it is widely applied across all medical fields, including orthopedics, neurosurgery, pediatrics, gynecology, thoracic surgery and dentistry. It is believed that various cases of patients receiving physical therapy will be secured. as mentioned earlier, there will be a large number of patients with physical therapy treatments, making big data analytics easier. based on this, physical therapy applications are thought to be helpful in the analogy of disease and the development of effective physical therapy and will ultimately promote the development of physical therapy.

Key Words: Physical Therapy, Aging Population, Big Data, Application

I. INTRODUCTION

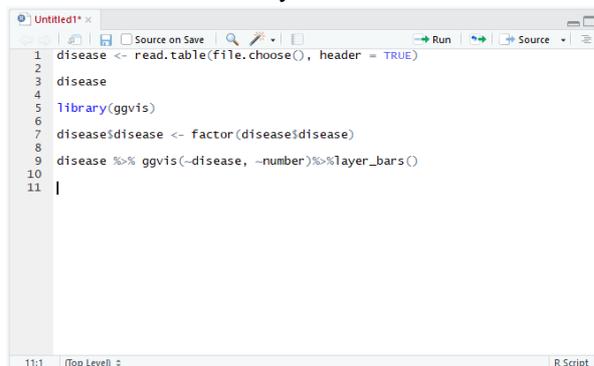
Nowadays, there was significant development in the field of Intelligent Big Data (IBD) analysis where a multicore platform based on a large computing cluster was used. Despite the improvement, too much complex information is still being provided for a single institute or a computing center for processing. Especially, the number of multimedia and user population will increase continuously and exponentially due to the rapid spread of smartphones and social networking sites [1].

According to World confederation for Physical Therapy (WCPT), Physical therapy is about relieving damage and functional limitations with planned, helpful, and adjustable therapeutic interventions. It also includes the stamina, health and quality of life for people of all ages. like this, physical therapy is a treatment that can be treated regardless of age or sex for various reasons that are restricting the body. The coverage of physical therapy is also very diverse. Physical therapy can be applied in a variety of fields such as orthopedics, neurosurgery, pediatrics, gynecology, thoracic surgery and dentistry. in particular, physical therapy by orthopedic accounts for the largest portion.(2008) among the health insurance benefits claims

for the 22 major diseases, the disorders of the digestive system (23.40 million cases) and the ailments associated with skin & subcutaneous tissue (over 12.00million+ cases) were the highest.

Besides these two, the incidents of diseases pertaining to the musculoskeletal, nerve, and circulatory systems, as well as the congenital deformities and the genito-urinary diseases are continuously occurring [Fig. 1-2].

In Fig. 2, No. 10 and 11 are of the disorders of the digestive system and the ailments associated with skin & subcutaneous tissue. It also shows that many other diseases have occurred additionally.



```
1 disease <- read.table(file.choose(), header = TRUE)
2
3 disease
4
5 library(ggvis)
6
7 disease$disease <- factor(disease$disease)
8
9 disease %>% ggvis(~disease, ~number)%>%layer_bars()
10
11 |
```

Fig. 1. R-Studio Analysis of insurance payments by 22 categories.

Manuscript received: September 17 2018; Accepted: September 26 2018. (ID No. JMIS-2018-0045)

Corresponding Author (*): Jun-Ho Huh, Catholic University of Pusan, Republic of Korea. +81-463-913130, 72networks@pukyong.ac.kr or 72networks@cup.ac.kr

¹Department of Physical therapy, Catholic University of Pusan, Busan 46252, Republic of Korea.

² Department of Software, Catholic University of Pusan, Busan 46252, Republic of Korea.

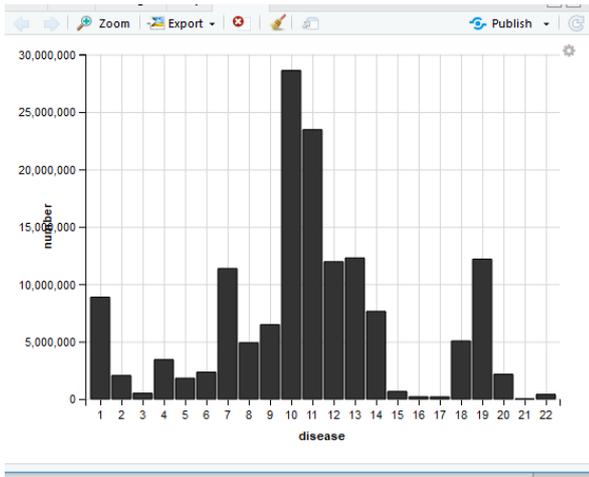


Fig. 2. A histogram of insurance payments by 22 categories.

Table. 1. is a public Big Data of the health care system including the health insurance sample cohorts collected by the National Health Insurance Corp., the patient dataset managed by the Health Insurance Review & Assessment Service, and the public health & nutrition survey conducted by the Korea Centers for Disease Control and Prevention. Big data is being utilized by such healthcare organizations so that it is expected that enough big data can be collected in the physical therapy field.

Table. 1. The public data of the health care system

Category	Agency in Possession	Contents	Opened/ Closed
Health Insurance Sample Cohort DB ¹	National Health Insurance Corp.	- Qualification DB: gender, age groups, districts, socio-economic variables, disorders, and deaths information of health insurance subscribers and recipients.	Limitedly opened
Patient Dataset ²	National Health Insurance Review & Assessment service	- The sample data of the patients treated for a period of one year from the date of starting the treatment based on the health insurance claims.	Limitedly opened
Korean Human Body Resources	Korea Centers for Disease Control & Prevention	-The information concerning the human-derived materials (DNA, tissue, blood, urine, etc.) & clinical data (therapy type surgery type, in addition to pathological test result & blood test result, etc.) and epidemiologic (gender, birth date, drinking/smoking history, etc.) and genetic (SNP, CNV, Exome, etc.) data	Limitedly opened

Community Healthcare Information	Social Security Service	-Information of health services (health/medical centers/branches & their administrative works. Obligatory electronic documents & treatment data.	Closed
Community Health Survey	Korea Centers for Disease Control & Prevention	-Investigation of community health conditions as a basis for establishment of a community healthcare plan & health service assessment index (health check-up, vaccination, disease contraction, use of medical systems, accidents & addictions, limited mobility, quality of life, use of healthcare systems, socio-physical environments, cardiac arrest, education & economic activities, etc.)	Opened
Public Health/Nutrition Survey	Korea Centers for Disease Control & Prevention	-Investigation present condition & trend of public health & nutritional balances. -Surveys on body measurements, obesity, high blood pressure, smoking/drinking, weight control in addition to nutritional data related to food & nutrition intake, dietary habits, & food supplements, etc.	Opened
Korea Health Panel	Korea Institute for Health & Social Affairs	-Personal health levels, factors associated with use of medical systems & medical expenditures, health activities, medical requirements, analysis of changes in the medical service demands. -Socio-economic characteristics, purchase of medicines, economic activities, health levels, medicine-intake forms, private medical insurances, health functional foods, health activities, etc.	Opened

On the basis of the health insurance subscribers of the Republic of Korea (ROK), the number of patients requiring physical therapy has exceeded 17 million [Fig. 3, 4 & 5] and it is estimated that the number has increased recently. On this basis, an attempt was made to engraft the concept of big data into such a content.

```

1 totalpatients <- read.table(file.choose(), header = TRUE)
2
3 totalpatients
4
5 library(ggvis)
6
7 totalpatients$year <- factor(totalpatients$year)
8
9 totalpatients %>% ggvis(~year, ~person) %>% layer_bars()
10
    
```

Fig. 3. A R-studio analysis of the patients who have received physical therapy (2005-2008).

	A	B
1	year	person
2	2005	14964575
3	2006	17203932
4	2007	17811875
5	2008	17764428

Fig. 4. The number of patients who have received physical therapy.

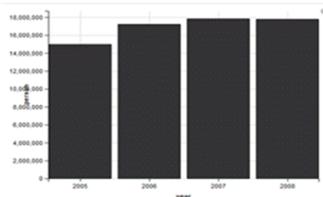


Fig. 5. The histogram of an R-studio analysis for the patients.

II. RELATED WORK

Accordingly, the market for the big data is becoming larger over time and the data is being used in different areas of our daily lives and much information is shared by the general population. However, since the analysis of big data is very complicated and difficult that sometimes it is quite hard to recognize its meaning and direction, the visualization of big data has come into the picture. Recently, the big data analysis is shifting from AMOS to R/TensorFlow [2].

Machine learning (ML) refers to studying various methods of achieving human-like learning ability through machines, and from the data analysis results, the program can learn about rules or new knowledge by extracting them

automatically by itself. The techniques related to machine-learning remaining at the basic level is now becoming more sophisticated due to the emergence of new data mining techniques which can maximize their potential. Recently, ML is one of the major areas of interest for the artificial intelligence systems, being at the intersection of informatics and statistics and closely related to the data science and knowledge discovery as well as the healthcare industry [3-4]. Especially, probabilistic ML is quite useful for the health informatics where most of the problem-solving process involves removing of uncertainties. The theoretical basis of the probabilistic ML was initially laid by Thomas Bayes (1701-1761) [5]. The probabilistic inference holds a key position in artificial intelligence and statistical learning where the inverse probability allows one to infer unknown facts, deducing them from the available data and making predictions [6-7].

Meanwhile, the scale of big data is much larger than that of the data generated from the analog environment of the past, shorter in generation cycles, and not only the numerical data but the character and image data are included in the big data as well. Since the use of PC, internet, or mobile devices has become part of people's daily routine, the volume of data left behind by them is increasing rapidly. Along with the fact that the volume of big data has increased explosively, the types of data have been also diversified such that people's behaviors, as well as their thoughts and opinions can be anticipated through positional information and SNS services. Many countries and companies are attempting to construct and utilize the big data system now. Accordingly, the market for the big data is becoming larger over time and the data is being used in different areas of our daily lives and much information is shared by the general population. However, since the analysis of big data is very complicated and difficult that sometimes it is quite hard to recognize its meaning and direction, the visualization of big data has come into the picture [8]. Wu et al. [9] have argued that a large volume of data (big data) can be problematic when frequent itemset mining has been used for the following reasons: (1) spatial complexity: the algorithm may not be run as the system memory deal with a large input data as well as large intermediate results and output pattern; (2) time complexity: many existing approaches depend on an exhaustive search or a complicated data structure to obtain a frequent pattern but this is not suitable for big data. Thus, they proposed an iterative sampling-based frequent itemset mining that samples the subsets instead of processing entire dataset all together and then extracting the frequent itemsets from them. Yang Luo et al. [10] maintained that segmenting the Left Ventricle (LV) from the cardiac MRI image is essential when computing the clinical indices such as stroke volume,

ejection fraction, etc. Thus, in this study, an automated LV segmenting method where the hierarchical extreme learning machine (H-ELM) is combined with a new location recognition method is proposed [1]. Big Data is not a new concept but refers to a massive data which exceeds the existing range of data storage, management, or analysis system can deal with, such as Relational Database Management System; RDBMS), etc. (J. H. Lee et al. 2014) [11-12]. Also, S. W. Cha (2014) noted that the technological characteristics of big data are high volume, high variety, and high velocity in terms of volume, variety, and velocity, respectively [13].

As such, it can be considered that big data is highly correlated with physical therapy on the basis of a large number of patients (high volume), a wide range of application (high variety), and the rapid increase in the number of patients (high velocity).

The proportion of the ROK's senior citizens over 65 reached 7.2% in 2000, which was then increased to 13.5% in 2016, approaching 14% satisfying the UN's definition of 'aging society' (H. J. Kang & G. H. Lee, 2017) [14]. Although the R-studio-based statistics did not indicate a large increase in the Korean population, it is evident from the fact that the rapid increase in the number of senior citizens is the evidence of becoming an aging society.

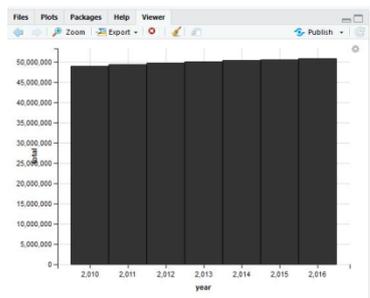


Fig.6. The total population of the ROK

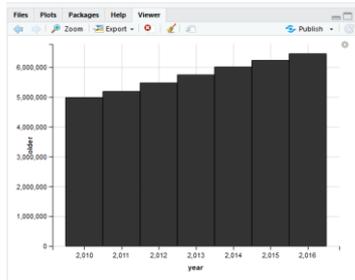


Fig.7. The population of senior citizens of the ROK.

For these reasons, the aging population will be at the risk of being exposed to a variety of diseases as they get older and the number of their visits to a hospital will be increased. The only currently available significant statistics of physical therapy-related data is one that has been compiled by the National Health Insurance Corp. from 2005 to 2008.

The same type of statistics has not been compiled for over 10 years now so that it would negatively affect the development of physical therapy. Thus, if an application based on the physical therapy big data is developed, the statistics will be taken actively and ultimately, the therapy itself will be improved in efficiency and quality.

III. Design of Physical Therapy-Based Application

3.1. Agreement on Sharing Physical Therapy Big Data

Observing the result of classification by risk types using a big data, the biggest risks are legal or institutional risks which occur frequently and have a specific issue of breaching one's privacy (S. O. Yoon, 2013) [15]. Additionally, according to Article 2 of the ROK's Privacy Act, the personal information is being defined as the information of a person which can be recognized through his/her name, resident registration number or images, including the information which by itself cannot be used to identify a person but can be combined easily with other information for identification. In this regard, the physical therapy big data may breach one's privacy or expose the details of the therapy and thereby resulting in a breach of the Privacy Act. Therefore, the most important part when developing an application based on the physical therapy big data is obtaining a patient's consent on sharing his/her personal information. Also, following Item 4 of Article 4 (Rights of Data Subject), the application will allow the patients to demand their information to be stopped from processing or revise/delete/destroy it afterward even if they have given their consent so that the users of the application will be free from the concerns of personal information leak.

The method of obtaining a consent from the user is by asking it on the first screen when starting to download the application.

The name of this application is AOPT which stands for All of Physical Therapy and the initial screen is shown in Fig 1 where the user's consent for sharing his/her personal information will be being asked. After obtaining the consent, the starting screen will collect the personal information including the types of diseases, gender, and age which are necessary for creating a physical therapy big data.

The main screen will be composed of Physical Therapy Report, Complaint-Type Search, Recommended Workout, Hospital Search, Talk with Physical Therapist, Community, and My Page.

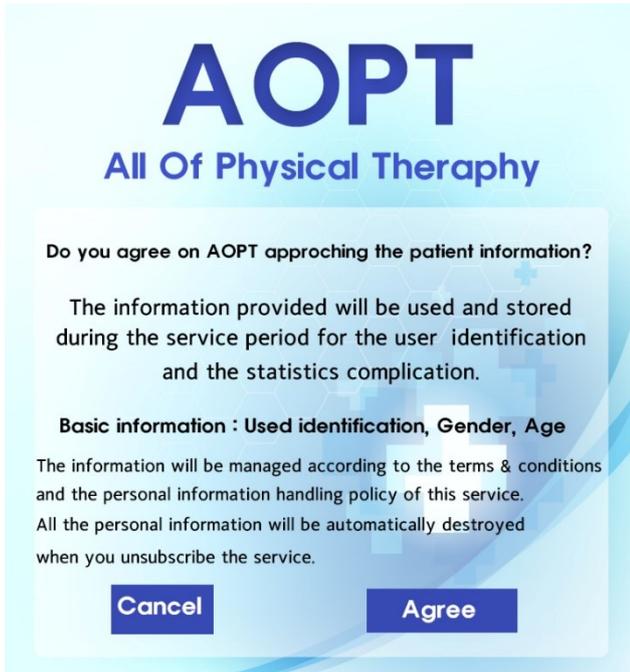


Fig. 8. Obtaining an agreement on sharing personal information on the initial screen.

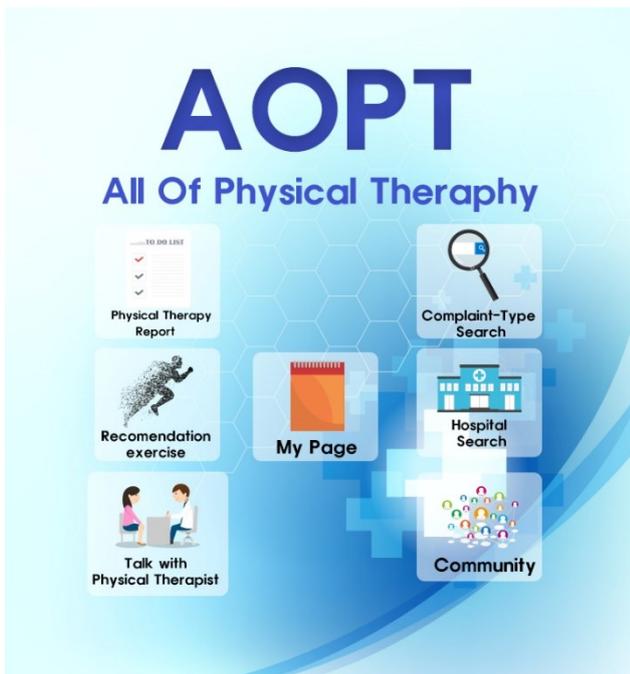


Fig. 9. The Composition of the Main Screen.

Fig. 10. is showing Physical Therapy Report where the total duration of the patient’s physical therapy, number of treatments, and the parts of the body being treated during his/her entire life will be statistically compiled and used for the generation of a physical therapy big data.

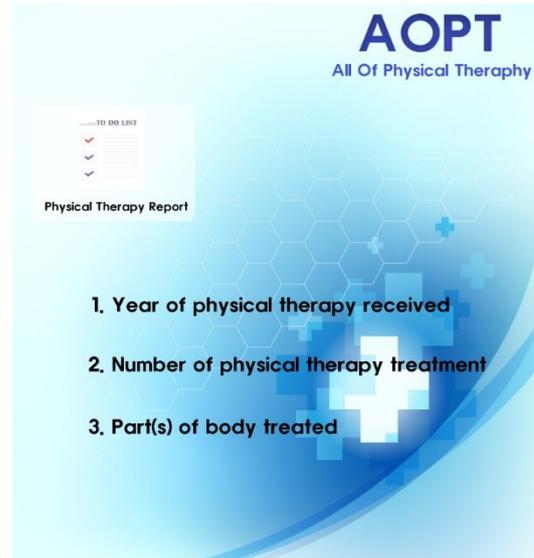


Fig.10. Physical Therapy Report.

Fig. 11 is showing Complaint-type Search with which a patient can search for his/her complaint about a certain symptom other than the main area(s) of treatment. In this step, a big data entered by the user will be used to infer the disease based on the symptom.

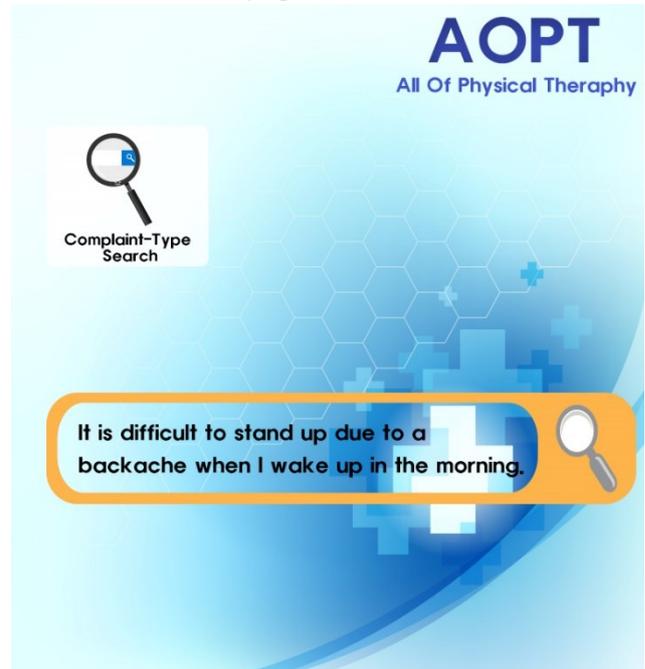


Fig.11. Complaint-Type Search.

Meanwhile, Fig. 12 shows Recommended Workout which recommends the self-workouts that can be performed at home after receiving a treatment according to the symptom.



Fig.12. Complaint-Type Search

Talk with Therapist in Fig.13. will be able to recommend the hospitals based on their specific physical therapy specialties considering patient’s present condition.



Fig.13. Hospital Search

Talk with Therapist in Fig. 14 will allow the patient to communicate with his/her therapist(s) directly about the treatment he/she has received. The talks with the therapists the who are involved with the treatment will be systematic and professional.

Finally, My page in Fig. 15 provides a space in which the present patient is required to enter or can modify his/her age, gender, and the type of physical therapy being applied to him/her. These data will be a basis for performing a big data analysis.

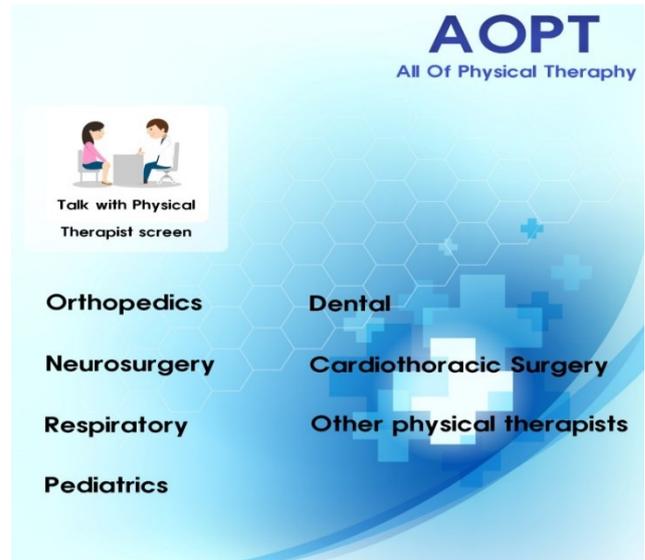


Fig.14. The Talk with Physical Therapist screen

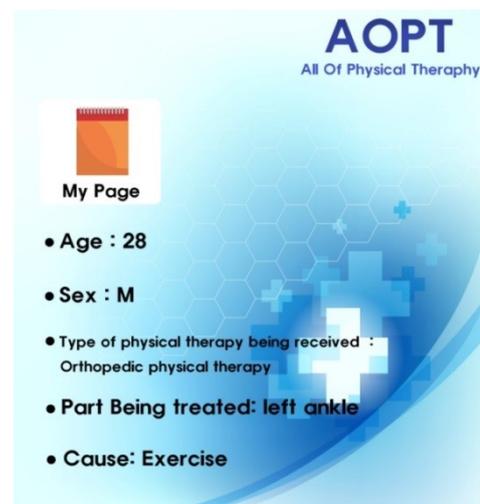


Fig.15. The Talk with Physical Therapist screen

IV. CONCLUSION

The application proposed in this study is highly expected to contribute to the development of physical therapy by assuming a big role in collecting the physical therapy-related big data. With this application, the patients will be able to receive physical therapy conveniently and acquire a variety of information whereas the physical therapists will be able to develop new types of treatment techniques with which they can conduct a prompt and systematic treatment by analyzing the Big Data.

REFERENCES

[1] Jun-Ho Huh,; “Big Data Analysis for Personalized Health Activities: Machine Learning Processing for

Automatic Keyword Extraction Approach,” Symmetry, MDPI, 2018, Vol.10, No.4, pp. 1-30.

- [2] Sangdo Lee,; Jun-Ho Huh,; “An effective security measures for nuclear power plant using big data analysis approach,” The Journal of Supercomputing, Springer US, pp.1-28, 2018.
- [3] Jordan, M.I.; Mitchell, T.M.; “Machine learning: Trends, perspectives, and prospects,” Science, 2015, 349, 255-260.
- [4] LeCun, Y.; Bengio, Y.; Hinton, G.; “Deep learning,” Nature, 2015, 521, 436-444.
- [5] Bayes, T.; “An essay towards solving a problem in the doctrine of chances,” Stud. Hist. Stat. Probab. 1970, 1, 134-153.
- [6] Hastie, T.; Tibshirani, R.; Friedman, J.; “The Elements of Statistical Learning: Data Mining, Inference and Prediction,” Springer Berlin, Germany, 2008.
- [7] Murphy, K.P.; “Machine Learning: A Probabilistic Perspective,” MIT Press: Cambridge, MA, USA, 2012.
- [8] Huh, J.H.; Kim, H.B.; Seo, K.; “A preliminary analysis model of big data for prevention of bioaccumulation of heavy metal-based pollutants: Focusing on the atmospheric data analyses,” Adv. Sci. Technol. Lett. SERSC, 2016, 129, 159-164.
- [9] Wu, X.; Fan, W.; Peng, J.; Zhang, K.; Yu, Y.; “Iterative sampling based frequent itemset mining for big data,” Int. J. Mach. Learn. Cybern. 2015, 6, 875-882.
- [10] Luo, Y.; Yang, B.; Xu, L.; Hao, L.; Liu, J.; Yao, Y.; Van de Vosse, F.; “Segmentation of the left ventricle in cardiac MRI using a hierarchical extreme learning machine model,” Int. J. Mach. Learn. Cybern. 2017.
- [11] Lee Yeon Hee.; “Use and Challenges of Public Big Data in the Field of Health and Welfare,” Health and Welfare Forum. 2015.
- [12] Lee Ji Hye et al.; “Big Data Utilization Trends in Healthcare Sector,” Journal of the Korean Institute of Communication Sciences, 32: 1, 2014, 63-75.
- [13] Breastfeeding,” “Protection of Big Data Environment and Privacy,” IT and Legal Research 8 (2014): 193-259.
- [14] Kang, Hyeong-jung, Lee, Kyung-hee,; “A Plan for Activating the Elderly Community in an Aging Society,” Journal of Korean Society for Living Environment 24, 3, 2017, 380-388.
- [15] Yoon Sang-Oh,; “A Study on the Classification of Risk Types of Big Data,” Journal of Korean Local Information Science 16: 2, 2013, 93-122.
- [16] Lee Hye Rim, Jo Jae Yeon, Yun Ji Won,; “Healthcare Security Issues from the Viewpoint of Cyber Physical System in Cloud Environment,” Journal of Information Security, 24.6, 2014, 7-13

Authors



Woo-Hyeok Choi was born in Gunwi-Gun, Gyeongsangbuk-do, Republic of Korea. Currently he is student at Department of Physical therapy of Catholic University of Pusan. His research interests are App, Algorithm, Optimization, High Availability Computing.



Jun-Ho Huh was born in Kyoto, Japan. He was finished the Cooperative Marine Science and Engineering Program, Texas A&M University at Galveston, United States of America in Aug. 2006. Received B.S. in Science Degree from Department of Major of Applied Marine Sciences (Currently Faculty of Marine Biomedical Sciences), And, B.S. in Engineering Degree (Double Major) from Department of Major of Computer Engineering from Jeju National University at Ara, Jeju, Republic of Korea in Aug. 2007. Also, completion of the Secondary School (Middle and High schools) Teacher Training Curriculum in accordance with Republic of Korea Secondary Education Act (Aug. 2007). Received M.A. in Education Degree from Department of Major of Computer Science Education, Graduate School of Education, Pukyong National University at Daeyeon, Busan, Republic of Korea in Aug. 2012. And completion of the Secondary School (Middle and High schools) Teacher Training Curriculum in accordance with Republic of Korea Secondary Education Act (Aug. 2012). Received the Ph.D. in Engineering Degree from Department of Major of Computer Engineering, Graduate School, Pukyong National University at Daeyeon, Busan, Republic of Korea in Feb. 2016. He received the Best Paper Award from Korea Multimedia Society nine times (Nov. 2014, May. 2015, Nov. 2015, May, 2016, Oct, 2016, May, 2017; three times, Oct, 2017). And he received Undergraduate Student Paper Bronze Medal (Corresponding Author) Awarded, Korea Information Processing Society (Apr. 28, 2017). Also, he received Best Paper Award The 10th 2016 International Interdisciplinary Workshop Series from HSST (Aug. 2016). Also, he received Best Paper Award The 16th International Conference on Control, Automation and Systems (Oct. 2016), ICROS with IEEE Xplore. He is Technical Committee at IFAC (International Federation of Automatic Control), CC 1 (Systems and Signals), TC 1.5. (Networked Systems) (2017-2020). And, Technical Committee at IFAC, CC 3 (Computers, Cognition and Communication), TC 3.2 (Computational Intelligence in Control) (2017-2020). Technical Committee at IFAC, CC 7 (Transportation and Vehicle Systems), TC 7.2. (Marine Systems) (2017-2020). Research Professor of Dankook University at Jukjeon, Yongin, Republic of Korea (July. 2016- Sep. 2016). Currently he is Assistant Professor of Department of Software, Catholic University of Pusan, Republic of Korea (Dec. 2016-). His research interests are Green IT, Smart Grid, Network Security, IoT, System Architecture.

