ISSN (Print) : 0974-6846
**Indian Journal of Science and Technology,** *Vol 9(26), DOI: 10.17485/ijst/2016/v9i26/97329, July 2016* ISSN (Online) : 0974-5645

# Second Language Learning Convergence System Based on Big Data

**Jae-il Yi***

Department of Liberal Education, SEMYUNG University, Republic of Korea;
nayltd@semyung.ac.kr

## Abstract

**Background/Objectives**: In this paper, we explore an effective way to figure out a simple and efficient means of Second Language Acquisition through Big Data utilization. The final goal is to develop a system that enables instant access to the information that is specifically arranged for certain individuals. **Methods/Statistical Analysis**: The ability for searching the information that second language learners have in mind differs by the individual's information literacy. Also, learners show different language aptitude and efficiency when they study the target language. Tracking a massive flow of information that individuals can take advantage of is the key feature to be investigated. Developing a system that takes account of individual's information literacy and aptitude can show positive effect in the learning process. **Findings**: Proposed system mode l, Big Data Language Education system (BDLEs) presents an effective language learning process through the utilization of Big Data that is subcategorized for its core purpose. BDLE system like the existing models has the function of determining the output for each individual hoping to find the right information they are looking for. It sends out a massive amount of information, however, it doesn't seem to fulfill the users' desires since the results appear different depending on the information literacy. The strong point of the proposed system is that it is focused on to design an advanced system that can complement the insufficient portion of the existing models. It diversifies the output by subcategorizing Big Data into a unit that enables the system to pick out a more specific and accurate information by analyzing and comparing the information literacy and learning aptitude of learners. **Applications/Improvement**: Utilization of well-organized Big Data complements the lack of information literacy. It will function as a time saving device for individuals in a busy modern society on condition that sensitive personal information is securely protected

**Keywords:** Data Categorization, Information Extraction Language Aptitude, Mobile Language Education, Second Language Learning

## 1. Introduction

Nowadays, the cost for storage and process of data has become much cheaper than the past. As a result, data that had to be disposed in the past can be gathered and processed and the widespread of Cloud Computing is also a cost-reducing factor and makes it possible to eliminate technical barriers. Due to an expansion of IoT infrastructure along with an explosive increase in multimedia content and IT industry, Big Data analytics is currently utilized in various fields including education sector. In other words, convergence has become a hot issue in the field of education. A convergence education represents combined professional practices and technological knowledge to help students in learning skills required for the internet-based practice of communication[1]. Also, new technology including the utilization of Big Data is occupying a large proportion in the education sector. As modern society is transformed into a highly information-oriented society, traditional learning methods that were limited to school boundary have dramatically changed into an e-learning environment. These environmental changes act as an important element in education since learners can immediately access the information no matter when and where they are. In addition, with the help of appropriate information system, they can surely rely on the information they requested with a bonus feature of related information.

* *Author for correspondence*

Therefore, this study proposes a system that utilizes Big Data for the purpose of setting up an effective and systematic Second Language Learning environment. An important factor in Big Data is gathering a considerable amount of information but utilizing the collected information for appropriate purpose is more important. To this end, we will first look through the basic notion of convergence education and Big Data then go through the case studies of Big Data utilization in education sector. Finally, we suggest Convergence Education System for Second Language Learning based on utilizing Big Data. We hope that suggested model could be an effective system for efficiently utilizing massive information in Second Language Learning which also possesses the power to give away not only the information requested but also some additional information related to the requestor that shorten the time spent in learning process accordingly.

## 2. Background and Related Work

### 2.1 Big Data and Mobile-Assisted Second Language Learning

Over the past years, an increase of vast amount of data has occurred[2]. The term Big Data refers to the chunks of data that goes beyond the limit of acceptable range of software that process, store and manage in a given time. This definition is very subjective since the volume of data to be recognized as Big Data is unsettled. The technology development is opening a path for the increase of datasets so a couple of terabytes can no longer be categorized as Big Data. Also, what kind of software are used or how much data is generally used in a certain industry can be a factor for the definition of Big Data. According to this principle, the volume of data which can be categorized as Big Data refers to a range of datasets from a few dozen terabytes to multiple petabytes. The main purpose of Big Data is collecting, processing and analyzing various data for its optimal purpose to estimate the flow of information that is normally difficult to guess[3]. Big Data has the following three dimensions as shown in Figure 1[4].

There are three aspects in Big Data categorized under the name of data volume, data variety and data velocity so-called three Vs. In addition, each of the Three Vs has its own ramifications for analytics[4].

- Volume

  Includes not only the physical size but also the datasets such as media, location information and video information and decides whether it can be processed or not.
- Velocity

  Include data processing speed and storage speed. If in need, returns the processed results to the request of numerous users in real time.
- Variety

  Classification of Big Data by its structure
  - Structured: data which is stored in a fixed field (Relational database).
  - Semi-Structured: non fixed field including the schema such as XML, HTML.
  - Unstructured: non fixed field data such as text, image.

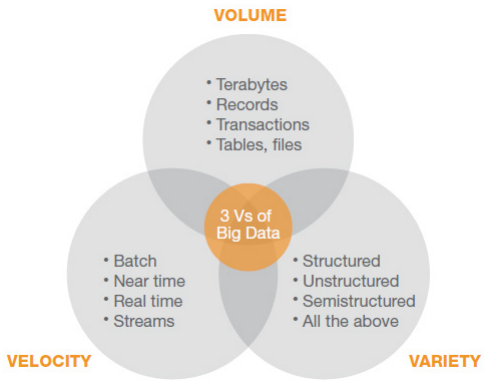


**Figure 1.** Three vs of big data.

From the perspective that these three Vs altogether have to be considered in order to become a valuable data, Value can be considered as another aspect of Big Data[5,6]. The importance of Big Data analytics keeps increasing as it is considered an important tool for improving efficiency and quality in many organizations[7].

### 2.2 Mobile Assisted Second Language Learning

Online education is a system that exchange information between distant computers via Internet to proceed with proper education. Following the establishment of high-speed communications network worldwide, more and more people avail themselves of this technology to utilize multimedia contents for their use. Through the activation of online education, learning system based on internet

service, so-called web-based learning or e-learning, draw people's attention. This phenomenon also had an effect on Second Language Learning. The traditional notion of a passive learner has developed into the notion of dynamic and active learner since it is essential for learners to search for the information or learning materials in order to make self-directed learning nowadays. This kind of online education based on internet or digital media offers an ideal learning environment for learners to mutually communicate by eliminating the time and space limitation of traditional education environment.

Not all learners are inherited with the same competency in Second Language Learning. Standardized measurement method such as Modern Language Aptitude Test (MLAT) and Language Aptitude Battery (LAB) are designed to show the standardized personal language aptitude difference among learners. Four aspects of this test are defined as follows[8]:

- Phonetic coding ability – ability to store new language sounds in memory.
- Grammatical sensitivity – the individual's ability to demonstrate his awareness of the syntactical patterning of sentences in a language.
- Inductive ability – ability to examine language material to notice and identify patterns and correspondences and relationships involving either meaning or grammatical form.
- Memory Abilities – measure of a special kind of rote-learning ability that appears to function in foreign language learning situations.

N. KarthickN and X. Agnes Kalarani introduced the method to extract the meaningful information from the large amount of data and provide the aggregated form of output to the users[9]. Therefore, establishment of the Big Data that can supplement learning ability through analyzing difference between each learner can bring out positive effects from each individual. Especially, in the world where time and space limitation have disappeared thanks to smart devices such as a smart phone and tablet PC utilization of web-based Big Data information can be a perfect assistive device for resolving various difficulties in Second Language Learning.

One of the aspects emphasized in mobile assisted learning is that development of language learning environment before the role of mobile technologies is decided[10]. Looking at the mobile language education in technical terms, it is an education that gives an option for learners to choose certain education programs depending on personal taste and aptitude utilizing various mobile devices, cloud computing environment based on high-speed Internet. Such a system makes it possible for learners to adopt an innovative system that allows learners to fit themselves into personalized learning system for optimal purpose in enhancing learning capabilities, training content and training methods. Educational activities carried out by students and teachers can be transformed into a vast amount of accumulated educational information through digitalization. And establishing an optimal learning system to provide individualized learning for person in need by utilizing Big Data processing technology to make use of these educational data is indispensable. For Big Data convergence education, all the learning content and learning activity data should be gathered in order to accomplish the conditions of Big Data and managed in a high capacity processing server under the elaborated privacy policy.

## 3. Big Data Language Education System

In this part, we explicate the structure of the suggested model explaining the constituent parts and the process which are designed in the system. Also, we illustrate the method for gathering and categorization of Big Data which is to be transmitted to the requestor. And finally, we explore the way to make the suggested model to become useful and effective in Second Language Learning.

### 3.1 System Outline

If we take a glance at the suggested model, tentatively named BDLEs, it is mainly designed to gather an enormous amount of information that is needed in Second Language Learning through various routes and categorize the information in a certain way to make it useful for individuals when it is transmitted to the requestor. To build Big Data for Second Language Learning, if we take English for example, standard information such as English grammar, intonation, pronunciation, idiomatic expressions including generalized slangs should be collected. And the information that are generally related to an English using environment e.g. the requestor's first language, major, interested related field etc. also

needs to be gathered. Thereafter, collected information is categorized to match the best sector for individual purpose. Finally, categorized information is divided into key information and related information then transmitted to the requestor. Figure 2 is a rough draft of the suggested model for Big Data Language Education System.

## 3.2 Process of the Proposed System

The following is a basic overview of BDLEs processing procedure.

- Intelligence vessel in BDLEs collects educational information from external database and internal database.
- If needed, BDLEs sends out requests for Learning Information to supplement insufficient information.
- Collected data in stored in 1st Big Data section.
- 1st Big Data is classified into subcategories through Collected information is classified into subcategories through Information Categorizer.
- Categorized information is stored in the 2nd Big Data Section.
- When there is a request for educational information from the requestor, the Priority Determiner sorts out information by priority ranking.
- Info Extractor receives the priority ranking from the Priority Determiner and transfers the information to the requestor by order.

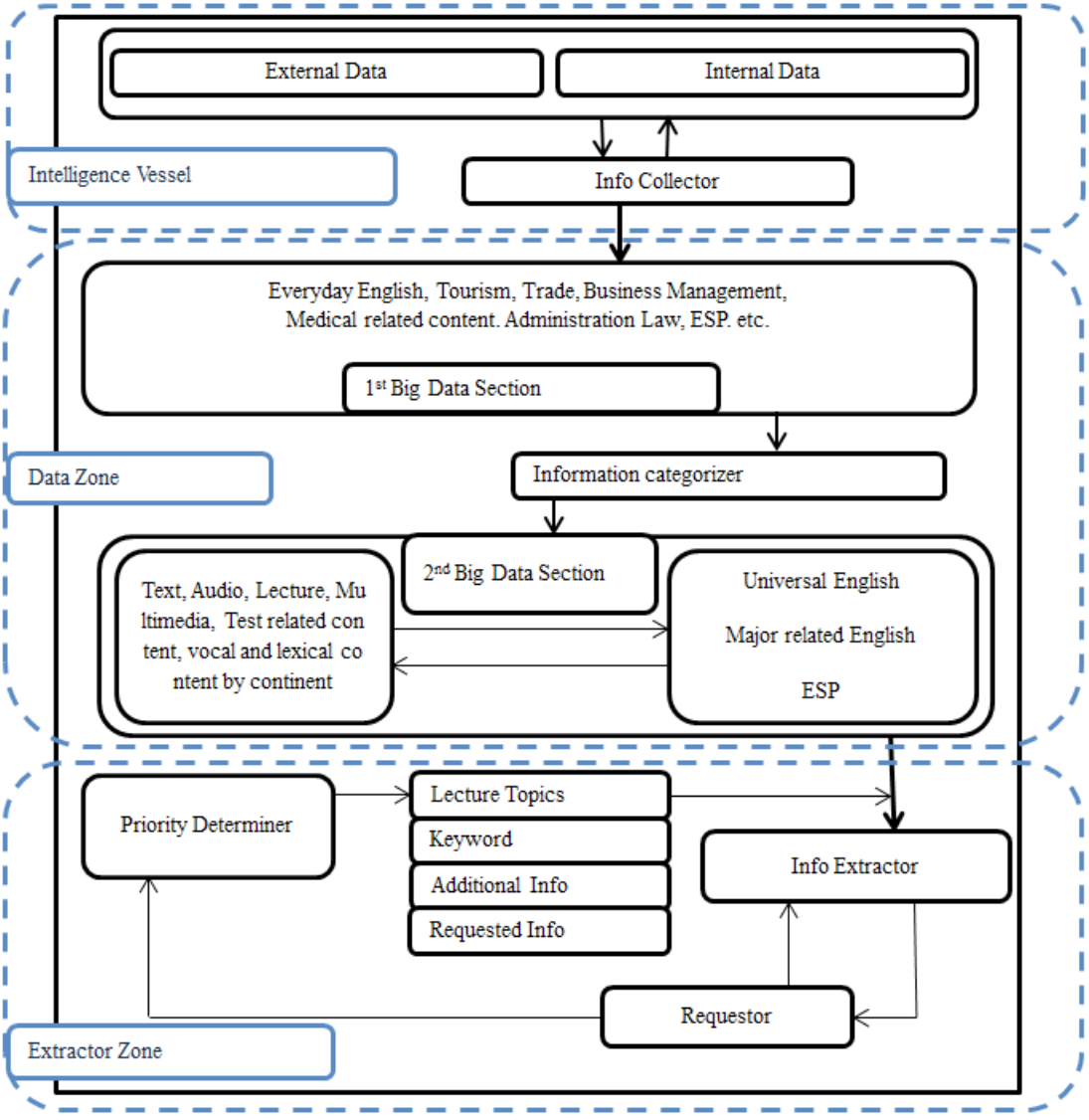


**Figure 2.** Big data language education system.

- Info Extractor extracts the optimal information for the requestor and then finally transfers the information to the requestor by order.

BDLEs is basically divided into three parts, which is Intelligence Vessel, Data Zone and Extractor Zone. First, Intelligence Vessel collects external data from online service provider and internal data from school database. Also, if additional information is required to accumulate enough Big Data, Info Collector demands for necessary information. Second, collected data which are somewhat at random are stored as Big Data in the form of a massive lump in the 1st Data Section. To make this information more optimal, Information categorizer classifies the massive lump of data into diversified subcategories. It is favorable to fit in a remanding process to review the information which is rejected by Info Collector to raise the usefulness of the collected data but for now we will concentrate on the system which is loyal to the basic function. Based on the collected data, Information categorizer classifies the information into two parts. The first part is subcategorized into text, audio, lecture, multimedia, test related data and then these data go through another sub-categorization to break it down by vocal and lexical difference between each continent. This is mainly due to the status that English possess. Currently, not many people can deny that English is the global language and intonation, pronunciation, lexicon etc. differs by region. So it is an important factor that cannot be missed out in the process to determine the priority in the Extractor Zone. Third, Priority Determiner automatically fetches the Requestors personal information when they log on to the system to set up the basic order of priority. Fourth, Requestors give the input for the information they need. Finally, the Info Extractor provides the most optimal information through a series of procedures. In this way, Requestors can reduce overlapping investment of time and energy searching for the right information they had in mind in the first place. One of the most important processes in the suggested model is the intervention of the Priority Determiner since it plays an important role in delivering the optimal information to the requestor.

## 4. Evaluations of Suggested Model

In this part, we will evaluate the efficiency of the existing system that we have proposed in the earlier study and the proposed system by comparing the difference between each model[11] (Table1).

**Table 1.** Comparative analysis of each model

| Category | Existing Model | Proposed Model |
|---|---|---|
| Information Collection | O | O |
| Information Analysis | O | O |
| Online learning contents | O | O |
| Priority Determiner | X | O |

Both of the existing model and the proposed model have been designed to contain the feature of developing an efficient system for Second Language Learning. On the other hand, the two models show difference in constructing systematic educational information. In the process of data collection, the existing model passively gathers only the information that is provided by an online service provider but the proposed model not only gathers a massive amount of information but also demand additional information in need through Info Collector. As a result, the proposed model provides a detailed and high-quality feedback for the requestor based on the accumulated Big Data. Also, the existing model extracts the educational information directly from the information database that is not categorized for specific purpose resulting in transmitting the information randomly to the requestor, whereas the proposed model sub-categorizes the collected Big Data to make it useful in ordering the priority rank which makes the transmitted information more efficient in educational purpose. In case of analyzing the requested information, the utilization and intervention of Priority Determiner helps constructing a database of each requestor which in turn leads to re-setting the most optimal education information to be extracted for the requestor.

## 5. Conclusion

In many occasions, especially when it comes to learning a second language, learners generally prefer making a fortune out of less effort. This study focused on the measures of establishing a system that is useful in Second Language Acquisition by utilizing Big Data to compliment the different language aptitude and information literacy of individuals.

The output of this study will allow foreign language learners to access the information in need by logging on to the network through the device in their hands which

in turn makes it possible to keep on studying the target language in real time whenever and wherever they are. This kind of favorable environment which is efficient for receiving the most optimal educational information will lead the learners to a higher level of language proficiency. Also, eliminating the unnecessary information by setting up an priority ranking shortens the time of learning to secure the time for other activities. Despite the many benefits, the weakness lies in the accumulated personal information. Since the system is devised to accumulate personal information, information leak by system malfunction or hacking can lead to a disastrous outcome. Thus, more research is required on figuring out a way to minimize collecting personal information without affecting the quality of the system and improving personal information protection technology.

## 6. References


1. Sarachan S, Jeremy J. The path already taken: Technological and pedagogical practices in convergence education. Journalism and Mass Communication Educator. 2011; 66(2):160-74.
2. Chen M, Mao S,Liu Y. Big Data: A survey. Mobile Networks and Applications. 2014; 19(2):171-209.
3. Sagiroglu S, Seref S, Sinanc S. Big data: A review. Collaboration Technologies and Systems (CTS). International Conference on IEEE; 2013. p. 42-47.
4. Russom R, Philip P. Big data analytics. TDWI Best Practices Report, Fourth Quarter; 2011. p. 1-35.
5. McAfee M, Andrew A. Big data. The management revolution. Harvard Bus Rev. 2012; 90(10):61-67.
6. Gantz G, John J, Reinsel D. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the Future; 2012. p. 1-16.
7. Bifet B, Albert A. Mining big data in real time. Informatica. 2013; 37(1):15-20.
8. Carroll C, John B. Cognitive abilities in foreign language aptitude: Then and now. Language aptitude reconsidered; 1990. p. 11-29.
9. Karthick N, Kalarani A. An Improved Method for Handling and Extracting useful Information from Big Data. Indian Journal of Science and Technology. 2015 December; 8(33):974-81.
10. Chinnery C, George M. Emerging technologies. Going to the mall: mobile assisted language learning. Language Learning and Technology. 2006; 10(1): 9-16.
11. Yi J, Lee S. A Study on Extracting Education Information based on IoT. Internation Journal of Applied Engineering Research. 2015; 10(90):2661-2700.