International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 13, Number 12 (2020), pp. 4177-4182
© International Research Publication House. http://www.irphouse.com

System Design of Big data based regional information service

Changbae Mun

[1] Professor, Dept. of Electrical, Electronic & Communication Engineering, Hanyang Cyber University, Seoul, Korea.
[1] ORCID: 0000-0002-3065-3307

Abstract

Along with the development of distributed processing systems, the LBS system using big data is continuously developing. The most frequently used LBS is navigation. The scope of LBS service has gradually expanded, such as API service for map information and ship route information. With the development of the improved system, the information to be dealt with is expanding to SNS and blog data for each user's location. In particular, in recent years, it has been used in various industries such as location-based marketing, mobile advertising of automobiles, Internet of Things (IoT) and Online to Offline (O2O) services. In this study, a method to analyze more accurate location information using big data was analyzed. In particular, a methodology was introduced to further increase accuracy by analyzing web-based information in real time and supplementing the data. Through the system proposed in this study, it is possible to construct the architecture of a new LBS service model integrated with various web services.

Keywords: Location based Service, LBS, Apache Spark, Big Data, Recommendation Service

I. INTRODUCTION

According to the LBS report of the Korea Internet & Security Agency, the share of LBS services in the current mobile industry has been increasing rapidly since 2012. The Korea Internet & Security Agency provides information on LBS services and contents to suggest the scope of the LBS industry. It was divided into service and information utilization service. It presents the scope of the LBS industry [1] According to Technavio's report on LBS business, the international LBS industry is continuously developing and is applied to new industry domains, especially based on IOT. In particular, the location-based marketing industry has a large proportion, and IoT and local advertising have been suggested as new domains [2]. In terms of public safety services, the technology of inquiring location information is also gradually developing in situations where a structure for users is required as a social safety net. LBS business based on smartphone application has developed. In the model of the mobile system that was initially introduced, the location information API was developed as a service that analyzes the location information of the user and provides various information to the user's application. As such, LBS is mainly used for location information in buildings and navigation of cars, and then gradually expands its scope, and is used for location services applying advertisements, location-based marketing, AR games, and drones. As described above, the LBS includes a content service that provides a user using the wireless Internet with various information related to a location that changes according to the user's movement. As described above, cases of utilizing user's location information are developing, but it is necessary to develop a technology that combines and normalizes the currently presented location information and reconstructs it into more accurate information [2]. This data categorizes security, map surrounding information, and local life information by utilizing the user's location information. Through this normalized information, the location information has expandability. In this study, we propose a new LBS system through normalized route information. This system continuously analyzes map API, SNS, and Internet search information.[4-5] Through this, data pre-processing is performed through a batch job of complex information on the user's area. Also, based on refined information, we present a method that can provide information tailored to the user's needs and a data processing technique that can integrate data. This paper deals with information scalability and refinement methods using big data analysis, and verifies the effectiveness according to the results of the experiment.

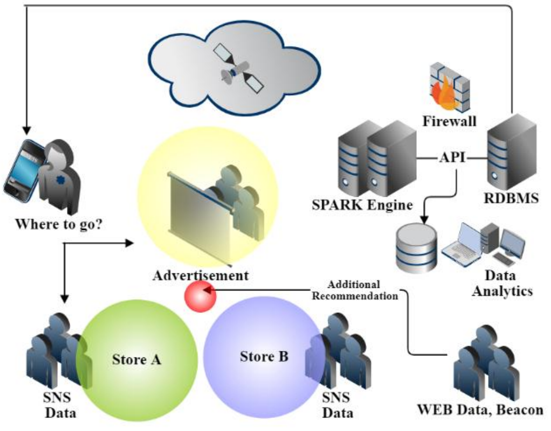


Figure 1. Overall system structure [3]

II. TECHNICAL REVIEW

2.1 Main function based LBS system

It is expected that the LBS industry will be active globally and will stand at the center of the development of the convergence industry that will arrive in the era of the 4th industrial revolution. The Korea Internet & Security Agency announced plans to revitalize the location information industry. In this information, a new LBS business area is prepared through improvement of laws and systems by establishing a plan until 2020. In addition, it is preparing a legacy that supports SMEs possessing various technologies in the location information industry. Through such institutional development, various technologies from the implementation of basic functions of the LBS service to new functions are supported. This system is the basis for the development of a new LBS Startup company. The LBS industry was used for military purposes in the early days of technological development. Initially, its functional effectiveness was proven, and now it is used in various systems (various public services such as transportation, logistics, location check, information search, advertisement, entertainment), and is being developed as a variety of application service applications. Since 2015, it has been combined with smartphones and tablets, and various services have been supported by opening a support platform along with hardware technical support. In terms of business, at the beginning of the introduction of smartphones, there was a continuous exploration of profit models other than map information API. Later, after 2011, map information applications were gradually developed, and new services appeared in various fields of map surrounding information, safety services, and entertainment. Accordingly, the LBS industry of the convergent technology domain is expanding. As such, the market for location information business has great potential for development, and various technologies are being tried. In the recommendation service, a model that provides customized location information based on information such as location, time, and age based on artificial intelligence has emerged. Carrot Market, a platform that provides local communities, collects local location information. Once the user's location is authenticated, they can trade objects and information with users in the nearby area. Especially, as in this application, location information is fused with other information to create new information in a combined form. You can see that it is evolving. In addition, in terms of personal information security, the importance of personal information security is increasing in the development of GPS-based applications, and technical alternatives are also provided.

2.2 Big data location information system

Through a big data distributed processing system such as Hadoop, I/O distribution of disk and load distribution for network input and output can be performed.Stable expansion of storage and systemic applicability by performing relocation of data nodes through data rebalancing This has the effect of increasing. In particular, Hadoop is an engine that can analyze large amounts of data at high speed, and is a platform for processing and analyzing big data. For data analysis, the system supports the use of fewer resources required for analysis. It is used for image search of portal services, data pattern analysis, and marketing through user analysis. For efficient processing of big data, system developers use a distributed file system and MapReduce platform within the Hadoop ecosystem. The MapReduce platform is an engine that quickly analyzes stored files using the CPU and memory resources of distributed servers. In terms of performance, data continuity is provided based on the network architecture. In terms of metadata, data redundancy is verified by supporting analysis processing in the Namenode.

In particular, Hadoop MapReduce supports map tasks without overlapping each other in HDFS, and provides a parallel processing system without overlapping. Since data is processed, duplicated, and stored in block units, the data type is managed with a schematic of the same type of key value, thus maintaining data consistency and ensuring the best result in the MapReduce process. In addition, it minimizes the data loss problem when dealing with failures to the data node. It overcomes the performance of the early version acting as a storage server in distributed processing services. Recently, it is possible to apply the performance of the most efficient analysis process by applying MapReduce of information related to the optimized analysis processing. A study was conducted to analyze the optimal path by analyzing the movement of a specific path mover through such a big data distributed processing system [3]. This study combines and analyzes location information processing methodologies and introduces the results of inference to this LBS service study, and proposes a new use in terms of convergence.

III. SYSTEM DESIGN AND VERIFICATION

3.1 Overall functional design of the system

In terms of the entire commerce industry, the proportion of mobile payments is increasing. Particularly in e-commerce, the aspect of customer experience for a specific brand is becoming more important. This aspect becomes an important factor according to the characteristics of each product that values communication with consumers. For this, beacon marketing has been introduced and a method of delivering discount coupons or promotion information to beacons through a smartphone app is widely used. This enables promotion information and products to be directly promoted to customers near the store. Therefore, this process can contribute to sales, and customers can conveniently receive information in the vicinity. In addition, it will be possible to conduct marketing and promotions based on customer behavior characteristics by synthesizing information such as user's purchase information, local information, and weather based on the purchase data and access log of the Internet mall. For example, a data set is created by analyzing the patterns of the movement of buyers in the store. This data can be used for online menu setting or for marketing based on visit history. Through this, the user's information is analyzed and a method to use the information most appropriately for marketing is statistically selected. When the shopping mall confirms consent to the provision of location information in the smartphone application, it analyzes and collects the customer's location signal. It can be analyzed mainly by analyzing the stores visited by customers and the route of movement, which can be used as major analysis data for the future distribution industry. Therefore, it will be possible to reestablish a product plan and use it for marketing by analyzing the customer's movement and analyzing the customer's purchasing pattern..

Table 1. Data set and key elements

DATA Level	DATA SET	Included elements
Initial	Latitude and longitude	Web Service
Managed	User's Data input and request, System exception handling	XML SAM FILE
Defined	The trajectory of the movement of Users, System key property information	XML Web Service
Quantitatively Managed	Integrated information system for travel routes, Data consistency verification information	XML External API Web Service

3.2 Composition of Big Data System

Apache Hadoop system based on open source is an open source Java framework that supports distributed application programs running on a computer cluster capable of processing large amounts of data. In short, Apache Hadoop is a software framework that can store, process, and analyze big data. As a distributed data processing technology, it operates as a computer cluster that bundles a small capacity of several servers rather than a single large capacity server. Hadoop's HDFS is used to store very large data, and MapReduce is used to perform operations using the collected data. When several companies bring to a centralized system, it is said to be an ecosystem with a system in which various software is added to increase the usability. Apache Hadoop runs on multiple nodes and consists of a master and one or more slave nodes. Each node has a different process depending on its type as follows. The role of the master node is the file system and job tracker. The role of the slave node is to save files and execute tasks (task tracker). Think about when dealing with large files. If you divide the file into a size that is easy to read, then save it to each node and read it from multiple nodes at the same time based on the split size, you will be able to process it with high performance. The key to HDFS handling large amounts of data is the distributed storage of large files.

MapReduce processes data in the saved file according to the MapReduce mechanism. It reads the input file and processes it in the map in key and value format. At this time, reduce collects and processes values with the same key. MapReduce passes data in this way between map and reduce, and you get the results you want. Apache Hadoop works by dividing the input file by the unit size (split), executing the map on multiple devices, and collecting the result back in reduce and processing it. This Apache Hadoop mechanism can be used to implement major frequency data extraction tasks on a large basis. Hadoop's distributed processing system provides data recovery and component recovery. Data recoverability means that even if a component of the system is down, system work is normally performed. Second, component recovery refers to a function that allows the system to run normally when a component of the system fails and is restored again.

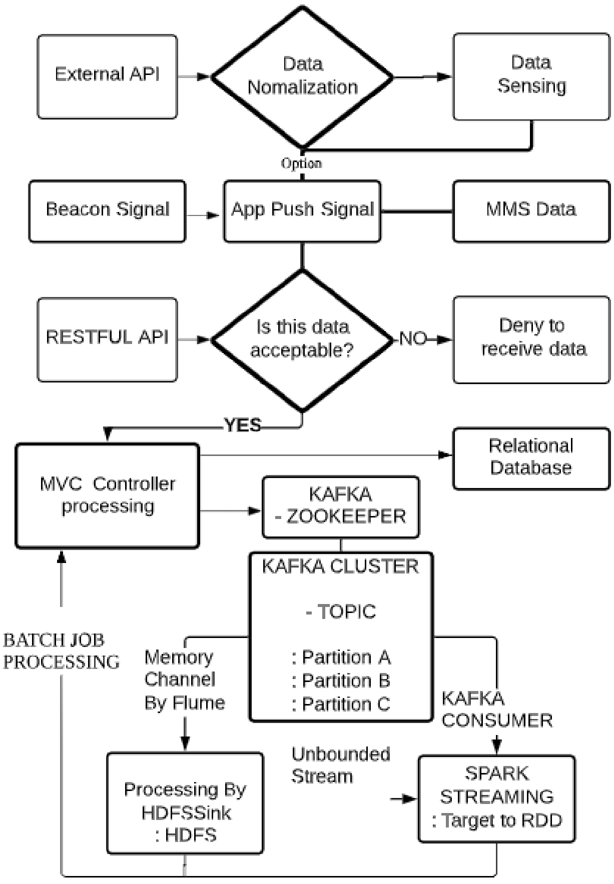


Figure 2. System process configuration [3]

3.3 Big data system processing logic

The schema of the whole system consists of data collection, analysis and user guidance. Managers are represented by experts who are familiar with the route between a specific origin and destination. The central server periodically analyzes the user's location and determines a branch spot using the provided information. The processing server will store all movement information from start to destination. Stores the user's store visit information provided while moving in a category unit. Then, the store categories visited by the user are subdivided and used in a collaborative filtering algorithm. This streaming of user movement information is configured through a distributed processing system. In order to provide API server to users who will actually use this system, Restful API server implemented with Spring framework is configured. This API server interlocks with a relational database to collect information on all movements. In particular, it was designed in such a way that the central server judges based on the legacy of the data while individual processing is taking place, separates the processing information from the storage and stores, and processes them in a distributed server.

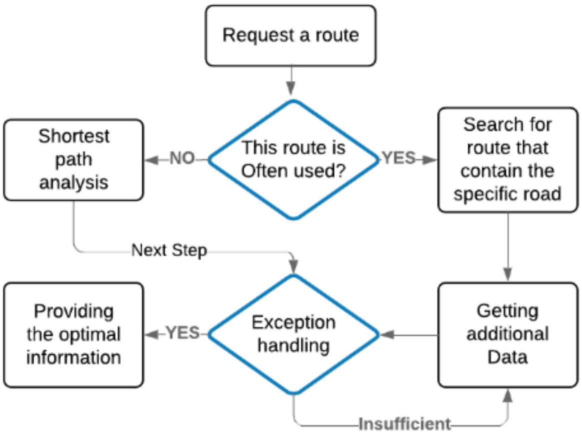


Figure 3. API's process schema [3]

In the schema of the entire system, Spark is an engine that runs at high speeds for large-scale data processing. Spark provides over 80 advanced operators to easily create parallel applications and can be used interactively in Python, R, etc. Hive, a Hadoop-based data solution, is an open source developed by Facebook that helps you analyze data without knowing Java. It provides a language similar to SQL called HiveQL to help you analyze data easily. Flume is a distributed service for collecting, aggregating, and moving large amounts of data. Flume consists of collectors that have agents installed on distributed servers like Chukwa and receive data from agents. The difference is that there is a master server that manages the entire flow of data, so you can dynamically change where data is collected, how it is transmitted, and where it is stored. Scribe is a data collection platform developed by Facebook, and unlike Chukwa, it transmits data to a central server, and the final data can be used as a variety of storage. Kafka is a distributed system to manage data streams in real time, and Kafka is developed for large-scale event processing. Hadoop is described as a distributed programming framework, but the Hadoop ecosystem is a group of various sub-projects that make up the framework. In other words, frameworks related to Hadoop are called Hadoop ecosystem. It consists of a Hadoop core project (HDFS, MapReduce) and a Hadoop sub-project (collection, analysis, mining, etc.). As shown in Fig. 2, in the first process of the branch, KAFKA exposes information with Spark Stream and Flume, and performs map reduction in Spark Stream. In the next step, the system uses Flume to log into HDFS and deliver the analyzed information to the user. Important data in this process has a separate backup data. Through this series of processes, the entire user data is analyzed.

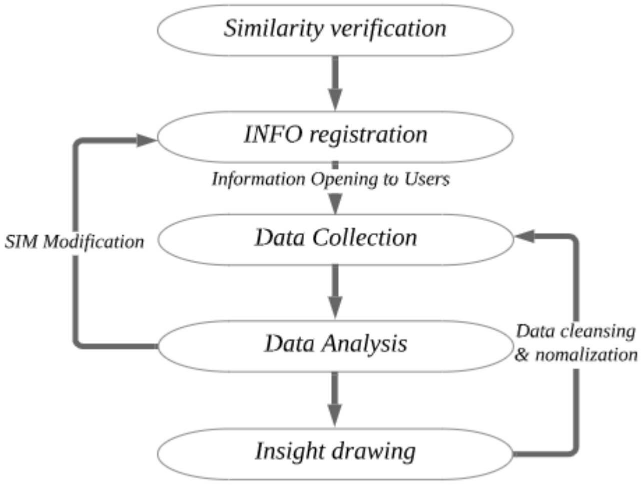


Figure 4. User's detection process schema

The schema of the entire system is divided into user manager function, route function, function, and statistical analysis function. It can be plotted as shown in 3. The central server was developed based on the Spring Framework, and considered the stability of the function development and the possibility of future definability. Mysql was used as the database, and all information processed inside the system was stored through a total of 8 tables. Mobile applications are of two types: administrator and user. The processing of data flowing into the system in real time constitutes the result value according to two branch logics. Fig. As shown in 3, the XML data introduced through the MVC Controller checks whether the count of the currently requested information is greater than 10. Through this, it is first determined whether the information has a history that has been previously collected. If this step is true, information including specific route information is transferred to the branch of step 2. If this step is true, the shortest distance information is delivered using the shortest distance algorithm like a map application. If additional information is required, transfer information provided by a small number of informants of less than 5 people. Through this, when it is determined how to organize the information transmitted in the first branch of the system, statistical processing of the data corresponding to the second branch is performed. In this process, when a data request is issued, all the collection and purification processes of the system are performed from the beginning and the result is reprocessed. Therefore, the system analyzes the history of store information visited by the user and identifies stores in the path. In this process, the system identifies stores that the user is likely to visit. Also, when the system is difficult to grasp such information, it searches for other users to match. The finding method follows figure.4 and the following algorithm.

The following algorithm is performed to find similar users.

The main process for the analysis is as follows.

1. Classification of multiple labels for path cases

2. Applying collaborative filtering for classification of people characteristics
3. Applying the input simulation technique of virtual subjects
4. Application of combination of transfer information
5. Applied in combination with SNS and blog information
6. Application/system feedback for real-time routing information
7. Virtualization group composition for moving data in the same region
8. Pedestrian indexing work for similarity analysis
9. Search/classify similar routes among actual subjects
10. Calculation of predicted scores for similar routes
11. Suggest a technique for reducing the load of the analysis system

In this aspect, as a distributed processing system including a communication module using open source, it has the advantage of easy maintenance. There is a flexible advantage in extensibility on Restful API by using Spring framework. In addition, when various movement data is accumulated, detailed paths are assigned to this path data, and the order is set and stored in units of paths. Through this process, it is used simply by running and moving a smartphone application without registering to a complex system. As an exception case of the system, there is a case that a specific person accidentally registers incorrect information in the function as an information provider. In this case, even though incorrect information is registered or a specific path is modified, the path may become more complicated than the actual possible path when the information is not updated, and thus exception processing proceeds.

IV. EVALUATION OF SYSTEM

In order to systematically verify the effectiveness of this system, It is necessary to check whether the user uses this system to accurately process the store information during the actual movement. In order to verify the effectiveness of such a system, a specific path was designated and 10000 cases of information from information providers were prepared. For the simulation test, the optimal route between the origin and the destination was searched to set the route for the test.

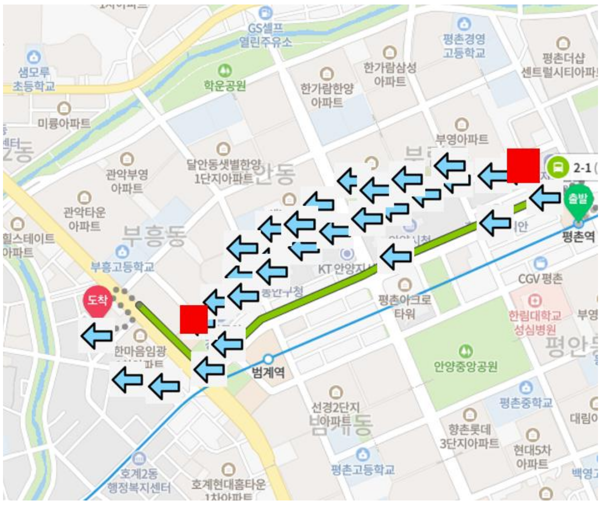
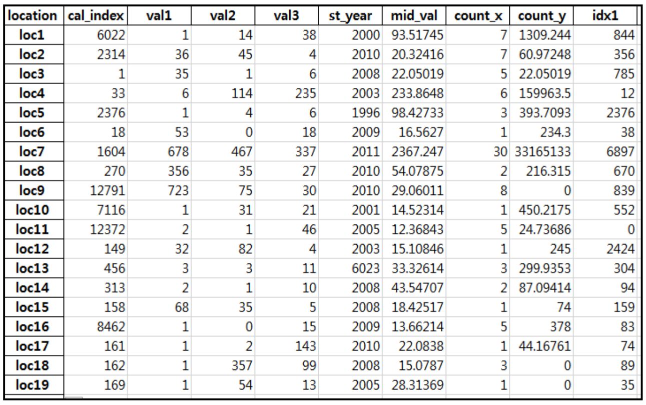


Figure 6. Example case of simulation results [7]

The operation of the entire system function is transmitted through HTTP communication between the central server and the mobile application in real time, so the integrated function test was conducted with a focus on error detection. During the day of May 9, 2020, 5 participants performed the simulation. In the entire process, it was confirmed that the entire implementation function was normally executed without errors. In order to measure the system efficiency from the user's point of view, the system operation was demonstrated with 50 expected users, and the questionnaire was selected from "very helpful" to "not helpful" according to the 5-point Likert scale. The response was received and analyzed with SPSS, a statistical analysis software, to verify the effectiveness of the response result. In terms of measuring the effectiveness of information system functions, we conducted in-depth interviews with three navigation experts, three marketing experts, and three system developers with more than 5 years of experience, as the main users, to discuss the effectiveness of functions, satisfaction in use, and limitations. Proceeded. For the items derived as limitations, feedback was conducted, and functions that can be developed and repaired and those that have limitations in application were classified.



location	cal_index	val1	val2	val3	st_year	mid_val	count_x	count_y	idx1
loc1	6022	1	14	38	2000	93.51745	7	1309.244	844
loc2	2314	36	45	4	2010	20.32416	7	60.97248	356
loc3	1	35	1	6	2008	22.05019	5	22.05019	785
loc4	33	6	114	235	2003	233.8648	6	159963.5	12
loc5	2376	1	4	6	1996	98.42733	3	393.7093	2376
loc6	18	53	0	18	2009	16.5627	1	234.3	38
loc7	1604	678	467	337	2011	2367.247	30	33165133	6897
loc8	270	356	35	27	2010	54.07875	2	216.315	670
loc9	12791	723	75	30	2010	29.06011	8	0	839
loc10	7116	1	31	21	2001	14.52314	1	450.2175	552
loc11	12372	2	1	46	2005	12.36843	5	24.73686	0
loc12	149	32	82	4	2003	15.10846	1	245	2424
loc13	456	3	3	11	6023	33.32614	3	299.9353	304
loc14	313	2	1	10	2008	43.54707	2	87.09414	94
loc15	158	68	35	5	2008	18.42517	1	74	159
loc16	8462	1	0	15	2009	13.66214	5	378	83
loc17	161	1	2	143	2010	22.0838	1	44.16761	74
loc18	162	1	357	99	2008	15.0787	3	0	89
loc19	169	1	54	13	2005	28.31369	1	0	35

Figure 6. Result of the simulation test

The process of analyzing the results of the simulation experiment proceeded.. In this process, the system analyzes the list of stores visited by the user. In the process, the stores along the route are identified, and the system identifies the stores that the user is likely to visit. And the system looks for other users to match. The multi label method is applied to simplify the entire process.

The multil-abel analysis technique based on the similarity between location and human was performed. When analyzing variance data by applying SPARK and KAFKA, batch processing of data groups to be applied first to the recommendation process was performed. As a result of this simulation experiment, it was confirmed that the performance improved by 16.4% compared to the existing methodology [7].

L =i=1∑nBCELoss(y^i,yi) (1)

con BCELoss(y^i,yi)

=−(yi×logy^i+(1−yi)×log(1−y^i))

L =i=1∑nCrossEntropy(y^i,yi) (2)

con y^i∈R3,yi∈{0,1}3 and ∣yi∣=1.

:(y^i,yi)=−yi×logy^i

=−j=1∑3yi,j×logy^i,j

As the actual empirical data is shown in 6. it can be seen that the route has changed to the shortest route bypassing the construction area. In addition, The process of analyzing the results of the simulation experiment proceeded. In this study, both viewpoints of user movement and store marketing are emphasized. At the same time, system stability as a big data system is also important, so a big data distributed processing system was introduced, and overall information processing speed and system stability were introduced. The big data collection engine collected the user's movement information as a log. For the stable processing of data as users increase in this system, additional research is needed on the data separation and storage policy and additional supplementation of data purification techniques.

REFERENCES

[1] Korea’s Internet Competitiveness, "LBS Industry Trend Report”, 2018.

[2] https://www.technavio.com/report/global-location-based-services-lbs-market

[3] Mun, Chang-Bae, and Hyun-Seok Park. "Big data-based Local Store Information Providing Service." The Journal of the Korea Contents Association 20.2 (2020): 561-571.

[4] Steenstra, Jack, et al. "Location based service (LBS) system and method for targeted advertising." U.S. Patent Application No. 10/931,309.

[5] Kushwaha, A., & Kushwaha, V. (2011). Location based services using android mobile operating system. International Journal of Advances in Engineering & Technology, 1(1),

[6] Virrantaus, Kirsi, et al. "Developing GIS-supported location-based services." Proceedings of the Second International Conference on Web Information Systems Engineering. Vol. 2. IEEE, 2001.

[7] Mun, Chang-Bae. " The LBS system with accuracy improvement function using big data." KAIS(Korea Academia-Industrial cooperation Society) Conference proceeding1 1.27 (2020).