The Evolutionary Process of IT Concept Words: A Case Study on Bigdata

Rieko Kataoka, Naoshi Uchihira, Yasuo Ikawa
Japan Advanced Institute of Science and Technology, School of Knowledge Science, Ishikawa - Japan

***Abstract*--In information technology (IT), new concept words appear every few years and affect the business environment. In several cases, the core technologies and architectures have remained the same despite minor changes in concepts. For example, grid computing is the forerunner of cloud computing and bigdata is now regarded as a part of the Internet of Things (IoT). The trend in concept words reveals an evolutionary pattern.**

In this study, we applied a text mining approach to analyze all the articles published in several popular IT magazines in the period of 2002-2015. This analysis revealed a gap between cloud computing and bigdata in the evolutionary process of IT concept words. An evolutionary model was identified that reached cloud computing, indicating that another episode of evolution might start from bigdata. We focused our analysis on the evolution of previous major concept words and examined emerging concepts, which reveal a trend from a human-oriented to a machine-oriented world; the former world is characterized by advancements in social networking and the latter is based on advancements in artificial intelligence. As a result of this analysis, we can determine a turning point in concept evolution, i.e., the change from computing series to data centric. Understanding this phenomenon facilitates detailed interpretation of concept evolution.

I. INTRODUCTION

In information technology (IT), new concept words appear frequently and affect the business environment. In several cases, the core technologies and architectures have remained the same despite minor changes in concepts.

The most popular concept word of the year is often adopted as the main theme at trade shows. At every corners and booths, the products related to the theme are displayed and seminars, which have the theme in their title, are held as an annex to the trade show. Such concept words are determined for reasons pertaining to business marketing rather than technology. These signboards attract both companies and people to the exhibition. For example, “SaaS” (Software as a Service) was the trending word in 2008 and it was succeeded by “cloud computing” and “bigdata” in 2010 and 2012, respectively. Currently, “IoT” has become popular and is frequently used in articles on the Internet and in newspapers. These concept words quickly become popular and are spread beyond the IT industry to the general public.

Foster [1] compared multiple aspects of both cloud and grid computing and concluded that cloud computing was not a new concept; in fact, it was derived from the theory of grid computing and the theory’s relationships with previous technologies. According to the definition of cloud computing[1], one of the service models of the cloud includes SaaS. Therefore, cloud computing can be considered a mere rephrasing of SaaS because its concept existed previously. Bigdata and IoT are closely related to cloud computing. Although not a part of cloud computing, they came into existence as a result of the development of cloud computing.

In our study, we focused on the following questions. Why is a signboard selected at a given time and replaced shortly thereafter? How does the new concept differ from other concepts? What aspects of a concept have been inherited and developed? By verifying and systematizing the process starting from bigdata, in addition to an analysis from grid computing to cloud computing, does the pattern of concept development become clearer?

We presented our previous work at PICMET2014 [2]. In this previous study, we analyzed all the articles in IT magazines published between 2002 and 2012 and organized the process of concept evolution from technical words to service words. With the case study of cloud computing, we created an evolutionary model. As a subsequent case study, we expanded the period of analyzing article data up to September 2015. In this series of studies, we investigated the IT keyword’ transition at a wider scale and added more details to the evolutionary process of IT concepts.

The purpose of the present study is to identify the evolution of a concept from a technology to a service by observing the IT keyword transition in the case study of bigdata. Further, we consider the results of this investigation from the viewpoint of knowledge science. In other words, the evolution of a pattern of IT keyword can be detected by analyzing and organizing the evolutionary process from grid computing to cloud computing. After collecting and generalizing these patterns, we will be able to formulate a method to measure the possibility of new concepts by observing their evolution patterns in the future. Moreover, we will be able to evaluate whether a new concept can be developed further or will disappear without further evolution. We approached our research via analyzing the case studies of cloud computing and bigdata. For this, we applied text mining techniques to analyze articles from popular biweekly IT magazines in Japan (the background and structure of our case studies are described later in this paper)..

The term “concept” is defined as the generalized meaning of a phenomenon that comprehends, abstracts, and generalizes common items of a certain matter and plays a part in categorizing, identifying, and classifying actual events and relationships between such matters. Keywords are the manifestations of these major concepts.

IT keywords from all articles published between 2002 and 2015 were analyzed using our text mining application. In the

[1] The definition in Special Publication 800-145 NIST (the United States National Institute of Standards and Technology).

case of the IT industry, almost all concepts originated in the United States. The purpose of our research was not to study the technical growth of concepts but to identify the trends of a concept by tracking the transition of keywords. In this paper, we present our analysis results from text mining and its implications.

II. LITERATURE REVIEW

In this section, we review previous studies in related fields.

A. Text mining case studies

First, we review the literature focusing on topic transitions. There are many case studies that have analyzed documents and articles on the Internet using text mining.

Moriwaki [3] understood that topics change with time and tried to identify the consumer trend transition by the number of appearances of a topic. Moriwaki chronologically displayed the number of appearances of keywords and concluded that a topic transition could be detected only by tracking several words with high rates of appearance. Moriwaki offered a standard rule to judge the transition by number; however, this did not include the time scale of the transition.

Shirai [4] insisted that keywords should be selected in advance to obtain the trend information from text data and maintained the text mining environment by selecting keywords related to each purpose. He demonstrated a method to select specific words from nouns in the text.

For her selection of keywords, Okuwada [5] used the method suggested by Shirai and then sorted the keywords by category. Subsequently Okuwada suggested another method to analyze text data via free writing.

Yamamoto [6] mapped the relationships of the keywords in specific technology areas by applying text mining to patent and paper data. Yamamoto created a correlation between keyword groups and years via a mapping tool. However no relationship from year to year was determined.

In addition, there are multiple precession documents that describe the method of analyzing text data using text mining; however, there are no descriptions of IT keyword transitions for time scales of more than 10 years.

B. Text mining techniques

Generally text mining techniques are not sufficiently precise because text mining is different from other software applications and it is not possible to obtain an effective result without sufficient knowledge of text mining. This is why data scientists are needed to handle these techniques.

Nasukawa [7] understood that the selection of the technique in text mining applications is important and suggested a procedure to analyze datasets using text mining in three steps, namely, “the trial phase,” “the core phase,” and “the application phase.” However, his procedure shows only a conceptual framework, therefore, a detailed discussion is needed within this framework when a real case in analyzed. Considering these factors, we attempted to analyze text data using the framework offered by Nasukawa and accordingly adopted it for this case study.

Tsumoto [8] insisted that several new concepts appear every year but only a few of them remain depending on the number of times the concept was adopted in documents released to the public. Tsumoto analyzed technology trends on the basis of the frequency of appearance of technical terms in medical care research dissertations. Tseng [9] studied patent application documents and detected several important results by using text mining. Patent application documents can be configured using a selection rule because they comprise structured information. Data can be automatically collected using such selection rules even though this was a slow, manual operation until recently. By studying the emerging patterns and connections between patents, the criticality or degree of influence of each patent could be grasped.

The literature cited above describes the methodology that describes the procedure of fitting the application software of text mining to domain-structured data, such as patent application documents, and fetching information. This approach is not applicable to text data, which includes unstructured data.

C. Theory of concept trends

We focused on previous studies on the relationship between concepts. Shirota [10] affirmed that bigdata is a keyword that followed cloud computing and it arose from the popularization of cloud computing. According to Villegas [11], cloud computing originated from grid computing. Foster [1] compared multiple aspects of both cloud computing and grid computing and concluded that the cloud is not a new concept; rather, it is a result derived from the theory of grid computing and the theory’s relationships with previous technologies. Even though these concepts include several different aspects of business models and security, the vision, architecture, and basic technologies are the same. Youseff [12] reported that the cloud evolved by converting multiple technologies such as grid computing, SOA (service-oriented architecture), and virtualization and then subdivided cloud computing into five layers to discuss the meaning of its existence.

SaaS was included as one of three service models in the definition of the cloud computing by the National Institute of Standards and technology (NIST).

There are multiple studies that describe the relationship between two concept words. In this case study, we identify the transformation of concept words that occurred during a period of more than 10 years as evolution and confirm their relationship.

Several studies have focused on terminology trends from the viewpoint of management consultants. Giroux [13] expressed that from this viewpoint, concept labeling is not based on mere interest but on the change and progressive elements of society. Alvesson [14] noted that knowledge management could be sorted and classified by determining the type of concept because knowledge and management are considered to be combined concepts. Previously, knowledge and management were recognized as two different concepts.

Yamamoto [15] arranged commodity sorting by including intangible commodities along the following two axes: (1) the source that generates use and (2) the movement of the proprietary of the source that generates use. Our previous study that we presented at PICMET2014 verified the concepts of grid computing, SaaS, and cloud computing using Yamamoto's sorting method. The meaning of a phenomenon changes from a tangible commodity and the information produced by a technology to the service being an entity. In other words, the character of the concept moves from being technology-related to service-related.

In this section, we reviewed previous studies in related fields. To conclude our literature review, we note that there is a description of IT keyword transitions for time scales of more than ten years has not been given yet. In the next section, we identify the transformation of concepts as an evolutionary process of IT concept words and attempt to confirm the relationships between IT concept words.

III. ANALYSIS OF CONCEPT TRANSITION

Because of popularization on the Internet, various text data such as the technical information of patents and papers and the mail and twitter data that people post daily are spreading and expanding. Consequently, many documents or data cannot be used effectively and may be thrown away in a short period of time. Komoda [16] judged that it was impossible to examine this enormous amount of document data; however, we have been able to extract and effectively use information from these sources using text mining. Text mining is not a tool where an effective result can be provided by merely inputting document data. Text mining applications are unlike normal application software. The analysis result depends on how people direct the analysis and interpret the output.

In this section, we illustrate a technique to extract knowledge with time scale information from enormous amounts of document data by effectively using text mining. We develop a technique not only to obtain the relationship between keywords, but also to obtain the transition of concepts with time. Normally, the text mining application program can offer information on the time scale; however, a suitable method for each purpose is required to reach an effective conclusion. First, we will explain the analysis method and techniques needed to extract the transition of concepts from a huge volume of document data in an IT magazine. Then, we will show our analysis result via text mining for a case study.

A. The analysis method

The design of this case study is based on the technique of Yin [17]. Yin listed five reasons that can be used as bases for the case study in this paper. The form of the case study is reflected in the decisions concerning five components: the nature of the research questions, propositions, analyses of the data, logic plan that links the data to the propositions, and criteria for the analyses. Yin identified the theory that will be examined in the case study, and we have listed a counter -hypothesis that might account for the data. Five reasons why single case designs might be selected are their criticalness, extremeness, typicality, revelatory power, and longitudinal possibility.

In this study, we examined the concept word transition towards bigdata that occurred during a period of more than 10 years, which is sufficient for a longitudinal study. In addition, we described that these concept words have extremeness and typicality and designed a single case study on these bases.

For the analytical method used in the case study, we adopted a text mining methodology. The target text data for text mining analysis are the articles in six major IT magazines (listed in Table 1) published by Nikkei Business Publications, Inc. during the study period.

All articles published after 2002 are available as searchable, readable, and downloadable content on the web home page of Nikkei Premium. For our study, only the summary section of each article was downloaded as the object data. There were a total of 84,147 articles published between January 2002 and September 2015; this was a sufficient amount of data for text mining.

For text mining, we used an IBM Content Analytics version 3.0 (ICA3.0) released in June 2012. ICA3.0 can handle several hundred million cases or petabyte-class information and is, therefore, appropriate for treating bigdata. ICA3.0 is a software product that collects only the necessary information from unstructured information, such as writings or documents on the Internet, and can consistently sort and analyze this information.

TABLE 1. LIST OF PUBLICATIONS.

Name of publication	Description
Nikkei Computer	A comprehensive IT information magazine that offers detailed descriptions, columns, and content regarding the use of two-way communications in networks.
Nikkei Communication	An IT information magazine that supports decision making for telecommunications and networks.
Nikkei Network	An IT information magazine that provides information about network technology at a basic level.
Nikkei Personal Computer	A comprehensive IT information magazine that offers current information and skills.
Nikkei Systems	An IT information magazine that cultivates the skill of system development.
Nikkei Information Strategy	An IT information magazine that uses IT to innovate management.

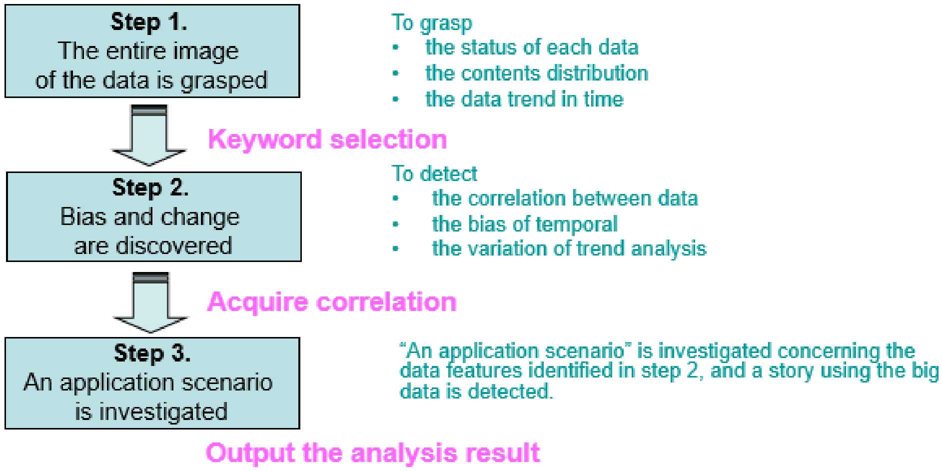

Figure 1. Flow of chart of the Nasukawa method.

To obtain effective results from the text mining process, the knowledge and techniques of the person implementing the text mining method are crucial. Without this ability or by just freely using the text mining application to analyze bigdata, obtaining meaningful analysis results is not possible. To solve this problem, we used for our analysis the Nasukawa method introduced by Nasukawa [7], who is an ICA3.0 developer and researcher. The Nasukawa method originated from the knowledge and experience that Nasukawa obtained from actual projects. He described his method in three steps that are designed to obtain an effective result using text mining (Fig.1).

The goal of Step 1 is to grasp the entire image of the data. We aim to grasp the features of the distribution by analyzing various output images produced by ICA3.0. The goal of Step 2 is to discover the bias and change and to execute the content bias, which is detected by correlation analysis; the time wise bias, which is detected by deviation analysis; and the trend analysis, which is detected by change analysis. If something is observed in this phase, further enhancements in the object data are needed to clearly grasp the features. In Step 3, an application scenario concerning the data features identified in Step 2 is investigated and a story using bigdata is detected.

1) Step 1

By making full use of the output form of the analysis from the ICA3.0 application, reading documents one by one using a list of the documents, and changing the facet items (e.g., nouns and verbs) and chronological order memory, we confirmed that the distribution kept the perspective of the data under control. The mining environment should be maintained according to the purpose of the analysis if it shows some directionality during this step..

Nouns and verbs include words that are not useful to grasp of the meaning of a sentence. We exclude such words because they are unnecessary information even if they are put on a map. In this study, we followed the method suggested by Moriwaki [3] and selected the keywords to be analyzed. Then, we used the keyword selection phase in Step 1 as a pre-transaction for Step 2. In ICA3.0, there is a function that performs text mining focused on a term; this is made possible by registering a specific keyword with "a dictionary" to gather the synonymous expressions of the keyword. We decided to use this function for keyword selection.

To enroll keywords in the dictionary for this study, we referenced a page of an article published in the Nikkei Computer magazine [18] that listed a glossary of keywords one should know. Nikkei Computer is a comprehensive IT magazine that has represented the IT industry for several decades. The selection criteria of the keywords from this page were as follows: (1) they should succinctly explain a technical term, (2) they could be checked as technical trends, and (3) whether they have been proposed as important words for a the phenomenon. If there is a keyword that appears multiple times in the same article, the count remains 1. Eventually, approximately 500 keywords were selected from 2002 to 2015 and were manually registered into ICA3.0 as the concepts to be analyzed.

2) Step 2

In this stage, we performed deviation analysis (chronological order deviation) for all the keywords registered in Step 1 and performed a detailed confirmation of the deflection and relationship change for each keyword. According to Komoda [16], it is important in text mining to read the implications of the increasing/decreasing and deflection trends of the distribution apart from just counting the number of appearances of the words; i.e., the detection of a characteristic by comparison is the fundamental value of text mining. As an example of the deviation analysis, we show the cloud computing-associated part in Fig. 2, which shows the number of appearances of the words, cloud computing, public cloud, and private cloud in each year of the study period. We performed similar confirmations for all the keywords.

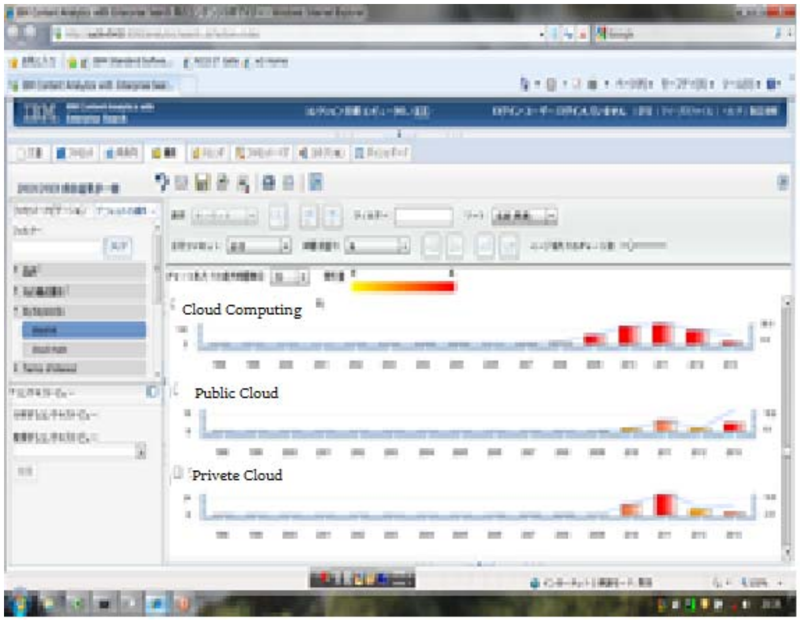

Figure 2. Deviation analysis for the example of cloud computing.

3) Step3

These keywords need several years to appear and do not necessarily appear together. This phenomenon occurs because the keywords appeared as one as they were related during these years. As a scenario in this study, we tried to convey this relationship as a diagram that expresses the change in the concept using the correlation level that emerged between the keywords in the standard described in Step 2. Having completed Steps 1 and 2 using the Nasukawa method, we decided to execute the connection analysis.

The connection analysis was performed according to the ICA3.0 Algorism [19] and was based on the correlation values, which automatically measure the correlations between the keywords. The correlation values have no direct relationship with the frequency but they reveal an aspect of the strength of the relationships between the keywords.

The correlation algorithm operates as follows. A correlation value is defined for two document populations A and B. D indicates the entire document population, and # indicates the number of documents. The left-hand and right-hand sides are equal.

$$\frac{\#(A \cap B)/\#A}{\#B/\#D} = \frac{\#(A \cap B)/\#D}{(\#A/\#D)(\#B/\#D)}$$

Example document populations include the following:

A = { “Commercial Product”: documents of the category keyword “PC”} and

B = { “Noun…want”: documents of the category keyword “Manual…acquire…want”},

Given these example A and B values, the left-hand side of the above equation is as follows:

The ratio of “Need the manual about PC”

The ratio of “Need the manual”

If 5% of people need the manual in all documents and 20% of people need the PC manual in all PC documents, the correlation value between “PC” and “Manual…acquire…want” is high, as shown in Fig. 3.

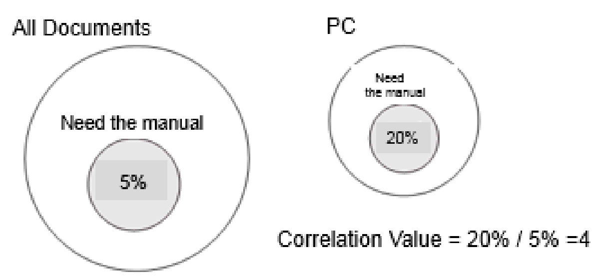

Figure 3. Illustration of the ICA3.0 correlation value.

The description above shows the logic involved in this system; however, with ICA3.0, the interval estimation is performed according to the reliability of the data. In contrast to point estimation, which involves only a single number, interval estimation is the use of sample data to calculate an interval of possible (or probable) values of an unknown population parameter, in contrast to point estimation, which is a single number.

In this paper, we adopted the time factor, which was the final output image of the connection analysis from the deviation analysis in Step 2. Using the correlation between each keyword of the concept within the time series, a map can be described, as illustrated in Fig. 4. The elliptical size of each keyword corresponds to its frequency. The x-axis indicates the time scale and shows the peak of the trend, whereas the y-axis has no meaning. A value of 1.0 indicates

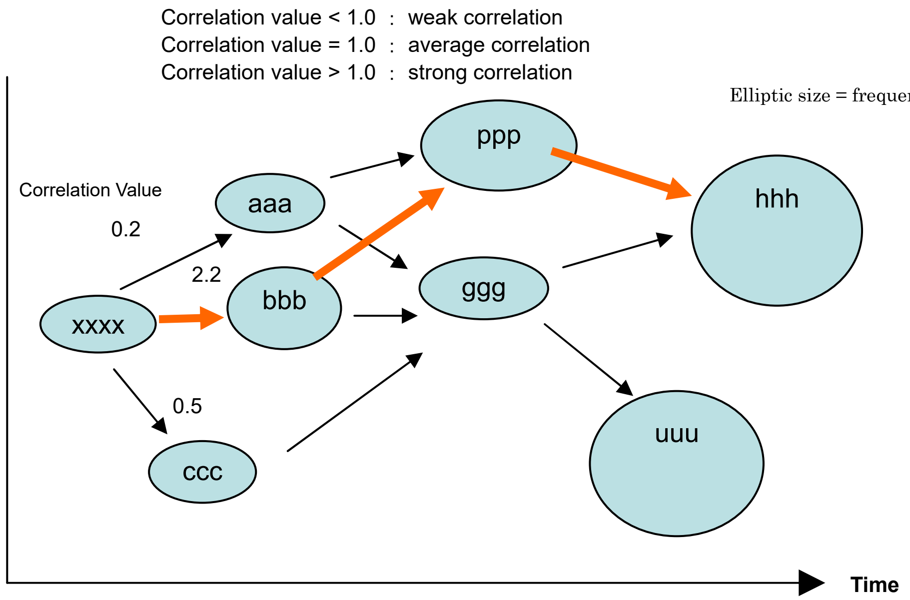

Figure 4. Sample output image of an ICA3.0 connection analysis.

an even correlation, whereas a value greater than 1.0 indicates a strong correlation, and a value less than 1.0 indicates a weak correlation or no correlation. The change in a given keyword can be observed by following the keyword that has the strongest correlation in the time series. The x-axis shows the year, whereas the y-axis has no meaning.

4) Output Result

Using the technique described above, we analyzed the result of the change in the concept trend in the IT industry (Fig. 5). Using the data from approximately 85,000 articles, we performed text mining based on the keywords that we registered in the dictionary; the output image is the connection analysis (Fig. 5). The x-axis shows the year in which the keyword peaked. We had already analyzed the concept evolution from 2002 to 2012 in a previous study. However, we did a subsequent comprehensive study by expanding the period of analyzing article data up to September 2015. The reason why there are less than 500 plots in Fig. 5 is that only the cases that had a high relationship between the keywords appear on the chart.

In Fig. 5, the correlation values from 2.0 to 4.0, 4.1 to 6.0, 6.1 to 15.0, and 15.1 and above are shown in yellow, orange, pink, and red, respectively.

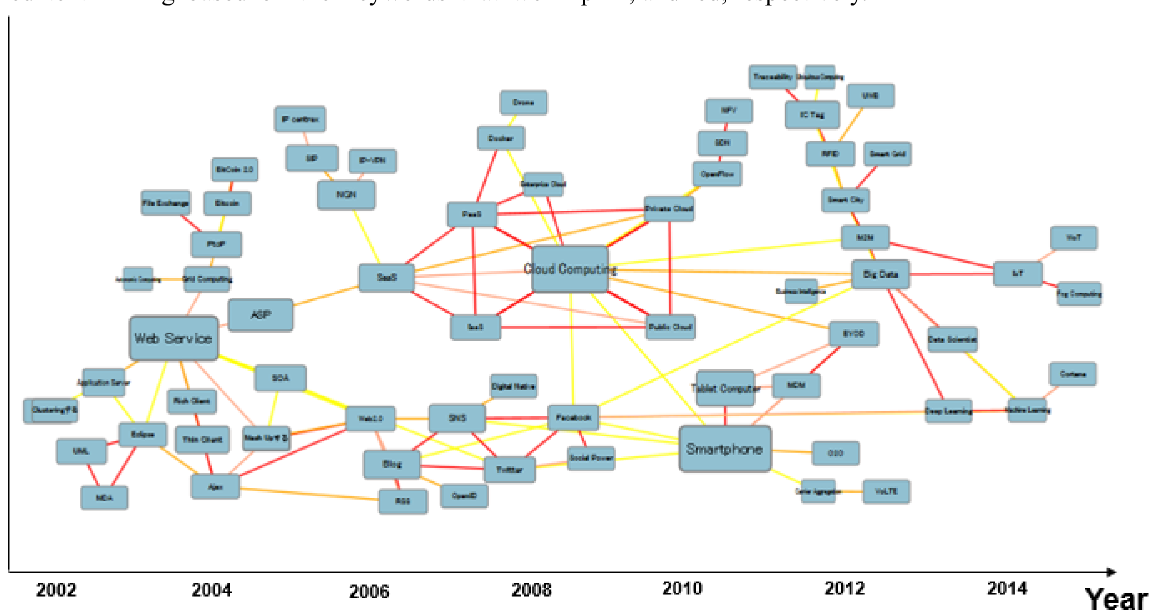

Figure 5. Connection analysis results for data spanning from 2002 to 2015.

B. A analysis of the output result

In this study, we applied a text mining approach to analyze all the articles published in six well-known IT magazines during the period of 2002-2015, as described in Section 3.1. The keywords between 2002 and 2015 are shown in Fig. 5 and 7. The concepts related to bigdata in 2012 were data scientist and M2M (machine to machine). The keyword IoT, which appeared in 2014, developed into "xx of Things" or "Internet of xx" in 2015.

In a previous study, we developed the concept evolution model, as shown in Fig. 6. We defined that the evolution comprises three stages; gathering, development, and specialization. We found that grid computing has a strong correlation with hardware and software keywords, such as web service and application service provider, which became service components in the next phase from 2002 to 2005. This phase is the congregating phase of components for service businesses.

Following the trend of SaaS, the keywords PaaS (Platform as a Service) and IaaS (Infrastructure as a Service) appeared in 2007. These are coined words devised for the IT server layer, application, middleware, and infrastructure. By applying this phenomenon to the structure of an IT server unit, keywords accomplished development in the vertical direction. In 2010, SaaS developed into cloud computing and then two smaller clouds, namely private and public clouds, were created. From the given articles, the cloud terminology became specialized in various industries, including a self-governing community cloud for government and municipal offices, medical clouds in hospitals, and other similar clouds for various industries.

Simultaneously, when we observed the developmental status of cloud computing via the output of the text mining analysis, we found that Social Network Service (SNS) led this concept evolution. SNS itself had developed with mobile enhancements.

The phenomenon in Fig. 7 is similar to that where some concepts appeared around grid computing and then PaaS and IaaS formed around SaaS. This means that the same evolutional pattern started from bigdata again after the specialization of cloud computing. Bigdata is in the gathering stage and IoT may be in the development stage ; i.e., the evolutional process of gathering, development, and specialization can also be applied to the concepts of bigdata.

According to this analysis, an evolutionary model has been recognized that arrived at cloud computing, and may start another evolution from bigdata. Our analysis focused on the evolution of previous major concept words and examined emerging concepts, which indicate a trend from the human-oriented to the machine-oriented world. The human-oriented world features the advancement of social networking and the machine-oriented world is based on advancements in artificial intelligence.

We need more time to further analize the concept evolution after the IoT. The next concept word following IoT is expected to appear within a few years and will be evolving in the specialization stage.

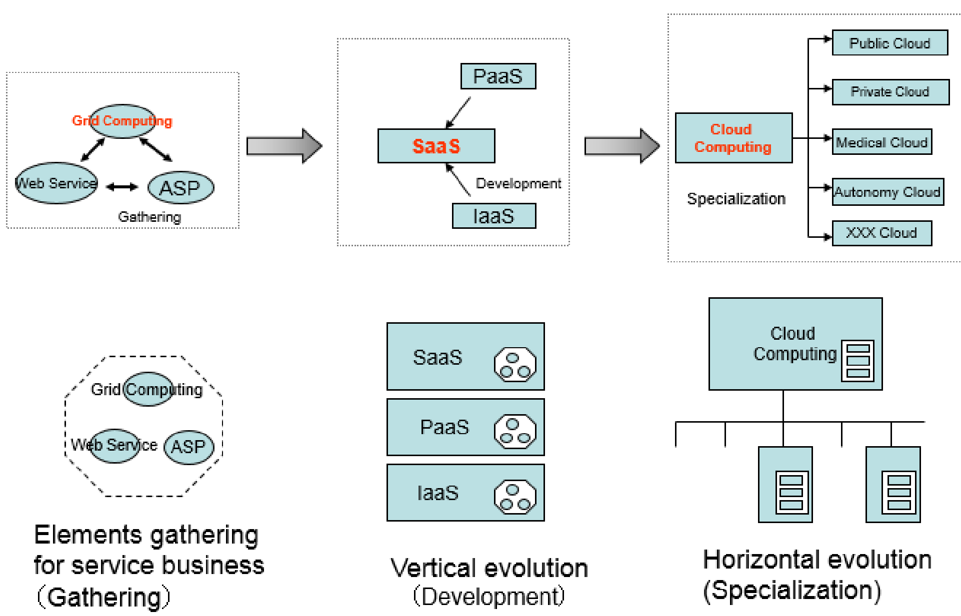

Figure 6. Evolutionary model of cloud computing[2]

[2] Kataoka, R., Ikawa, Y. and Uchihira, N. "The evolutionary process from a technology concept to a service concept –a cloud computing case study–," PICMET2014, p.1855-1865.

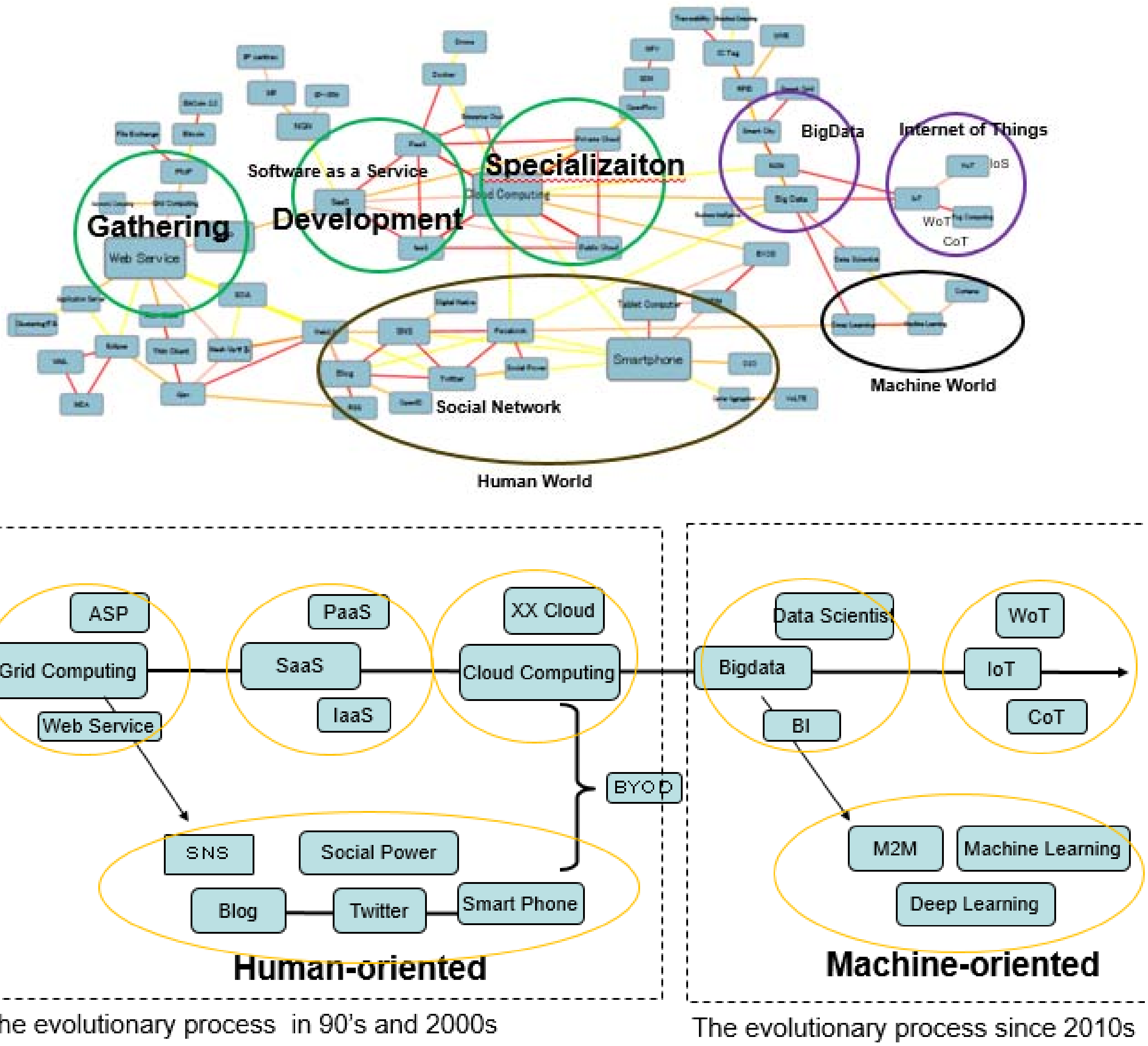

Figure 7. Evolutionary process of concepts.

IV. INVESTIGATIONS

In a further analysis, we found that IT concepts developed from cloud computing to big data and IoT. We applied a text mining approach to analyze all the articles published in several well-known IT magazines during the period of 2002-2015. According to this analysis, an evolutionary model was identified that cumulated at cloud computing; this indicates that another evolution might start from bigdata. We focused our analysis on the evolution of previous major concept words and looked at emerging concepts, which indicate a trend from the human-oriented to the machine-oriented world. We found that the human-oriented world is marked by the advancement in social networking and that the machine-oriented world is based on advancements in artificial intelligence. Consequently, we will describe a turning point in concept evolution, i.e., the change from computing series to data centric. Understanding this phenomenon allows detailed interpretation of concept evolution.

In a previous paper [2], we sorted the commodities of grid computing, SaaS, and cloud computing according to the method suggested by Yamamoto [15], which is summarized in Fig. 8. We verified that the meaning of a phenomenon changes from a tangible commodity and information to a nontangible commodity including service. In other words, the character of the concept moves from being technology-related to service-related.

Yamamoto (2010) Sort of commodity

	The source which generates use is the physical commodity	The source which generates use is the non-physical commodity
There is moving the proprietary	Tangible commodity	Information
There is NO moving the proprietary	Utilization right of tangible commodity	Service

Type of commodity		Example	Grid Computing	SaaS	Cloud Computing
The source which generates use is the physical commodity	There is moving the proprietary	Tangible commodity	Server, Storage, etc.	None	None
	There is NO moving the proprietary	Utilization right of tangible commodity	None	Utilization right of Disc space (Monthly rate)	None
The source which generates use is the non-physical commodity	There is moving the proprietary	Information	Application SW	None	None
	There is NO moving the proprietary	Service	None	Utilization right of application SW (Monthly rating)	Utilization right of computing (One minute unit rating)

Figure 8. The commodities of Grid, SaaS, and cloud computing.

From bigdata, we identified the physical commodity as data and applied the sorting method shown in Fig. 9 to highlight the differences between each keyword. The "X" in the figure will be replaced when the next-generation keyword appears in the future. According to a Mitsubishi Research Institute Inc. report [20], the difference between bigdata and IoT depends on value creation. Moreover, IoT provides services made from data analysis whereas bigdata provides analysis results as information.

Bigdata is a new system treating larger-capacity data; this has been possible because of the spread of the Internet and the current evolution of IT technology. Bigdata is characterized not only by its volume but also by various kinds of data, including unstructured data such as text data, sound data, and movie data. In the generation of bigdata, data can be transferred through personal devices using M2M technology. Such transfers have become easier with the development of high-speed mobile network infrastructures and the advancement of cloud computing.

In addition, by connecting things with the Internet, the generation of IoT creates a new value that is quite different from the information gathering of bigdata. Sensors can be installed in smart phones, electrical home devices, vehicles, buildings, shops and factories and can be automatically connected to generate valuable information. For this, submission of data is not necessary and only an environment for analytics with high-speed communication is required. Connecting things is not a new concept. The difference between IoT and bigdata is that IoT has two-way communication and creates value for users; only IoT can offer knowledge as a service. In other words, the service provided by the IoT generation is that humans do not need to manage every minor detail; instead, the things can autonomously determine the best solutions.

Type of commodity		Example	BigData (BYOD,M2M)	IoT (WoT, IoC, IoE･･･)	XXX
The source with generates use is the physical commodity (data)	There is moving the proprietary	Tangible commodity	Data	None	None
	There is NO moving the proprietary	Utilization right of tangible commodity	None	Utilization right of data	None
The source which generates use is the non-physical commodity	There is moving the proprietary	Information	The analysis result of data	None	None
	There is NO moving the proprietary	Service	None	Utilization right of the analysis result of data	Utilization right of XXX ?

Figure 9. Commodities of bigdata and IoT.

V. CONCLUSIONS

In any industry, it is important for companies to differentiate themselves from their rivals to secure their competitive predominance. Therefore, originality is required from each company. In the industry, new concept words are a technique to achieve differentiation.

In addition to analyzing the evolution pattern of cloud computing from a previous study, we analyzed the evolution of concepts in this study. Because the environment of concept evolution has switched from SNS to machine learning, we assumed that one evolution pattern ended at cloud computing and a new evolution started from bigdata. However, a detailed analysis of the keywords shows that the evolution pattern comprises gathering, development, and specialization and is moved towards a service concept from a technology concept.

Cisco Systems [21] said, "We have discovered the possibility of IoT power and have arrived at the entrance of the generation of a new Internet economy. In future, IoT will develop from a horizontally unified IoT into an independent verticalization-type IoT on the basis of a common infrastructure called the Internet. In this environment, a new value will be created and innovation will happen." They predicted the advancement of IoT and called it IoE (Internet of Everything). However, we may need to wait for a new concept to achieve the independent verticalization-type IoT.

The theoretical implication of this study is that the evolutionary process of a concept in the IT industry has been demonstrated and modeled with a focus on cloud computing by analyzing data using text mining techniques.

The practical implication is that planning business strategies and actions by referring to the evolutionary case of the cloud computing concept may be possible. In addition, a few appealing concepts might attract other researchers in such a way that the method leads to implementations in the industry.

The concept evolution model can be further generalized by collecting data on the further developing business environmental situation. By repeating this study to observe the short-term changes in keywords, the status of the developmental stage of a service business and its possibility for future evolution could be determined.

ACKNOWLEDGEMENTS

It is gratitude to Dr. Tetsuya Nasukawa and Dr. Issei Yoshida, IBM Research of Tokyo, whom guidance it receives concerning text mining technique and to everyone of knowledge management course in JAIST who offers many advices when this research is advanced. The authors would like to thank Enago (http://www.enago.jp) for the English language review.

REFERENCES

[1] Foster, I. et.al. "Cloud Computing and Grid Computing 360-Degree Compared," *Grid Computing Environments Workshop*, Conference Publications, 2008.

[2] Kataoka, R., Ikawa, Y. and Uchihira, N. "The evolutionary process from a technology concept to a service concept –a cloud computing case study–," PICMET2014, p.1855-1865.

[3] Moriwaki, H.; "Text mining technology and its case study of big data", *Unisys Technology review* Vol.115, 2013, (in Japanese).

[4] Shirai, Y. et.al. "The text mining environment for information trend survey," *MRI Technical Report* 5.14, p.110-123, 2009, (in Japanese).

[5] Okuwada, K. et.al. "Trial to extract a meaningful opinion from the free description according to the field in a technology policy automatically," *The Japan Society for Research Policy and Innovation Management, Annual General Meeting Proceedings*, Vol.22, 69205, 2007.

[6] Yamamoto, T.; "The effectiveness of the text mining in the matching-related analysis of the industrial cooperation," *Science and technology of the information*, Vol.59. 6. 2009, (in Japanese).

[7] Nasukawa, T.; "Text Mining−Technology for Usage", *Tokyo Denki University Punishment Book*, 2006, (in Japanese).

[8] Tsumoto, S. and Abe, H. "Comparing Similarity of Concepts Identified by Temporal Patterns of Terms in Biomedical Research Documents," *Rough Sets and Knowledge Technology, 7th International Conference, Proceedings*, pp.232-241. 2012.

[9] Tseng, Y. et.al. "Text mining techniques for patent analysis," *Information processing and management 43*, p.1216–1247. 2007.

[10] Shirota, M.; "The impact of Big data, " Toyo Keizai Inc. 2012, (in Japanese).

[11] Villegas, D.; "The Role of Grid Computing Technologies in Cloud Computing," *Handbook of Cloud Computing*, Part 1, pp.183-218. 2010.

[12] Youseff, L., Butrico, M. and Silva, D. "Toward a Unified Ontology of Cloud Computing," *Grid Computing Environments Workshop*, 2008. GCE '08

[13] Giroux, H.; "It Was Such a Handy Term; Management Fashions and Pragmatic Ambiguity," *Journal of Management Studies* 43, 6 September. 2006.

[14] Alvesson, M. et al. "Odd Couple: Making sense of the curious concept of knowledge management," *Journal of Management Studies* 28, 7 November. 2001.

[15] Yamamoto, S.; "New Edition−Service Quality," Chikura Book, 2010, (in Japanese).

[16] Komoda, F. and Nasukawa, T. "Text mining as technical strategy," Chuo Keizai Inc. 2014, (in Japanese).

[17] Yin, R.; "Case Study Research: Design and Methods (4th Edition)," Sage Publications. 2008.

[18] Nikkei Computer Editing Bureau; Nikkei Computer, Japan: Nikkei BP Publishing Company, Edition in 2002-2015, (in Japanese).

[19] Introducing OmniFind Analytics Edition; "Customizing Text Analytics," An IBM Redbooks publication, http://www.redbooks.ibm.com/abstracts/sg247568.html, January 31th, 2016.

[20] Mitsubishi Research Institute Inc.; "All about IoT", Nikkei Publishing Inc., 2015, (in Japanese).

[21] Cisco Systems IOT Incubation Laboratory, "The impact of Internet of Everything", *Cisco Systems Series, Next Publishing*, on demand paperback. 2013.