International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.2 (2014), pp.319-330
http://dx.doi.org/10.14257/ijmue.2014.9.2.32

Development of Electric Trading System Using Big Data

Cheol-Woong Lee[1], Sae-Hong Cho[2], Jong-Wook Kim[3] and Dae-Hoon Hwang[4]

[1]*Dept. of Digital Culture Technology & Contents, at Hansung University*
[2]*Dept. of Multimedia Engineering, at Hansung University*
[3]*Dept. of Media Software, at Sangmyung University*
[4]*Dept. of Computer Science, at Gachon University*

cyber@hansung.ac.kr, chosh@hansung.ac.kr, jkim@smu.ac.kr, hwangdh@gachon.ac.kr

Abstract

Development of Smart Apparatuses, represented by Smart-phone, and development of internet and information & communication technology has changed the pattern of our life. With wireless communication and smart apparatuses, people now can create, share, and consume information at any time and any where. In addition, development of technology has allowed the large-scale data of image, voice, and image to be spreaded and shared, from the existing data consumption pattern of users that was mainly consisted of the text. With this, the quantity of data to be consumed by individual has been geometrically increased. At last, the age of Big Data has been emerged. The big data existing in the life have included a vast information. Now, analysis and utilization of these data in what kind of aspect has been more important. In this study, the status of Big Data have been looked through and the method of its utilization in the global online business field has been suggested.

Keywords: *Big data, U-trade Hub, Data mining, Data analysis*

1. Introduction

A paradigm shift from the PC era and to the Internet era, mobile era and smart era has rapidly increased the creation, utilization and use of data. IDC (International Data Corporation) said global digital information volume in 2011 was about 1.8 zettabytes and it will reach 35.2 zettabytes in 2020 with exponential growth of data [1].

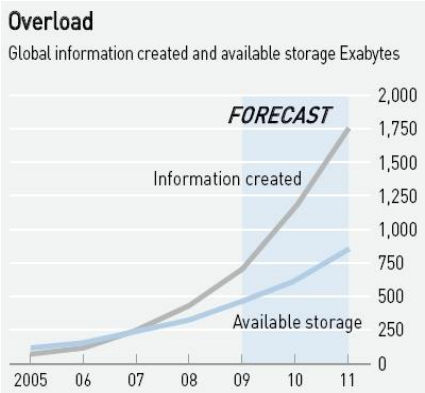

Figure 1. Data overload

ISSN: 1975-0080 IJMUE

The big data possessed by all the businesses has reached the size enough to extract huge value, and thus, those businesses are faced with a situation that which company will extract the value ahead of others may determine firm's success or failure [2]. The World Economic Forum pointed out big data as a technology to take note of in 2012 and selected a solution to the data overload problem and data utilization through data capitalization as the top priority pending issues [3]. The quantitative hugeness of big data is expected to shift impossibilities to possibilities in numerous fields, and Google's big data solution has been confirmed as a magic myth transforming IBM's failure into success [2]. Valuable data, however, accounts for just a small portion of data including text, document, phone call record and large scale e-commerce list. Nowadays when all the businesses face with a situation that the success or failure of a company depends on which business will extract the value first out of the Big data possessed by all the businesses, a technology to analyze large capacity data and develop it as meaningful data is needed [2]. In this context, a need to effectively utilize big data arises through research on big data analysis technology and application model.

2. Concept and Characteristics of Big Data

2.1 Concept of Big Data

Although, the definition of big data started from the data size and technology aspects, there is no specially and quantitatively agreed definition on the big data. Actually, the trend in the definition of big data is that the meaning of big data expands to the value and utilization effect aspect [4]. The Big data Strategy Research Center of National Information Society Agency has defined big data as hardware and software composing big data and a huge platform meaning all the processes encompassing the hardware and software [5]. SERI (Samsung Economic Research Institute) defines big data as ultra large capacity data, which is difficult to manage and analyze with the existing technology, because the forms of big data are diverse and have fast life cycle [6]. IDC defines big data as the next generation technology and architecture designed to extract economically necessary value through high speed capture, data search and analysis from vast volume of data comprised of a variety of data [8]. Gartner defines big data as large volume of data coming in various forms continually in real time [9]. McKinsey has defined big data as large scale data difficult to collect, store, manage and analyze with conventional database software [7].

Actually, big data includes enormous data gathering difficult to handle with the existing management and analysis systems, and the platforms and analysis techniques to solve the problem. Big data passes through such processes as data creation → collection → storage → analysis→ expression. And, detailed areas and relevant technologies emerge in each process. The analysis technologies include statistics, data mining, machine learning, natural language processing, pattern recognition, social network analysis and video, audio and image processing. Infrastructure encompassing the utilization, analysis and value of big data includes BI, DW, cloud computing, distributed database (NoSQL), parallel distributed processing, distributed file system (HDFS) and MapReduce [9].

Table 1. Technological Area by Big Data Processing Process [9]

Flow	Area	
Source	Internal Data	Database, file management system
	External Data	File, multimedia, streaming
Collection	Crawling	Data collection using search engine robot
	ETL (Extraction, Transformation, Landing)	Source data extraction, transmission, conversion and storage
Storage	NoSQL Databases	Unstructured data management
	Storage	Big data storage
	Servers	Ultra lightweight server
Processing	MapReduce	Data extraction
	Processing	Multiple work processing
Analysis	NLP (Neuro Linguistic Programming)	Natural language processing
	Machine Learning	Data pattern detection through machine learning
	Serialization	Data serialization
Expression	Visualization	Expression of data or diagrams graphically
	Acquisition	Acquisition and reinterpretation of data

Most analysis techniques have already been used in the statistics and computer science fields, especially, machine learning/data mining fields and the algorithms of these analysis techniques are improved to be suitable for large scale data processing so as to be applied to big data processing. The big data analysis techniques include the following:

• Text Mining

Text mining is a technique to extract and process useful information, based on natural language processing from unstructured/semi-structured text data. Through the text mining technique, meaningful information can be extracted from the vast text bundle, and linkage with other information is identified and the category that text has can be found: namely, results beyond simple information search can be obtained.

Large capacity language resources and statistical and regular algorithms are used for a computer to analyze languages (natural language) that humans use and find out the information hidden in the languages. Major applied fields include document classification, document clustering, information extraction and document summary.

• Opinion Mining

There is a technique called opinion mining or sentiment analysis in the field related with text mining. Opinion mining is a technique to distinguish between the positive, negative and neutral preferences of structured/unstructured text such as social media. The opinion mining is utilized for market size prediction and the consumer response and viral analysis of specific service or goods. For accurate opinion mining, the accumulation of expressions/words indicating preferences by experts is required.

• Social Network Analytics

Social network analytics is rooted in the graph theory of mathematics. The social network analytics is mainly utilized to find a user playing a role of viral center or hub on the social network by measuring user reputation and influence, based on social network connection structure and strength. A user having influence on the social network is called an influencer and the monitoring and management of the influencer is important from a marketing standpoint.

• Cluster Analysis

Cluster analysis is used to discover the groups having similar characteristics by combining individuals with similar characteristics. For example, there can be a user group mainly talking about photos/cameras and a user group interested in cars on Twitter. These user groups, according to interest or hobby, can be classified through the cluster analysis.

2.2 Characteristics of Big Data

An IT market research agency, Gartner, explained big data by adding complexity to 3V in its report, "Big Data Analytics" published in January 2011. IBM defined big data as a new type of data having 3V, through which an opportunity to gain insight that could not be answered can be acquired. SAS presented 4V as basic concept by adding "value" to 3V.

Table 2. Four Characteristics of Big Data

Classification	Details
Volume	• Digital information volume surges exponentially each year, due to technological development and routinized IT life, and thus, a zettabyte era has emerged.
Velocity	• Real time information increase such as object information (sensor, monitoring) and streaming information • Increase in velocity of data creation and movement (distribution), because of real time attributes • Data processing and analysis velocity are important for large scale data processing and valuable present information (real time) utilization
Variety	• The types of data including log records, social network, location, consumption and reality data augment. • Diversification of unstructured data types like multimedia in addition to text
Complexity	• Issues including unstructured data, difference of data storage methods and redundancy • Increase in management subjects, due to data type increase and external data utilization • The complexity of data management and processing deepens and new techniques are required.

In addition, big data can be classified as follows, according to the degree of standardization.

Table 3. Types of Big Data

Category	Details
Structured	Data saved in the fixed field For example: Relationship-type database and spreadsheet
Semi-Structured	Although it has not been saved in the fixed field, it is the data including meta data or schema. For example: XML or HTML, text
Unstructured	Has not been saved in the fixed field. For example: text document and image/video/voice data, etc. available for text analysis.

Reports including the report of Gartner explained the different characteristics of big data from the existing data as follows [8]:

• Quick decision making is required relatively less.

: A long-term and strategic approach is necessary by focusing on the analysis of large capacity data. Unlike immediate processing velocity required for existing data processing, r immediate decision making is required relatively less.

• Processing complexity is high.

: Processing complexity is very high, due to various data sources, complex logic processing and large capacity data processing, and thus, a distributed processing technology is needed to solve such a high processing complexity.

• Data volume to process is huge.

: For clickstream data, customer information collection and analysis, for an instant, should be conducted for the long-term, and thus, the data volume to process is enormous, compared to existing methods.

• Unstructured data ratio is high.

: The ratio of unstructured data files including social media data, log file, clickstream data, call center log, communication CRD log is very high, which is a factor to increase processing complexity.

• Processing/Analysis flexibility is high.

: The flexibility of processing and analysis is high, because of no well-defined data model, correlations and procedures, compared to the existing processing methods. Flexibility needs to be guaranteed to accept new and various processing methods.

• Simultaneous throughput is low.

: Data volume requiring simultaneous processing is low, due to the characteristics of large capacity and complex data processing. Therefore, big data processing is not suitable for a data analysis that assures real time (semi-real time) processing.

2.3 Status of Big Data Utilization

Global companies bolster their capabilities by reshuffling organizations, centered on data, to preempt and dominate big data markets. For capability consolidation, they simultaneously

undertake aggressive M&A to preempt markets, as well as technology development[9]. The utilization of some companies is presented below:

Google

• A flu forecasting system is offered through an analysis ofcold-related search word.

• As a result of comparison with the data from the U.S. Centers of Disease Control and Prevention, a very close correlation was confirmed between search frequency and patients showing flu symptoms. Through this, a flu diffusion early warning system, "Google Flu Trends," has been devised by investigating search word query frequency related to flu including flue and influenza on Google homepage.

• Offers flu-prevalence information by time and region ahead of the U.S. health authorities.

Zara

• Pursues fast fashion through internal production and inventories management in real time.

• Analyzes the sales status of Zara's special offer products in real time by developing an inventory optimal distribution system through which maximum sales can be achieved with an analysis of sales and inventory data of stores worldwide jointly with MIT (Masscults Institute of Technology).

• Devises a system that can supply clothes in line with quickly changing trends by linking with the production system through an analysis of popular products in the markets.

Coca Cola

• Analyzes data on SNS and the Internet message boards in real time and utilizes the analyzed data as an internal asset.

• Real time response can be conducted such as fortifying PR in the regions or countries, where unfavorable information to Coca Cola increases through the analysis of data obtained from various social media and by reflecting the data in sales-linked decision making.

Walmart

• Operates "@Walmartlabs" using each branch's mobile and social shopping characteristics.

• Synergy is exerted between distribution and e-commerce by managing social network and content through the acquisition of a social media company, Kosmix.

• Social Genome is a technique to promote the sales of goods using real time-interpreted and extracted information by collecting large scale data through social media. This technique can prevent unnecessary inventory waste and supply customer-desired goods sufficiently by analyzing rapidly changing consumer patterns and quickly providing goods required at the right time and in the right place. Because, all these are connected to customer satisfaction enhancement per branch, the social genome plays a role of a virtuous circle for the firm's development.

SAS

• Predicts social change through an analysis of feelings or mood.

• Analyzes the feelings or mood of social media data on the Internet chatting, blogs, Facebook and Twitter in the U.S. and Ireland for the past two years.

• Confirms that unemployment rate soars in four months, if chatting including the wordings of depressed, being pissed off, etc. increases in the U.S.

• In Ireland, the uneasy or restless feelings have spread five months before unemployment rate goes up and the wording of convinced or assured has sharply dropped two months before unemployment rate goes up.

3. Concept & Status of U-Trade Hub

3.1 Concept of U-Trade Hub

Article 2.6 of the Foreign Trade Act defines e-Trade (U-trade) as trade made fully or partially using devices having information processing capability like computers and information communications network. Article 4.1 of the Act on the Shaping of Trade Basis defines e-trade (U-trade) as the trade carried out through information communications network including the Internet. The Act on the Promotion of e-Trade (U-trade) defines as trade, under Article 2 of the Foreign Trade Act, handled by e-Trade (U-trade) documents in full or in part. Kim defines e-Trade (U-trade) as the trade processed by e-trade (U-Trade) documents fully or partially, where trade is conducted using the devices having information processing capability, such as a PC, and information communications network. According to Kim, e-Trade (U-Trade) is conducted via a single window with which a trading company can process in blanket without discontinuity from market research to contract, customs clearance and payment on the Internet environment through information and communications technology development[10]. UN classifies the development level of e-Trade (U-Trade) into five levels. The start level is to implement customs clearance automation. At this level, only customs clearance is conducted electronically and paper documents are used for remaining work. In the growing level, the automation of customs clearance and export/import work is made. The primary level is to utilize e-Trade (U-Trade) through integration of the customs and export/import related agencies (bank, port, logistics company). In the medium level, the basis, where export/import companies handle all trading work through national integrated platform, is devised. In the advanced level, linkage service between countries is embodied, based on national integration platform.

3.2 Characteristics of U-Trade Hub

The trading areas expansion and environmental change acceleration can be essential, according to wireless communications environment and smart information device-related IT technology utilization, and thus, the evolvement velocity of e-Trade (U-Trade) gets rapidly fast. Such an evolution includes all trading elements such as service, technology, labor force and capital, beyond the limitation of existing goods trade. From this point of view, the characteristics of u-Trade are as follows:

• Utilizes the increase of the wireless Internet and Web technology through ubiquitous computing environment.

• Bases on the wire+wireless network of mobile equipment utilizing cellphone, PDA, wearable computing and next generation hand-held appliances including existing PC-based wire network.

• Creates an opportunity of a commercial transaction unconsciously and executes it by using devices and objects (machines) having autonomous computing function without human's consciousness.

• As on/offline integrated commercial transactions are possible, the efficiency enhancement of order/processing of transactions can be actualized in real time.

• Information disclosure becomes easy, transparency improvement and goods' operation path are traced in real time, and processing suitability can be checked.

• Intelligent marketing that can predict, trace and discover customer needs is available, because real time recognition and tracing of customer situation information & goods situation information and communication with the information is possible anywhere, anytime by the sensor, chip, tag and label embedded in terminals (hand-held gadgets) and objects.

• The participants in trade increase, according to environmental change, in which the situations of all members related to trade such as general consumers, commercial transaction expert groups and distributors are traced and interfered as real time global ubiquitous network environment.

3.3 Status of U-Trade Hub

Concerning Korea's U-Trade (e-Trade), the Ministry of Knowledge Economy is the main agent of trade automation for commercial trade sector, and the Bank of Korea and the Korea Telecommunications & Clearings Institute are the main agents for foreign exchange sector trade automation. The Ministry of Land, Infrastructure and Transport and the Korea Insurance Development Institute are the main agents of transportation and insurance sectors automation, respectively. Through organic cooperation with KTNET, the general trade automation project was conducted step-by-step. In this manner, general trade automation project systematically linking overall international trade sectors like customs clearance, commercial transactions, foreign exchange, logistics and insurance was completed [11].

Table 4. Details of the Next Generation U-Trade (e-Trade) System (2009-2012) [11]

Year	Task	Details of embodiment
2009	System linkage	• Linkage of information with MLIT (logistics), Korea Customs Service (customs clearance) and trade-related agencies (marketing) like local governments, KOTRA and Korea International Trade Association
	Pilot linkage of e-Nego	• Promote linkage targeting companies having plenty of exports/imports
	Supplementary expense payment system	• Unification of supplementary expense payment including freight and various commissions
	Airway bill distribution management system	• Build airway bill operating system that accounts for 54% (annually 3.6 million cases) of exports/imports
2010	System enhancement and user linkage	• E-B/L right register dualization, development and diffusion of user linkage system solution

2011	User linkage and overseas linkage	• Build FTA certificate of origin management system, global payment pilot project
2012	System enhancement and overseas linkage	• Enhancement of trade logistics system and linkage between overseas e-trade network and e-Nego pilot service

4. Proposals for e-Trading System Usingg Big Data

Big data can be utilized in the various industrial fields according to collection and analysis of information. McKinsey has introduced effect and examples of utilizing Big Data for each 5 field, such as medical, retail, manufacture, GPS information, and public service, and IBM has increased productivity and efficiency by applying Big Data to the entire industrial fields through campaign.

With utilizing Big Data, the following system model is proposed to solve the problems of the electric trading integration system and reinforce the function as the single window.

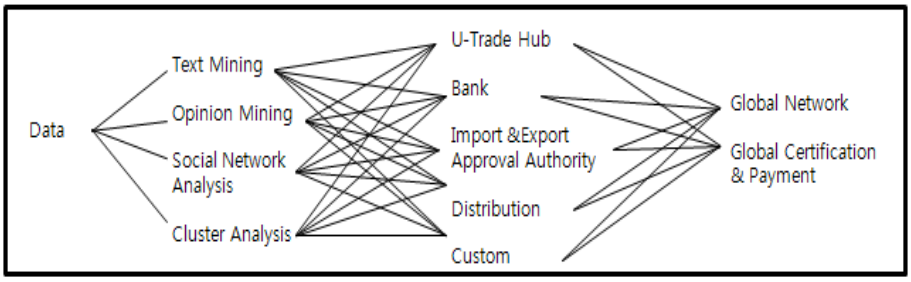

Figure 2. Single Window System

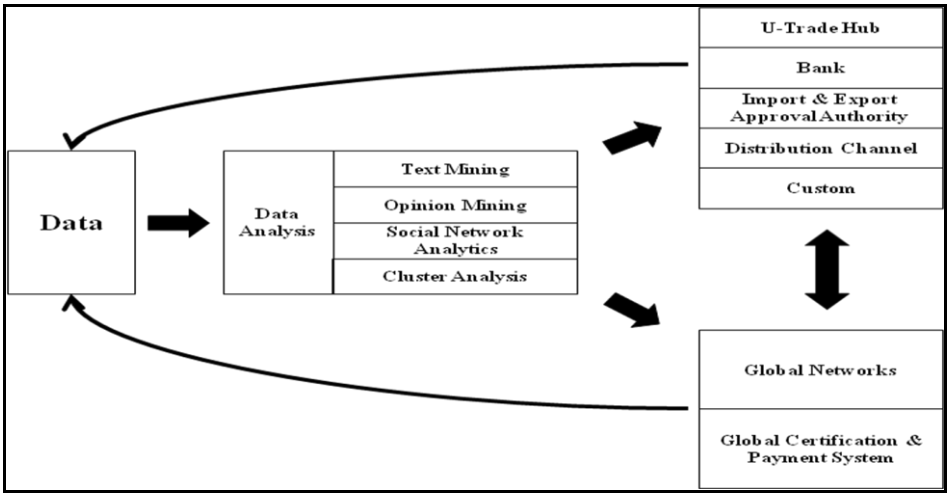

Figure 3. Single Window System detail

After analyzing the Big Data related the electric trading and trade to be variously collected through various analysis technique, the professional information is forwarded to each detail system by classifying with information appropriate for each sub-system of U-Trade Hub with arrangement of information through a consistent and standardized rules. Through a series of process, various information is though the collecting, arranging, classifying procedures and systemized. This systematized information is utilized as the detail information for the system

integration and service more efficient than in each specialized sub-system. In addition, the range of the system integration can be expanded to worldwide through sharing data with global web site and searching site.

5. Conclusions

Development of IT technology and wireless communication technology has caused the flow of the gigantic information data and the situation that analysis and classification of information are more important than production and collection. Development of new Big Data analysis technique will bring innovation in various fields utilizing Big Data, and suggest the way to deal with the vast volume of information that could not be dealt with so far. Big Data is a new paradigm throughout the entire industrial fields.

In this study, the Big Data, represented as non-structured, diversification, fast data generation, and complexity, has been intended to be analyzed and utilized with focus on the part , increase of efficiency of the system and integration of electric trading system.

Acknowledgements

This research was financially supported by Hansung University.

References

[1] IDC & EMC, Digital Universe Study 2011, **(2011)**.
[2] J. -s. Kim, "study for Big data technique and industry", Korea content association, vol. 10, no. 1, **(2012)**, pp. 34~40.
[3] WEF (World Economic Forum), "Big Data, Big Impact: New Possibilities for International Development", **(2012)**.
[4] J. -C. Koh, H. -U. Lee, J. -Y. Jeong, K. -S. Kim, "Correspondence Strategy for Big Data's New Customer Value and Creation of Business", Korea safety management &science, vol. 14, no. 4, **(2012)** December, pp. 229~238.
[5] Big data lab, "Big data for New future", National Information Society Agency, **(2013)**.
[6] SERI, "Big data: The core of Industry Change", **(2012)** May.
[7] McKinsey Global Institute, "Big Data: The next frontier for innovation, competition, and productivity", **(2011)**.
[8] S. -w. Cho, "Technology of Big Data Generation", KT Intelligent Knowledge Service, **(2011)**.
[9] C.-W. Ahn and S. -K. Hwang, "Big Data Technologies and Main Issues", Journal of the Korean Data & Information Science Society, vol. 30, no. 6, **(2012)** June, pp. 10~17.
[10] P. Warden, "Big Data Glossary", O'reilly Media, **(2011)**.
[11] C. -H. Kim, "Proposals for Developing Task of e-Trade and International Trade Customs in u-Trade Era", Korean Academy of International Commerce, vol. 25, no. 2, **(2010)** December, pp. 81~100.
[12] W. -S. Lee, "A Study on the Realization of Next Generation for e-Trade System in Korea", The Journal of Korea Research Society for Customs, vol. 13, no. 2, **(2012)** May, pp. 149~173.

Authors

Cheol-Woong Lee

Hansung University

Dept. of Digital Culture Technology & Contents

Interest Area: Multimedia, Electric Trading System, Big Data

e-mail: cyber@hansung.ac.kr

Sae-Hong Cho

Professor, Hansung University

Dept. of Multimedia Engineering

Interest Area: Multimedia, Virtual Reality, Big Data, Digital Contents

e-mail: chosh@hansung.ac.kr

Jong-Wook Kim

Assistant Professor, Sangmyung University

Dept. of Media Software

Interest Area: Web Data Mining, Information Retrieval, DB Systems

e-mail: jkim@smu.ac.kr

Dae-Hoon Hwang

Professor, Gachon University

Dept. of Computer Media

Interest Area: Multimedia, Big Data, Virtual Reality, Digital Contents

e-mail: hwangdh@gachon.ac.kr

