Real-time Interpreting and Translating System Based on Analyzing Big Data for Tourists

Seokhoon Kim¹ and Sang-Jin Lee²*

¹Department of Mobile Communications Engineering, Changshin University, Republic of Korea; shkim@cs.ac.kr
²Department of Hospitality and Tourism, Changshin University, Republic of Korea; lee3137@cs.ac.kr

Abstract

In modern life, a travel has been one of the most important lifestyles. Although there are many reasons why the travels have important meaning, it is mainly caused by evolution of ICTs and transportations. Actually, the improvements of transportations have been a trigger of rapid deployment of the travels. However, modern travels aren’t similar to past travels in various aspects. The most different thing is that the travelers or tourists can freely utilize the ICTs in their destination. It means that they can easily search some useful information in anywhere and anytime, and it can provide high elasticity in their trips. Nevertheless, it can’t perfectly solve the language problems until now, because the current interpreting and translating system are based on a machine language. However, if the interpreting and translating system can use exact location and big data, which are used in native speakers of the region, the system will be a best solution. Therefore, we propose the real-time interpreting and translating system based on analyzed big data. It utilizes an exact location of the tourists, and then it will support the most suitable sentence by using big data when they need. Based on the proposed system, we verified that the system outperforms than legacy scheme in various features, and we can show these results in this paper.

Keywords: Big Data, Interpretation and Translation, Real Time, Tourists

1. Introduction

Nowadays, most of people think much of the quality of life. It has been one of the most valuable things in our life, and it is very related to be increased a tourism and leisure business. It is also connected to an introduction of ICT evolutions, which are representing to a rapid deployment of smart phone. That is, people who want to enjoy a trip can easily get information that they need, by using a smart phone which is based on the ICT. For example, there are many things that they should consider to make a happy trip. First of all, they have to plan their itinerary for the trip. At that time, they are usually referring to other's schedules, and deciding to their own plan based on the various information in the Internet. In these steps, they use a lot of ICTs, even though they don’t aware of the detail technologies such as cloud, web, SNS (Social Network Service), search engine, LBS (Location based System) and so on. These scenarios are very common instance in these days. However, the tourists may suffer some troubles, especially language problems, when they go to other countries. To solve the language problems in the travel, they can use some interpretation and translation solutions, such as Google Translation, Bing Translation, and so on. But, it isn't enough to utilize properly when they need. But, it isn't enough to utilize properly when they need. If the solutions can utilize analyzed big data based on the tourist locations, it will be better solutions, which are specialized to the tourists. Unfortunately, these tries aren't very few until now. This is one of the reasons that we propose the real-time interpreting and translating system based on analyzing big data.

On the one hand, in the past few years, LBS services have been one of the most popular services in real life, and it is still and steadily evolving. There are many reasons why LBS services have been popular. First of all, LBS can create new businesses which are very valuable. The most representative example is the navigation device.
in a vehicle. Base on this trend, the related industry of navigation system has been a major business. Secondly, the various services, which are based on the LBS, can be revitalized a stagnant business because these services can be used for various marketing points and new approaches. Thirdly, these LBS services can lead to convenient life, including to help a second class citizen. Finally, it can be a strong catalyst for some businesses. For instance, the MICE (Meeting, Incentive trip, Convention, and Exhibition or Event) industry is a typical business which is the best suitable LBS services.

Actually, the LBS aren't a new technology anymore. It just has been constantly evolving to improve the accuracy and precision. However, this service can be applied to various fields based on the improved accuracy and precision. According to advancing the accuracy and precision, there have been the advents of new LBS base services. It is very similar to the various convergence services which are based on the IT technologies, and it can be also applied to the tourism businesses.

There is one other keyword which we have to keep an eye on the recent ICT trends. It is a big data technology. Most of people have been focused on the potential and possibility of big data, because it will be created a new market to aggregate/store/analyze/apply the tremendous amount of information, various data type and source, and so on. This is the reason why big data technology can be converged on various businesses as a valuable platform. However, the big data technology is in a start step, and there are still some problems we have to solve. Due to the fact that the researches about big data are one of the most active things in the various theme. Although there are many tries which want to utilize a big data technology in recent ICT trend, it is very rare to use a big data to a travel until now. That is, individual or private utilization of big data in their trip is still difficult thing. If they can freely use the technology, it can be more powerful solutions to people, especially interpretation and translation system.

Although the LBS and Big data services can't make the groundbreaking turning point in the tourism business, it can support more convenient tour to the foreign tourists, especially interpreting and translating, and providing tour information. As previously described, there are already some interpreting and translating programs such as Google Translator, Bing Translator, and so on. In addition, there are various interpreting and translating apps which can be used on a smart phone besides the Google Translator or Bing Translator. However, these aren't linked to the LBS services yet, and also it does not specialize in the tourist business. Besides, the various solutions can't properly use big data yet. That is, the foreign tourists should interpret and translate by themselves, whenever they need to express something. To solve these problems, we propose the real-time interpreting and translating system based on analyzing big data for tourists.

The proposed scheme automatically interprets and translates to local tourists where they are. The proposed system automatically interprets and translates for a tourist where they are. During these processes, the proposed system uses the location information and data, which are analyzed information based on the big data.

The rest of this paper is organized as follows. Section 2 describes the related works such as LBS and big data technologies, and Section 3 discusses the proposed interpreting and translating system and its performance evaluations. Finally, we offer a conclusion in Section 4.

## 2. Related Works

### 2.1 Location based System

Generally, LBS is providing information, which are necessary to a user based on the current position. Modern LBS services are usually supported by mobile and telecommunications carriers, location based service providers. Generally, LBS is providing information, which are necessary to a user based on the current position. Modern LBS services are usually supported by mobile and telecommunications carriers, location based service providers. The evolution of mobile communication environments and the rapid deployments of smartphone with wireless and mobile technology are not only spreading converged networks, but also interworking with legacy networks and future networks. These environmental evolutions lead wireless and mobile Internet service paradigms, and it is improving to a life-oriented service from the contents-oriented service.

According to the generalization of LBS services, the LBS component technologies have been evolving to 4 factors, information, identity, position, and terminal, as shown as Table 1. The core component technologies are supported by a mobile and telecommunication company, terminal manufacturer, LBS service provider and so on.

Based on these evolutions, several LBS services have been developed or developing to create a new business. The representative service areas are social network
service based on location, advertising and marketing service based on location, indoor location service, and social safety net implementation based on location information. There have been some substantial efforts in the areas. For example, Skyhook, Inc. has developed the WPS (Wi-Fi Positioning System) solution, Taiwan has been proceeding a M-Taiwan project, Bell-Labs have been developing robot based RF surveying technology, and etc. In addition, Google has progressed a project which called to “Project glass”, and Microsoft Research Group also has a RADAR solutions. The Senion Lab in Sweden has developed indoor location service and user pattern analysis technology. ETRI in Korea has been developing a platform of location based service. The above various examples of LBS mean that the LBS services have been evolving and will be evolved. That is, it caused by using a mobile and wireless infrastructure, and the trends will be more accelerated in the near future. Moreover, indoor location determination technology is very important technology, because most of citizen spend a time in some places where are located to indoor. Due to the fact that we have to focus on the seamless LBS, which can be easily coupled with two technologies.  

### 2.2 Big Data Technology  
There are some hot topics in the ICTs. Among these issues, big data technology is one of the hottest keyword, because big data will be an intellectual platform which can lead a paradigm shift. However, there are still some prior tasks, which are related to big data, such as concepts, technologies, business models, exact applications and so on. To create a value added big data platform, we have to clarify such things. In addition, we should invest to a human resource, technology development, related laws, implementation of infrastructure, etc.

In general, the technologies related to big data are divided into 2 categories. One is a big data analyzing technique, the other is a big data processing technique. The big data analyzing techniques are text mining, opinion mining, social network analytics, and cluster analysis, and then the big data processing techniques are Hadoop, NoSQL, and R, representatively. The detailed descriptions are as followings, as shown in Table 2 and Table 3.

In big data processing techniques, the Hadoop is a kind of key technique. The Hadoop platform is mainly consisted to HDFS (Hadoop Distributed File System), MapReduce, and Hbase. In these components, the

<table>
<thead>
<tr>
<th>Component Technology</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>GIS, Enhanced GIS, Web Data Mashup, Personal Information</td>
</tr>
<tr>
<td>Identity</td>
<td>Device Identity, Network Identity, Service Identity</td>
</tr>
<tr>
<td>Position</td>
<td>High Resolution Positioning with Wi-Fi, Indoor Positioning</td>
</tr>
<tr>
<td>Terminal</td>
<td>Dedicated Terminal, Smartphone</td>
</tr>
</tbody>
</table>

Table 1. Infrastructure component technology  

<table>
<thead>
<tr>
<th>Technique</th>
<th>Features</th>
</tr>
</thead>
</table>
| Text Mining | - It extracts and processes useful information based on natural language processing in unstructured or semi-structured data.  
- It can extract meaningful information in tremendous text data.  
- It can catch a connectivity and correlation to other information.  
- It can figure out the category, which is related to the text.  
- Based on these features, it can make a better result than simple information searching. |
| Opinion Mining | - It can distinguishes reputations in unstructured or semi-structured data into 3 categories, such as positive thing, negative thing, and neutral thing.  
- Based on these functionalities, it can measure a preference.  
- It is usually used in social media services.  
- It is also utilized to analyze market volume estimation, consumer reaction, etc. |
| Social Network Analytics | - This technique is based on the graph theory of mathematics.  
- It measures/estimates user’s reputations or influences in the social networks.  
- It usually uses a connection structure or connection strength, when it measures/estimates the things.  
- Especially, it is used to verify a user, who has a hub role. |
| Cluster Analysis | - It is usually used to combine each object, which has a similar aspect, to find a group, which has an analogous feature.  
- It can categorize to identify a user group, which can be indicated by an interesting, hobby and so on.  
- It is usually used in Twitter, Instagram. |
the legacy system. First of all, the Hadoop has superiority than the legacy database management system, especially it can handle an unstructured data such as video, audio, e-mail, etc. Secondly, the Hadoop will be steadily increasing their territory because it can be easily expanded.

3. Real-Time Interpreting and Translating System

As previously mentioned in Section 2, there are many technologies which are related to big data, and the technologies have been already used in big data analyzing and processing. That is, the analyzing and processing techniques with Hadoop ecosystem have been widely introduced and applied. However, the ICT trends are dramatically evolving to the convergence environments, which need to a swift decision making. However, a traditional Hadoop base analyzing and processing technique can't effectively cope with these circumstances, because it doesn't have included real-time aspects. There are some alternatives so that it is developed to correspond these cases. For example, Oracle CEP, Business Events, IBM InfoSphere Streams, Esper, etc. are representative instances to provide a real-time feature.

In fact, the legacy interpreting and translating system don't be popularly used in real life until now, because the existing system couldn't suggest exact word or sentence. This is the reason why it has very poor interpreting and translating quality in the accuracy aspect. However, if the system can apply to the big data techniques, it will be one of the best solution in this field. Actually, the system with big data techniques might be provided enough translation quality in a translating system. On the other hand, it isn't suitable for a interpreting system, because the system can't support real-time aspects. Therefore, the proposed interpreting and translating system is designed to provide both of them, real-time aspect and high quality interpretation & translation.

In the proposed system, we are introduced the CEP (Complex Event Processing) concepts to provide a real-time attribute. The CEP technique handles data as an event. In addition, it analyzes data in real-time, and it recognizes a meaningful pattern among the data which are consistently occurred, and it immediately reacts a result. That is, the CEP based on In-Memory scheme handles various high speed event stream in real-time, without saving the data to a database or Hadoop. Since we adopted

<table>
<thead>
<tr>
<th>Table 3. Big data processing technique</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technique</td>
</tr>
<tr>
<td>------------</td>
</tr>
</tbody>
</table>
| Hadoop | - It is an open source, and distributed processing project. 
- It is also most preferred big data solution until now. 
- It can handle structured/unstructured/semi-structured data. 
- It includes to HDFS, Map Reduce, Hbase, Nutch, Pig, Hive, HCatalog and so on. 
- Basically, it is configured to x86 server, which has a high cost-effectiveness. 
- Large virtualized storage is set in the x86 server. 
- Map Reduce based on the JAVA can process the tremendous data set in the HDFS. |
| NoSQL | - NoSQL means Not-Only SQL or No SQL. 
- It is different from the traditional Relational Data Base Management System (RDBMS). 
- It is designed to support non-relational database. 
- There are some representative NoSQL solutions such as Cassandra, Hbase, MongDB, etc. 
- It doesn't be fixed the table scheme. 
- It doesn't support to the Join between tables. 
- It has advantages about horizontal scalability. |
| R | - It is a kind of open source project. 
- It provides a language and development environment. 
- Based on these environments, user can easily use a modeling method, data mining, and so on. 
- The implemented results can be visualized to a graph in the R. 
- It can easily link to other language such as JAVA, C, Python, etc. |

HDFS is a file system which is fundamental element, the MapReduce is a framework for distributed processing, and Hbase is a kind of distributed database. That is, the whole Hadoop frameworks don't need a software license by Apache software license policy. Due to the fact that the Hadoop has been an original technology in big data businesses which are used in JP Morgan, IBM, Booz and Co., New York Times and so on. Although there are some difficulties when people want to adapt to the Hadoop technology, the Hadoop technology has certainly benefits than
the CEP technique, the proposed system can provide the real-time aspects. Additionally, the proposed system works with location information when starts interpretation and translation, because it is very useful functions for the tourists.

As shown in Figure 1, the proposed system provides real-time interpreting and translating results for convenience of tourists. At that time, the proposed system verifies the user's exact location based on the GPS, and the information will be included in the request message. The cloud system, which received the request message from the user, will transmit the reply message to the user. In the reply message, it includes an interpreting and translating result, which is best suitable word or sentence based on the analyzing and processing of big data. The detail steps are as followings.

- Step 1: Verifying the exact tourists location
- Step 2: The tourists or users will send an interpreting and translating request message to the cloud system. In this case, the request message includes an user location information based on GPS.
- Step 3: The cloud center immediately starts the analyzing and processing of big data in real time, whenever the center received the request message from the tourists. In this step, the big data center in cloud uses the tourist location information to analyze and translate the request message. That is, the system utilizes various information when performing the analyzing and processing of big data, which are related with a lot of information such as local/region information, commonly used word or sentence by native speakers or tourists and so on. Based on these processes, the system can find a best suitable word or sentence. In addition, the system is basically working on the CEP technique to provide an exact result in real-time.
- Step 4: The system sends the reply message which have an interpreting and translating result.
- Step 5: The above steps (Step 1 ~ Step 4) will be repeated.

On the one hand, the system also has another functionality. In this case, the system consistently suggests a best suitable word or sentence based on their location information. That is, although there are no request messages from the user, the system can recommend a best word or sentence. For example, if the tourists visit a store to buy something, the system will send a best suitable word or sentence to the tourists based on their location information and analyzing/processing big data.

These are the results comparing with Google translation and Bing translation, as shown in Table 4. We compare the system with these tools, because these tools are very popular tools until now.

As shown in Table 4, the final results can't provide a perfect accuracy yet. However, the proposed system outperforms the accuracy than others, because it can use an exact location information and analyzing/processing big data. In the performance and analysis test, we used two languages, Korean and English, in interpretation and translation. Although, the proposed system's response time is slower than others, it isn't critical problem because it is enough acceptable ranges.

4. Conclusions

The LBS service is not only one of the inevitable IT trends, but also one of the representative examples. It can also be an applicable word in big data. Although the LBS and big data services aren't an only solution for raising business
effectiveness and convergence, there is no doubt that it is one of the best solutions until now. Based on the proposed system, we expect that the tourists can utilize a lot of useful information in their location. For make a better system, the proposed system should be supported various languages.

5. Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1A2060035).

6. References