Received: 12 February 2018 | Revised: 17 April 2018 | Accepted: 27 May 2018
DOI: 10.4218/etrij.2018-0059

ORIGINAL ARTICLE

WILEY ETRI Journal

Exploring the dynamic knowledge structure of studies on the Internet of things: Keyword analysis

Young Seog Yoon[1,2] | Hangjung Zo[1] | Munkee Choi[1] | Donghyun Lee[3] | Hyun-woo Lee[4]

[1]School of Business and Technology Management, Korea Advanced Science and Technology, Daejeon, Rep. of Korea.
[2]Hyper-connected Communication Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea.
[3]Department of Business Administration, Korea Polytechnic University, Siheung, Rep. of Korea.
[4]Broadcasting & Media Research Laboratory, Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea.

Correspondence
Hangjung Zo, School of Business and Technology Management, Korea Advanced Science and Technology, Daejeon, Rep. of Korea.
Email: joezo@kaist.edu

Funding information
ICT R&D program of MSIP/IITP, Grant/Award Number: 2015-0-00533

A wide range of studies in various disciplines has focused on the Internet of Things (IoT) and cyber-physical systems (CPS). However, it is necessary to summarize the current status and to establish future directions because each study has its own individual goals independent of the completion of all IoT applications. The absence of a comprehensive understanding of IoT and CPS has disrupted an efficient resource allocation. To assess changes in the knowledge structure and emerging technologies, this study explores the dynamic research trends in IoT by analyzing bibliographic data. We retrieved 54,237 keywords in 12,600 IoT studies from the Scopus database, and conducted keyword frequency, co-occurrence, and growth-rate analyses. The analysis results reveal how IoT technologies have been developed and how they are connected to each other. We also show that such technologies have diverged and converged simultaneously, and that the emerging keywords of trust, smart home, cloud, authentication, context-aware, and big data have been extracted. We also unveil that the CPS is directly involved in network, security, management, cloud, big data, system, industry, architecture, and the Internet.

KEYWORDS
IoT, keyword, co-occurrence, knowledge structure, emerging technology

1 | INTRODUCTION

Since the introduction of the Internet of Things (IoT) [1] and cyber-physical systems (CPS) [2], numerous studies have been conducted in these fields. However, because each study has its own individual goals, it is necessary to provide an overview of the current status and to establish a future direction from a comprehensive viewpoint. In addition, IoT does not indicate a specific technology, but is a novel paradigm consisting of a set of technologies [3]. As a result, a holistic understanding of the IoT and CPS has been limited [3,4]. For researchers and policymakers to allocate resources more efficiently, it is necessary to investigate IoT and CPS from a more comprehensive perspective.

To fill in such gaps, previous studies have conducted surveys on IoT and CPS to assess their evolutionary path and expected research topics [3–7]. However, most of these studies have adopted a narrative review, which overviews and summarizes prior studies from the author's subjective point of view [8]. Although their insights provide significant implications, it may be beneficial to assess the knowledge structure of IoT and CPS based on scientific evidence.

This study adopted quantitative methods of keyword frequency, co-occurrence, and growth-rate analyses. Based on such analyses, to trace the changes in IoT research and extract emerging technologies for IoT, this study explores the dynamic knowledge structure of IoT. Emerging technology is defined as "a radically novel and relatively fast growing technology characterized by a certain degree of coherence persisting over time and with the potential to exert a considerable impact on the socio-economic domain [9]." This is not necessarily new because it may be originated from another domain [10], and may cause a radical change in business, industry, or society [10].

The purpose of this study is to visualize and unveil the dynamic knowledge structure generated through academic research. Our analysis results may be helpful to understand current state-of-the-art developments and to discuss future research directions. In addition, we can also confirm a set of emerging technologies. Specifically, this study addresses the following research questions:

1) How have emerging IoT technologies evolved and converged?
2) Which emerging technologies have been developed and adopted for IoT?
3) Which research topics are promising for future application?

2 | LITERATURE REVIEW

2.1 | Definitions of IoT and CPS

Although we do not intend to compare IoT and CPS, it is necessary to define them clearly. As discussed in [3] and [4], the definitions and concepts of IoT and CPS remain vague and overlap because: 1) they are used as trendy brands, 2) they are composed of different sets of technologies, 3) they share a common vision, and 4) they are still rapidly progressing.

Although the previous concept of IoT mainly focused on identifying and monitoring connected things, in recent times, the focus has been on the control of physical things and systems [7]. As shown in Table 1, the IoT is mainly interested in networking and interconnection among different things, whereas CPS is concerned with exchanging information and controlling physical things [11]. However, the distinction between them has become blurred over time [7–12]. Intuitively, the control of physical things is premised on connections between them. Accordingly, the current study mainly focuses on analyzing IoT because their boundaries almost overlap.

TABLE 1 Definitions of IoT and CPS

Domain	Definition	Source
IoT	Networked interconnection of all objects, which are often equipped with ubiquitous intelligence	[13]
	A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies	[14]
	A "thing" in IoT indicates a physical or virtual object that connects to the Internet, and which has the ability to communicate with human users or other objects	[15]
	An open and comprehensive network of intelligent objects that have the capacity to auto-organize	[16]
	A worldwide network of interconnected objects that are uniquely addressable, based on standard communication protocols	[17]
	A conceptual framework that leverages the availability of heterogeneous devices and interconnection solutions, as well as augmented physical objects providing a shared information base on a global scale	[3]
CPS	Physical and engineered systems whose operations are monitored, coordinated, controlled, and integrated through a computing and communicating core	[18]
	Integrations of computations with physical processes	[19]
	Physical, biological, and engineered systems whose operations are integrated, monitored, and/or controlled using a computational core	[11]
	A new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities	[2]
	Transformative technologies for managing interconnected systems between its physical assets and computational capabilities	[20]

2.2 | Literature survey of previous IoT and CPS studies

Several studies have discussed the history, emerging technologies, research trends, research clusters, and future research directions by investigating previous studies on IoT and CPS. However, most of them have relied on a narrative review, which is not a quantitative analysis, as a methodology.

Although their findings provide remarkable implications, their conclusions do not perfectly coincide with each other. Table 2 summarizes the IoT and CPS elements in previous works. Although most of the IoT elements identified in previous studies are identical, there are some differences among them. For example, some of them consider technologies for the analysis of data as key elements for IoT [4,16–22],

TABLE 2 Proposed IoT/CPS elements derived from previous studies

ResearchMethod	IoT / CPS elements (Enabling technologies)	Source|number of articles reviewed
Narrative review	Radiofrequency identification (RFID), UID, Spimes, Smart items, NFC, objects, WISP, WSA, connectivity, communication, things, IPSO, Internet, web of things, smart semantic middleware, semantic technologies, reasoning over data, semantic execution environments	[4]|95
Narrative review	RFID, IP, electronic product codes (EPCs), barcodes, Wi-Fi, Bluetooth, ZigBee, NFC, actuators, wireless sensor networks (WSNs), artificial intelligence (AI)	[16]|27
Systematic review	Identification, sensing, communication, computation, services, semantics	[6]|195
Narrative review	RFID, WSN, addressing schemes (IPv6, uniform resource name (URN)), data storage and analytics (AI, machine learning), visualization	[21]|72
Narrative review	RFID, WSN, cloud computing, IMT-advanced, Bluetooth, Wi-Fi, Nano devices, MEMS, Li-Fi, BiDi	[23]|50
Narrative review	Sensor nodes, actuator nodes, communication networks, cyber systems	[7]|132
Narrative review	Application, architecture, sensing, data management, computation, communication, security, control/ actuation	[22]|80

whereas others do not exclusively include them. Rather, they place more weight on networking elements [7–23].

These different perspectives originate from the fact that their individual conclusions are derived from the use of limited data and each author's insights. However, inconsistency is not a scientific problem because their findings aim to improve and expand knowledge. In this manner, current data-driven studies can contribute to existing knowledge.

2.3 | Narrative, systematic review, and keyword analysis

Generally, researchers have adopted two types of methodologies (narrative and systemic review) to overview the dynamics of a research stream. A narrative review covers a wide range of issues, but it hardly state or justify any rules about the searching results [24]. They describe and summarize existing findings from the literature without statistical or quantitative validation. Meanwhile, a systematic review uses explicit methods to methodically search, critically appraise, and synthesize the available literature on a specific issue [24]. A systematic review can contain a statistical method (eg, a meta-analysis).

However, both methods suffer from selection bias and subjective interpretation [25]. Recently, keyword analysis has been widely adopted to overcome these problems because it is based on quantitative and statistical methods by analyzing a significant number of studies [26–32].

3 | RESEARCH METHODS

3.1 | Research framework

This study investigates the knowledge structure of IoT in order to trace the dynamic research trends, extract emerging technologies, and identify convergence and divergence of research topics.

For these purposes, we conducted keyword frequency, keyword mapping, and keyword growth-rate analyses. A keyword frequency analysis provides useful clues for detecting the essentials of the knowledge structure. Keyword mapping analysis has been widely used as a suitable way of investigating dynamic changes from a holistic perspective [26,27,30,33,34], whereas a keyword growth rate analysis is an appropriate method for detecting dynamic changes of specific emerging keywords [30–35]. The research framework used for this study is shown in Figure 1.

3.2 | Data collection and refinement

We retrieved bibliographic data from the Scopus database, which is the world's largest multidisciplinary database of scientific studies covering 20,000 peer-reviewed journals [36–39]. The collection of data from multiple databases causes a data redundancy problem because there are a considerable number of duplicate records between databases.

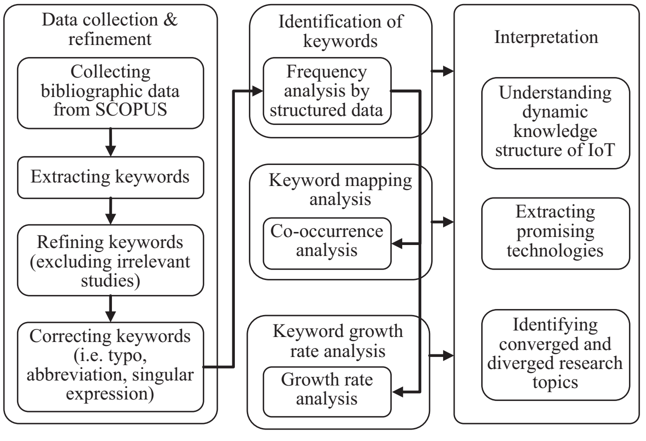

FIGURE 1 Research framework

Indeed, it is necessary to conduct data cleansing, which is an extremely time- and labor-intensive task. For this reason, previous studies have relied on a single database to collect and process structured data from articles [37,39,40].

Initially, we collected bibliographic data (keywords, authors, and year of publication) on 17,616 academic studies that were published from 2000 to 2016, which include one of the following keywords: “IoT,” “Internet of Things,” “Internet of Thing,” “WoT,” “Web of Thing(s),” “Internet of Everything,” “IoE,” and other relevant terms (eg, WoTs). Because the underlying philosophy of IoT, IoE, and WoT are almost identical [41], we considered them as unified terms of the IoT.

After that, two IT experts refined the collected data because the term IoT is also used in other research domains. For example, it also represents “Intensive Outpatient Treatment” or “Institution-based Occupation Therapy” in medical studies. In addition, some papers were redundantly indexed during the crawling process. After eliminating irrelevant studies and duplicate articles, we conducted a keyword revision process by modifying and correcting the collected keywords into unified terms when considering typos (eg, thign → thing), abbreviations (Internet of Thing → IoT), and singular expressions (Things → Thing).

Other than this revision process, we did not modify the concept of the keywords described by the authors in order to reflect their original intention without distortion or debate [26]. The refinement process was conducted using R scripts. Finally, there remained 54,237 keywords in 12,600 papers. Table 3 summarizes our dataset collected from the Scopus database.

3.3 | Analysis methods

3.3.1 | Keywords frequency analysis

Keyword frequency analysis is a part of the field of descriptive statistics, and measures the frequency of each keyword in each document. This method is considered as the basis of other methods, and thus various data-mining techniques rely on this keyword frequency to define the concept of core topics [42]. However, it cannot extract hidden or latent topics because authors tend to use synonyms, and a simple count of each word is not identical to a research stream. In addition, we cannot identify the structural relationship between keywords. Thus, the following analyses are required.

TABLE 3 Summary of collected data

Attributes	Description
Source	Scopus database
Keywords	Internet of Things and relevant terms (eg, IoT, IoTs, WoT, and IoE)
Types of literature	Scientific journals, books, and conference proceedings
Retrieved bibliographic data	**Keywords, year of publication**, authors, title, volume, issue no, pages, affrications, and other data (***items in bold were used in the analyses***)
Year of publication	From Jan. 1, 2000 to Nov. 3, 2016
Number of collected studies	Initial collection: 17,161→ after refinement: 12,600

3.3.2 | Co-occurrence analysis

A co-occurrence (co-word) analysis is a method that describes the network of interactions among keywords [43]. This analysis allows researchers to identify research directions for a specific research topic based on relationships among keywords [26,33,43–45]. Indexed keywords are valuable pieces of information that represent fundamental concepts in a study because the authors carefully list up the selective keywords for the benefit of readers. The co-occurrences of words in different parts of different articles were counted and analyzed [43]. This analysis calculated the co-occurrence frequency of each word pair, and obtained correlations and similarities between words [33].

To measure the similarities, we adopted the association strength, which is used to normalize the strength of the links between words [44]. If two keywords appear simultaneously in the same article, they are highly correlated with each other [33] and their association strength is increased. According to [44], the association strength s_{ij} between keyword i and j is defined by (1).

$$\text{Association strength}(s_{ij}) = \frac{c_{ij}}{w_i \cdot w_j}. \tag{1}$$

In (1), c_{ij} denotes the number of co-occurrences of items i and j, and w_i and w_j denote either the total number of occurrences of items i and j or the total number of co-occurrences of these items [44]. The association strength has been widely adopted in previous studies [26–33,43–47].

We adopted VOSviewer as a tool for visualizing a keyword map because it is tailored to visualizing bibliometric maps and for handling complicated maps with large-scale data [44].

3.3.3 | Keyword growth-rate analysis

We then conducted a keyword growth-rate analysis that is suitable for extracting promising dynamic keywords, whereas a co-occurrence analysis is appropriate for assessing their relative importance and relation. A keyword growth-rate analysis is widely used in drawing technology roadmaps [30] and in determining keyword popularity over

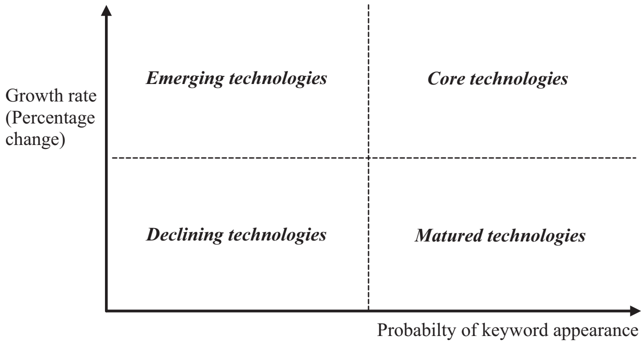

FIGURE 2 Dimensions of keyword growth rate map

time [35]. For such a map, the horizontal axis indicates the probability of a keyword appearance, and the vertical axis shows the growth rate. According to a previous study [30], keywords can be assigned into four groups, namely core, emerging, declining, and matured technologies, as shown in Figure 2.

4 | ANALYSIS RESULTS

4.1 | Keyword frequency analysis

Although IoT technologies have been evolving seamlessly, we divided the analysis period into three generations in order to summarize and trace the comprehensive changes in technologies, as proposed in [3]. In addition, another study [35] divided published research articles into three-year periods to facilitate interpretation.

According to [3], the first IoT generation was dominated by wireless communication technologies such as Radio Frequency IDentification (RFID) and Wireless Sensor Networks (WSNs). In the second generation, detailed network protocols, namely, 6LoWPAN, RPL, 802.15.4, and CoAP, were developed. The keywords social IoT (SIoT), future Internet, and cloud represent the third generation because social interactions among things have been extensively discussed in these areas.

Although their approach is written under a network-oriented perspective and does not clarify a specific period, our analysis adopted periodic classifications of the three generations.

Because the term "IoT" was coined in 1999, it was reasonable to collect keywords from studies published after the year 2000. Thus, retrieved keywords were classified into the first (2000–2005), second (2006–2011), and third (2012–2016) generations. Table 4 summarizes the basic statistics of retrieved keyword data.

As shown in Table 4, it appears that IoT keywords have converged and diverged simultaneously. As the number of listed keywords increases, the IoT keywords diverge. While the number of listed keywords in the first generation is counted as 447, the number in the second generation is counted as 3,737. In addition, the fact that the ratio of unique keywords is incrementally decreasing indicates that the IoT keywords converged.

It may be useful to keep track of the top keywords in order to investigate the dynamics of IoT studies. Appendix 1 reports the frequency of the top 30 words in each generation. A keyword frequency analysis confirms that previous studies [4,16,21,22] have described the state-of-the-art elements of IoT very well. Because RFID, WSN, and Bluetooth are included within the top keywords representing the first generation, wireless communication technologies have clearly triggered IoT studies.

Although wireless communication technologies were still highly ranked in the top-30 keywords in the second period, new keywords emerged as well. For example, 6LoWPAN, CoAP, ZigBee, and other wireless network protocols appeared in the top 30. Simultaneously emerging keywords such as context-aware, business model, smart city, and smart object were also ranked in the top 30. Thus, it can be inferred that the identification of the potential IoT applications has been an urgent issue in the second generation. In the third generation, the volume and variety of keywords grew rapidly, with the total number of keywords counted reaching 48,263. Important keywords, including WSN, RFID, cloud, security, IoT sensors, big data, M2M, privacy, and other emerging technologies have been widely studied. Meanwhile, it should be noted that emerging keywords such as trust, smart home, ontology, pervasive computing, and interoperability have also appeared. These keywords have been the focus of attention in several IoT studies. We discuss these keywords in more detail in the next section.

TABLE 4 Basic statistics of IoT-related keywords

	First generation	Second generation	Third generation	Total
Number of reported studies	121	1,330	11,149	12,600
Number of retrieved keywords (1)	516	5,458	48,263	54,237
Number of listed keywords	447	3,737	22,632	26,816
Number of unique keywords (2)	414	3,248	17,999	19,776
Ratio of unique keywords (2)/(1)	80.23%	59.51%	37.29%	36.46%

4.2 | Keyword mapping analysis

4.2.1 | Overview

To trace the evolution of emerging technologies for IoT, it is necessary to investigate the relationships among relevant keywords from a holistic perspective. We extracted 54,237 keywords from 12,600 academic articles. The process of constructing a map follows three steps: calculation of the keyword similarity matrix, minimizing the weighted sum of the squared Euclidean distances between all pairs of keywords, and performing a transformation of the results to resolve the local optimization problem [44]. The visualized IoT keyword map (2000–2016) is shown in Figure 3.

VOSviewer constructed seven clusters, each of which has distinctive characteristics. Cluster 1 (architecture and service) is composed of the terms agent, architecture, context, AR, interoperability, REST, platform, semantic, SIoT, SOA, social network, web, and web services, among others. This cluster indicates a large number of architectures and services in the IoT research field based on social interactions. Cluster 2 (IoT network) includes the keywords 6LoWPAN, M2M, WSN, mobility, protocol, scalability, standard, and RPL. Cluster 3 (CPS and big data) includes big data, CEP, control, CPS, industry, RFID, SCM, system, data, and data analytics. Technologies in this cluster are associated with how to utilize data for industrial application how to utilize data. Cluster 4 (IoT sensors and monitoring) includes the keywords Arduino, device, IoT sensor, monitoring, Raspberry Pi, and sensor home automation. It is inferred that a field of study related to how to apply IoT technologies in terms of lightweight realization has developed. Cluster 5 (business model and cloud) includes the keywords business model, cloud, cloud computing, cloud manufacturing, cloud service, smart city, SDN, and resource. Cluster 5 indicates a research stream considering IoT services on the cloud. Cluster 6 (privacy, security, and trust) includes the keywords security, privacy, trust, access control, authentication, detection, and cryptography. The existence of cluster 6 indicates that one research stream of IoT has considered privacy, security, and trust. The final cluster (Internet and things) is composed of the two keywords Internet and things. This cluster indicates that two of the considerations of IoT research are the Internet itself and things.

Interestingly, each visible IoT application belongs to a different cluster. Specifically, the smart home is widely adopted in cluster 4 (IoT sensor and monitoring), smart city and e-health are associated with cluster 5 (business model and cloud), smart objects and the environment are involved with cluster 1 (architecture and service), and the smart grid is associated with cluster 3 (CPS and big data). This finding indicates that each cluster has adopted an independent service as a proof of concepts.

One interesting finding is that several emerging technologies for wireless connections are not deeply integrated because each of them belongs to a different cluster. While RFID, WSN, and ZigBee are linked with each other, they are assigned into different clusters as seen in Figure 3. Thus, integration, interoperability, and scalability issues are significant research challenges, as reported in previous studies [5,6,22].

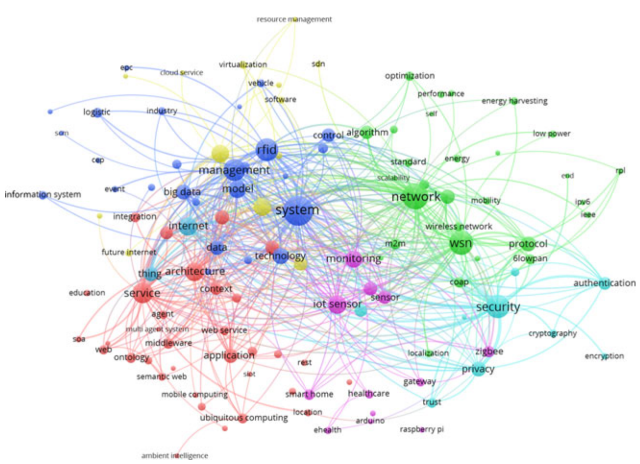

FIGURE 3 IoT keyword map (2000–2016)

4.2.2 | First generation (2000–2005)

A total of 516 keywords were collected from 2000 to 2005. In the first generation, the most frequently observed IoT-related keywords include RFID (12), WSN (8), cloud computing (6), 6LoWPAN (4), application (4), IoT sensors (4), M2M (4), architecture (3), big data (3), CPS (3), industry 4.0 (3), and privacy (3), as shown in Appendix 1. A keyword map constructed using VOSviewer is depicted in Figure 4.

As shown in Figure 4, few studies were conducted in this generation. There are four IoT research clusters. Cluster 1 is composed of data, M2M, and privacy; cluster 2 is composed of CPS, cloud, and industry; cluster 3 is composed of application and architecture; and cluster 4 is composed of context and RFID.

While the linkage among keywords is quite limited, researchers have studied RFID, architecture, applications,

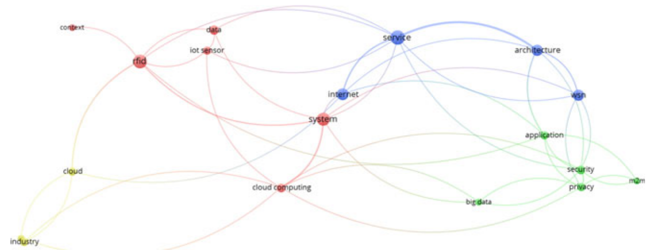

FIGURE 4 IoT keyword map (2000–2005)

and cloud for conceptualizing IoT. In addition, it was observed that CPS considered an industrial application based on the cloud after it emerged. Although the observed keywords are not tightly coupled, the essential keywords emerged during this generation.

4.2.3 | Second generation (2006–2011)

Compared to the first generation, the number of IoT-related keywords grew dramatically during the second generation (2006–2011). During this period, a total of 5,458 keywords were indexed in 1,330 articles. As shown in Appendix 1, the most frequently observed IoT-related keywords include RFID (144), WSN (101), IoT sensors (68), cloud computing (63), security (52), M2M (46), smart city (30), 6LoWPAN (26), ZigBee (26), big data (23), middleware (23), privacy (21), context-aware (20), CPS (20), and CoAP (19). These keywords indicate that specific studies on IoT have progressed. A visualized IoT keyword map is shown in Figure 5.

As shown in Figure 5, during the second generation, researchers conducted diverse studies and attempted to integrate a number of enabling technologies into IoT. Compared to the first generation, previously separated research fields were connected in a complex manner. In addition, the network density increased with the emergence of new keywords.

During the second generation, seven research clusters were identified. Cluster 1 is associated with wireless communications, and includes the keywords 6LoWPAN, SDN, WSN, ZigBee, wireless network, and access control. Cluster 2 includes the keywords cloud, CPS, platform, management, pervasive computing, resource management, smart city, and information technology. The presence of cluster 2 indicates that researchers were highly interested not only connecting but also controlling the resources based on the IoT platform. Cluster 3 includes the keywords big data, data mining, cloud computing, cloud manufacturing, RFID, logistics, and information. This cluster appears to indicate that one research stream contributed to the analysis of big data on the cloud. Cluster 4 is closely associated with IoT applications because the keywords in this cluster include architecture, CoAP, rest, sensor, smart home, data, and location. Studies in cluster 5 contributed to authentication and security because the keywords in this cluster include authentication, model, ontology, resource, privacy, and security. It should be noted that security and privacy issues in the authentication and construction of knowledge through ontology were considered as one of the research streams. Cluster 6 includes the keywords device, event, Internet, middleware, monitoring, systems, and web service. Thus, it is inferred that one of the major issues in the second generation was monitoring of things existing in IoT system on the Web. Finally, one of the research streams in the second generation is related to business and services because cluster 7 includes the keywords business model, interoperability, simulation, SOA, social network, and trust.

4.2.4 | Third generation (2012–2016)

During the third generation, a total of 48,263 keywords were retrieved. The most frequently observed IoT-related keywords include WSN (932), RFID (900), cloud computing (606), security (511), IoT sensor (495), big data (320), M2M (294), smart city (275), privacy (215), cloud (201), ZigBee (194), smart home (187), CPS (174), trust (170), 6LoWPAN (167), CoAP (164), and context-aware (160).

Figure 6 shows the IoT keyword map during the third generation. During the third generation, seven clusters were identified. Cluster 1 includes the keywords architecture,

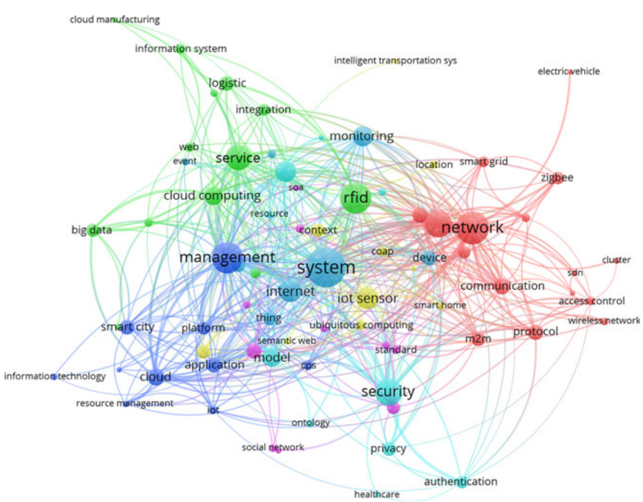

FIGURE 5 Keyword map of IoT (2006–2011)

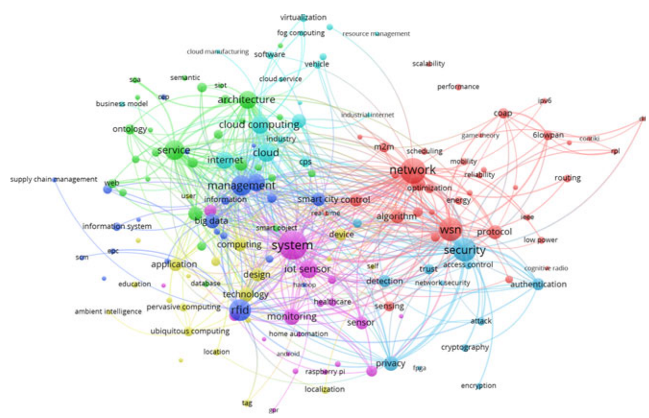

FIGURE 6 Keyword map of IoT (2012–2016)

service, data, SIoT, web, cloud computing, automation, ontology, semantic, and SOA. This cluster indicates that there are a research stream aimed at building a web-based service based on a semantic ontology in cloud computing. Cluster 2 includes the keywords RFID, management, big data, data mining, data analysis, SCM, and logistics. It was inferred that studies in this cluster have the consistent goal of managing things by analyzing big data from the RFID tag in the SCM field. Cluster 3 includes system, IoT sensor, monitoring, ZigBee, and smart home. Researchers in this cluster may be interested in a system for monitoring things using IoT sensors based on the ZigBee protocol, which is suitable for low-power, low data rates, and close proximity. Thus, the smart home may be an appropriate application. Cluster 4 consists of several keywords for wireless network technologies and standards, and includes WSN, 6LoWPAN, CoAP, protocol, routing, mobility, reliability, and scalability. Studies in cluster 4 have been dedicated to building reliable wireless communication environments. Cluster 5 includes the keywords Internet, business model, cloud, CPS, thing, industry, and cloud manufacturing. Based on these keywords, it is inferred that a way of utilizing IoT in industry and business is one of the main issues. Cluster 6 includes keyword application, device, design, education, and environment. Because its keywords are not directly related to specific technologies, cluster 6 appears to focus on providing future IoT research directions. For example, the keywords technology, application, design, and device appear to be abstract terms. Cluster 7 have focused on studying security, authentication, privacy, trust, and detection issues, which were discussed during the second generation.

Meanwhile, it should be noted that the term "standard" connects security and architecture in cluster 1. This indicates that security issues of the IoT architecture were considered using a standard. In addition, one of the interesting findings is that the keyword "trust" is connected to security and privacy in cluster 7, and to network, WSN, and system in other clusters. In contrast, in the second generation, "trust" was not associated with those keywords, but this was the case in the third generation.

In addition, the role of CPS is noteworthy because it links core technologies in other clusters. Specifically, CPS is directly connected to the keywords network, security, management, cloud, big data, system, industry, architecture, and the Internet. It appears that CPS is a realized form of IoT because the linked technologies are almost identical to the IoT elements proposed in a previous study [6]. However, CPS appears to place more weight on commercial applications because the keyword industry is directly involved in CPS, as shown in Figure 7.

Although it seems that significant progress from the second to the third generation has not been made at a glance,

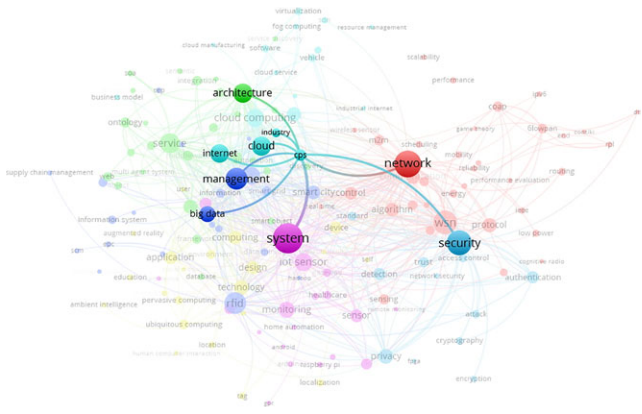

FIGURE 7 CPS-centric keyword map (2012–2016)

there were several significant changes during this period. First, the diversity in the studies was explicitly shown, as is clearly shown by the emerging keywords (namely trust and context-aware). Second, the research density increased during the third generation, as indicated by the increasing number of specific topics. For example, the number of studies on CPS was counted as 20 in the second generation, whereas it was counted as 174 in the third generation. This indicates that related studies on a certain research topic have progressed and matured. Third, the relationships between the keywords become stronger. For example, the relationship among the keyword architecture, cloud, Internet, service, big data, management, system, IoT sensor, and RFID has been clearly strengthened. In other words, various studies on IoT are connected with each other.

4.3 | Keyword growth-rate analysis

4.3.1 | From the first to second IoT generations

Figure 8 shows a keyword growth-rate analysis map from the first generation to the second generation. Rather than the absolute number of keywords, we adopted the probability of keyword appearance because the total number of keywords during the first generation is different from that during the second generation. We traced the changes in the top 20 keywords during the first and second generations, and then depicted the keyword growth map. The keywords authentication, middleware, CoAP, ubiquitous computing, smart city, ZigBee, context-aware, and security were assigned to keywords that represent emerging technologies.

4.3.2 | From the second to third IoT generations

Figure 9 shows a keyword growth-rate analysis map from the second to third generations. As seen, the keywords trust,

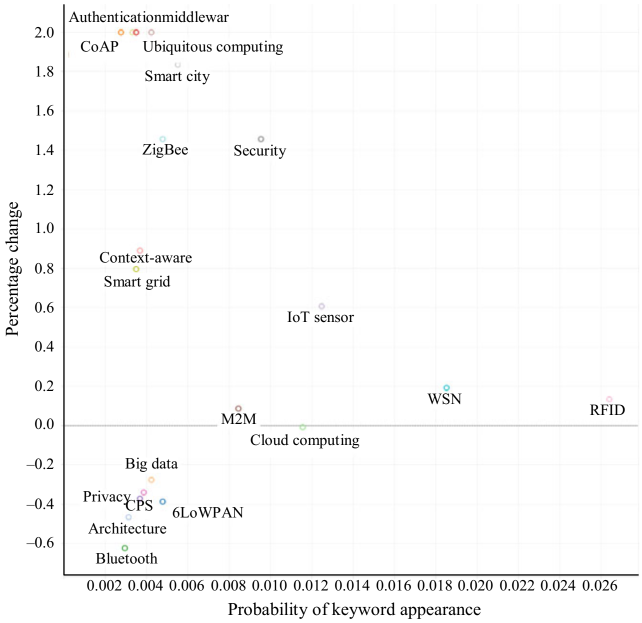

FIGURE 8 Keyword growth rate map from the first to second generations

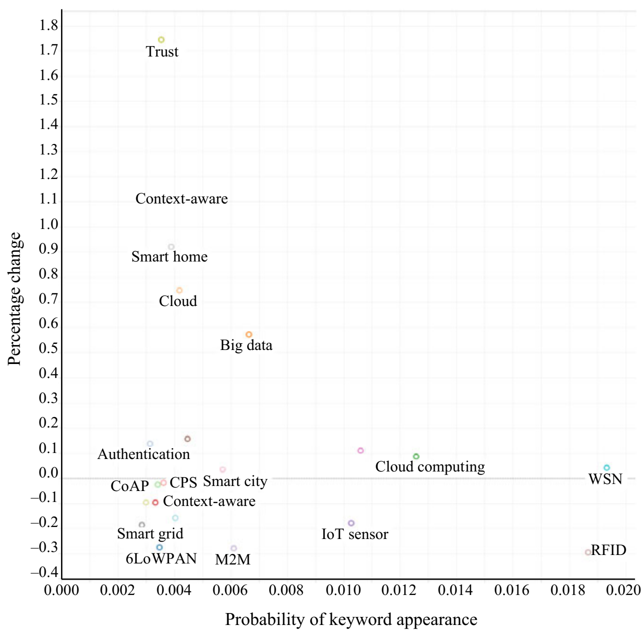

FIGURE 9 Keyword growth rate map from the second to third generations

smart home, cloud, authentication, and big data are classified into emerging keywords. Notably, emerging technologies that are represented by those keywords indicate the attention of researchers. Although the growth rate of CoAP, CPS, smart city, context-aware, smart grid, 6LoWPAN, and M2M have slightly declined, it should be noted that further research opportunities still exist because these keywords are identified as fundamental keywords in co-occurrence analysis.

After passing through two generations, emerging technologies for wireless connections, namely WSN and RFID, have matured, and network protocols (eg, 6LoWPAN and CoAP) announced relevant specifications. As such, the map also confirms that the IoT is rooted in wireless communication technologies. It is remarkable that, trust, the smart home, cloud, CPS, context-aware, and big data may be valuable research topics in the fourth generation.

4.4 | Betweenness centrality of keywords

Although we present and reveal the dynamics of an IoT knowledge map as well as emerging technologies, one may wonder about the relations among keywords. To address this issue, we computed the betweenness centrality (BC) among keywords to identify the relational dependency. The BC has been widely used in previous works because it predicts preferential attachments by new entering technologies [48] and measures the ability to control the information flow [49]. We calculated the BC for keywords that appear more than 40 times, as shown in Table 5.

One of the interesting findings is that the values of BC are relatively low. This indicates that dependencies among technologies for IoT are quite weak because they are connected to each other through multiple paths. Thus, this infers that the IoT standard and dominant technologies have not yet emerged because the values of BC for each keyword appear not to be constrained to other technologies.

5 | CONCLUSIONS

After the concept of IoT was first introduced, researchers began to focus on developing various enabling technologies in multiple domains. Their efforts have improved individual technologies of IoT. However, it should be noted that researchers are prone to the error of not “seeing the forest for the trees.” To overcome the absence of a holistic view, previous studies have attempted to present and summarize the history, concept, and components of IoT. However, most have relied on narrative views. In other words, their approaches may be biased by the authors’ particular insights.

To fill in the gaps, this study explores the IoT knowledge structure to investigate state-of-the-art elements and future directions. For this purpose, we retrieved bibliographic data from the Scopus database. We retrieved 54,237 keywords from 12,600 studies.

Based on a keyword frequency analysis, we found that IoT studies are simultaneously converging and diverging. In addition, IoT has successfully motivated the participation of numerous researchers. Remarkably, these findings indicate that researchers recognize IoT as a significant research opportunity.

TABLE 5 Top 30 keywords according to betweenness centrality

Keyword	BC (ranks)	Keyword	BC (ranks)	Keyword	BC (ranks)
Network	0.0127 (1)	Architecture	0.0095 (11)	Sensor	0.0067 (21)
Management	0.0124 (2)	RFID	0.0086 (12)	Smart city	0.0066 (22)
System	0.0123 (3)	Model	0.0081 (13)	Device	0.0065 (23)
WSN	0.0112 (4)	Thing	0.0080 (14)	Context	0.0065 (24)
Internet	0.0110 (5)	Monitoring	0.0079 (15)	Platform	0.0060 (25)
Security	0.0106 (6)	Big data	0.0076 (16)	Protocol	0.0059 (26)
IoT sensor	0.0105 (7)	Technology	0.0076 (17)	Design	0.0058 (27)
Cloud computing	0.0104 (8)	Application	0.0074 (18)	Control	0.0057 (21)
Cloud	0.0103 (9)	Communication	0.0073 (19)	CPS	0.0057 (22)
Service	0.0096 (10)	Computing	0.0069 (20)	Data	0.0055 (23)

We conducted a co-occurrence analysis and revealed the evolution of IoT studies. Based on investigations using a visualized knowledge structure, we found that relevant studies have become more sophisticated and profound with time. In particular, in the third IoT generation, seven research clusters were found.

Several findings are noteworthy. First, technologies for IoT have converged and diverged simultaneously. The relationship among keywords is becoming stronger, and emerging research stems have been added to IoT. Second, we found that RFID and WSN are not deeply associated with each other. Thus, the integration and interoperability of wireless communication technologies may have potential as future research opportunities. Third, privacy, security, trust, authentication, and CPS are highly expected research topics. Because there are inherent risks in connecting to unknown things, both trust and security may be required. In addition, methods that focus on controlling and managing things are being considered, as highlighted by authentication on the map. Fourth, IoT studies have concentrated on the wireless network domain.

Based on a keyword growth-rate analysis, we extracted emerging technologies. Specifically, trust, smart home, cloud, authentication, context-aware, and big data were classified as emerging technologies. In addition, the next step of IoT may be data analysis, as shown through "big data" of the growth-rate analysis. These findings straightforwardly guide further research directions.

Although this study provides valuable insights, there are several limitations. First, our data were limited to the Scopus database. Although it is the largest academic literature database, further studies replicating our study using data from another database such as the Web of Science (WoS), which covers select peer-reviewed journals with high impact factors, is required. Second, our analysis was limited to academic studies, and the use of keywords in the other knowledge structure (eg, a patent analysis) may be a valuable research opportunity. Finally, it may be useful to explore the IoT knowledge structure with different scientific methods because keywords or research topics have not yet been fully investigated. For example, a further research opportunity may be to conduct a latent Dirichlet allocation (LDA) or event study analysis to discover hidden topics or the dynamics of topics. Toward the fourth IoT generation, it is necessary to consider how to assemble individual technologies into a complete form of IoT.

ACKNOWLEDGMENTS

This work was supported by the ICT R&D program of MSIP/IITP [R0190-15-2027, Development of TII (Trusted Information Infrastructure) S/W Framework for Realizing Trustworthy IoT Eco-system], Republic of Korea.

ORCID

Hangjung Zo http://orcid.org/0000-0002-2892-1659

REFERENCES

1. K. Ashton, *That 'internet of things' thing*, RFiD J. **22** (2009), no. 7, 97–114.
2. R. Baheti and H. Gill, *Cyber-physical systems*, Impact Cont. Technol. **12** (2011), 161–166.
3. L. Atzori, A. Iera, and G. Morabito, *Understanding the internet of things: Definition, potentials, and societal role of a fast evolving paradigm*, Ad Hoc Netw. **56** (2017), no. Sup. C, 122–140.
4. L. Atzori, A. Iera, and G. Morabito, *The internet of things: A survey*, Comput. Netw. **54** (2010), no. 15, 2787–2805.
5. A. M. Ortiz et al., *The cluster between internet of things and social networks: Review and research challenges*, IEEE Internet Things J. **1** (2014), no. 3, 206–215.
6. A. Al-Fuqaha et al., *Internet of things: A survey on enabling technologies, protocols, and applications*, IEEE Commun. Surveys Tutor **17** (2015), no. 4, 2347–2376.
7. V. Gunes et al., *A survey on concepts, applications, and challenges in cyber-physical systems*, Trans. Internet Inform. Syst. **8** (2014), no. 12, 4242–4268.

8. B. N. Green, C. D. Johnson, and A. Adams, *Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade*, J. Chiropr. Med. **5** (2006), no. 3, 101–117.
9. D. Rotolo, D. Hicks, and B. R. Martin, *What is an emerging technology?*, Res. Policy **44** (2015), no. 10, 1827–1843.
10. M. Halaweh, *Emerging technology: What is it?*, J. Technol. Manag. Innovation **8** (2013), no. 3, 108–115.
11. H.-D. Ma, *Internet of things: Objectives and scientific challenges*, J. Comput. Sci. Technol. **26** (2011), no. 6, 919–924.
12. A. Koubâa and B. Andersson, A vision of cyber-physical internet, *Int. Workshop Real-Time Netw.* **(RTN'09)**, Dublin, Ireland, June 2009, pp. 1–6.
13. F. Xia et al., *Internet of things*, Int. J. Commun Syst **25**, (2012), no. 9, 1101–1102.
14. ITU-T, *Itu-T Standard Y.2060*, 2012.
15. Z.-K. Zhang et al., IoT security: Ongoing challenges and research opportunities, *IEEE Int. Conf. Serv.-Oriented Comput. Applicat.* Matsue, Japan, Nov. 2014, pp. 230–234.
16. S. Madakam, R. Ramaswamy, and S. Tripathi, *Internet of things (IoT): A literature review*, J. Comput. Commun. **3** (2015), no. 5, 164–173.
17. A. Bassi and G. Horn, Internet of things in 2020: A roadmap for the future, *Working Group RFID of the ETP EPoSS*, Version 1.1, 2008, pp. 97–114.
18. R. R. Rajkumar et al., Cyber-physical systems: The next computing revolution, *Proc. Design Autom. Conf.*, Anaheim, CA, USA, June 13–18, pp. 731–736.
19. E. A. Lee, Cyber-physical systems - are computing foundations adequate, *Position Paper NSF Workshop Cyber-Phys. Syst.: Res. Motivation, Tech.* Roadmap, Austin, TX, USA, 16–17, Oct. 2006, pp. 1–9.
20. J. Lee, B. Bagheri, and H.-A. Kao, *A cyber-physical systems architecture for industry 4.0-based manufacturing systems*, Manuf. Lett. **3** (2015), 18–23.
21. J. Gubbi et al., *Internet of things (IoT): A vision, architectural elements, and future directions*, Future Gener. Comput. Syst. **29** (2013), no. 7, 1645–1660.
22. H. Shah Ahsanul, A. Syed Mahfuzul, and R. Mustafizur, *Review of cyber-physical system in healthcare*, Int. J. Distrib. Sensor Netw. **10** (2014), no. 4, pp. 217415:1–217415:20.
23. M. Farooq et al., *A review on internet of things (IoT)*, Int. J. Comput. Applicat. **113** (2015), no. 1, 1–7.
24. J. A. Collins and B. C. Fauser, *Balancing the strengths of systematic and narrative reviews*, Oxford, Oxford University Press, 2005.
25. M. Oraee et al., *Collaboration in bim-based construction networks: A bibliometric-qualitative literature review*, Int. J. Project Manag. **35** (2017), no. 7, 1288–1301.
26. H.-N. Su and P.-C. Lee, *Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in technology foresight*, Scientometrics **85** (2010), no. 1, 65–79.
27. J. Choi and Y.-S. Hwang, *Patent keyword network analysis for improving technology development efficiency*, Technol. Forecast. Soc. Change **83** (2014), 170–182.
28. Z. Hu, S. Fang, and T. Liang, *Empirical study of constructing a knowledge organization system of patent documents using topic modeling*, Scientometrics **100** (2014), no. 3, 787–799.
29. G. Kim and J. Bae, *A novel approach to forecast promising technology through patent analysis*, Technol. Forecast. Soc. Change **117** (2017), 228–237.
30. S. Lee et al., *Using patent information for designing new product and technology: Keyword based technology roadmapping*, R&D Manag. **38** (2008), no. 2, 169–188.
31. I. Park et al., *Exploring promising technology in ICT sector using patent network and promising index based on patent information*, ETRI J. **38** (2016), no. 2, 405–415.
32. K. Song, K. Kim, and S. Lee, *Identifying promising technologies using patents: A retrospective feature analysis and a prospective needs analysis on outlier patents*, Technol. Forecast. Soc. Change **128** (2017), 118–132.
33. B.-N. Yan, T.-S. Lee, and T.-P. Lee, *Mapping the intellectual structure of the internet of things (Iot) field (2000–2014): A co-word analysis*, Scientometrics **105** (2015), no. 2, 1285–1300.
34. S. Lee, B. Yoon, and Y. Park, *An approach to discovering new technology opportunities: Keyword-based patent map approach*, Technovation **29** (2009), no. 6, 481–497.
35. C. F. Mela, J. Roos, and Y. Deng, *A keyword history of marketing science*, Marketing Sci. **32** (2013), no. 1, 8–18.
36. B. Fahimnia, J. Sarkis, and H. Davarzani, *Green supply chain management: A review and bibliometric analysis*, Int. J. Production Economics **162** (2015), 101–114.
37. B. Cheng et al., *Research on e-learning in the workplace 2000–2012: A bibliometric analysis of the literature*, Educational Res. Rev. **11** (2014), 56–72.
38. M. E. Falagas et al., *Comparison of scimago journal rank indicator with journal impact factor*, FASEB J. **22** (2008), no. 8, 2623–2628.
39. L. Heilig and S. Voß, *A scientometric analysis of cloud computing literature*, IEEE Trans. Cloud Comput. **2** (2014), no. 3, 266–278.
40. D. Mishra et al., *Vision, applications and future challenges of internet of things: A bibliometric study of the recent literature*, Ind. Manag. Data Syst. **116** (2016), no. 7, 1331–1355.
41. G. Mulligan, *The internet of things: Here now and coming soon*, IEEE Internet Comput. **14** (2010), no. 1, 35–36.
42. P. Berkhin, *A survey of clustering data mining techniques*, Grouping Multidimensional Data, ed., Berlin, Heidelberg, Springer, 2006, pp. 25–71.
43. M. Callon, J. P. Courtial, and F. Laville, *Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry*, Scientometrics **22** (1991), no. 1, 155–205.
44. N. J. Van Eck and L. Waltman, *Software survey: Vosviewer, a computer program for bibliometric mapping*, Scientometrics **84** (2010), no. 2, 523–538.
45. B. Yoon, S. Lee, and G. Lee, *Development and application of a keyword-based knowledge map for effective R&D planning*, Scientometrics **85** (2010), no. 3, 803–820.
46. H. Peters and A. F. van Raan, *Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling*, Res. Policy **22** (1993), no. 1, 23–45.
47. Y. Ding, G. G. Chowdhury, and S. Foo, *Bibliometric cartography of information retrieval research by using co-word analysis*, Inform. Process. Manag. **37** (2001), no. 6, 817–842.
48. A. Abbasi, L. Hossain, and L. Leydesdorff, *Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks*, J. Informetrics **6** (2012), no. 3, 403–412.
49. M. E. Newman, *A measure of betweenness centrality based on random walks*, Soc. Netw. **27** (2005), no. 1, 39–54.

AUTHOR BIOGRAPHIES

Young Seog Yoon received a BS degree in industrial engineering from Pusan National University, Pusan, Rep. of Korea, in 2003. He received his MS. degree in information and technology management from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 2005. He is a PhD candidate in the School of Business and Technology Management at KAIST. After 2015, he has worked for the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea, as a senior researcher. His research interests include consumer behavior, data science, big data analytics, e-commerce, the Internet of things, cyber-physical systems, and smart factories.

Hangjung Zo is an associate professor of IT management in the School of Business and Technology Management at Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea. He obtained his PhD in IT management from the University of Wisconsin-Milwaukee, USA. His research interests include digital social innovations and the impacts of digital technology on business and society. His papers have appeared in the *IEEE Transaction on Systems Man and Cybernetics, Decision Support Systems, Journal of Business Research, Electronic Commerce Research and Applications, and Computers and Education*, among others. He is the director of the Research Institute for Social Technology and Innovation at KAIST.

Munkee Choi received his BS degree in applied mathematics from the College of Engineering, Seoul National University, Rep. of Korea, in 1974, and his MS degree in industrial engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 1978, and his PhD degree in engineering from North Carolina State University, Raleigh, NC, USA, in 1989. From 1978 to 1998, he worked for the Electronics and Telecommunications Research Institute, Daejeon, Rep. of Korea. Since 1999, he has been with the School of Business and Technology Management, KAIST, Daejeon, Rep. of Korea, where he was a professor and now he is the Professor Emeritus. His main research interests are network service business and global technology commercialization.

Donghyun Lee received his BS degree in information and communication engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 2009, and his MS and PhD degrees from the Innovation and Technology Management from the KAIST in 2011 and 2016, respectively. He worked for the big data research team at the Korea Environment Institute, Sejong, Rep. of Korea, in 2016. He is now an assistant professor in the Department of Business Administration, Korea Polytechnic University, Rep. of Korea. His research interests include data science, big data analytics, and deep learning for environment and social science.

Hyun-woo Lee received his MS degree and PhD degrees in 1995 and 2005, respectively, from the Korea Aerospace University, Koyang, Rep. of Korea. He is currently an assistant vice president of the Media Research Division of the Electronics and Telecommunications Research Institute, Deajeon, Rep. of Korea. His main research interests include trust information infrastructure, smart media platform, ATSC3.0 services, and an OTT platform in next-generation broadcasting. His current research interests also include cognitive IoT platforms and an edge computing.

APPENDIX 1

TOP 30 RETRIEVED IOT RELEVANT KEYWORDS FROM 2000 TO 2016

	2000-2005 (Number of keywords: 516)		2006-2011 (Number of keywords: 5,458)		2012-2016 (Number of keywords: 48,263)		2000-2016 (Number of keywords: 54,237)	
Rank	Keywords	Number of appearances	Keywords	Number of appearances	Keywords	Number of appearances	Keywords	Number of appearances
1	RFID	12	RFID	144	WSN	932	RFID	1,056
2	WSN	8	WSN	101	RFID	900	WSN	1,041
3	Cloud computing	6	IoT sensor	68	Cloud computing	606	Cloud computing	675
4	6LoWPAN	4	Cloud computing	63	Security	511	IoT sensor	567
5	Applications	4	Security	52	IoT sensor	495	Security	565
6	Bluetooth	4	M2M	46	Big data	320	Big data	346
7	IoT sensor	4	Smart city	30	M2M	294	M2M	344
8	M2M	4	6LoWPAN	26	Smart city	275	Smart city	306
9	Architecture	3	ZigBee	26	Privacy	215	Privacy	239
10	Big data	3	Big data	23	Cloud	201	ZigBee	221
11	CPS	3	Middleware	23	ZigBee	194	Cloud	215
12	Industry 4.0	3	Privacy	21	Smart home	187	Smart home	199
13	Privacy	3	Context-aware	20	CPS	174	6LoWPAN	197
14	PUF	3	CPS	20	Trust	170	CPS	197
15	Availability	2	CoAP	19	6LoWPAN	167	CoAP	183
16	CEP	2	Smart grid	19	CoAP	164	Context-aware	181
17	Cloud manufacturing	2	Ubiquitous computing	18	Context-aware	160	Trust	177
18	Culture	2	Architecture	17	Authentication	151	Authentication	166
19	Design	2	Bluetooth	16	Ubiquitous computing	144	Ubiquitous computing	162
20	Energy saving	2	Authentication	15	Smart grid	137	Smart grid	157
21	GPS	2	Future Internet	14	Internet	136	Middleware	153
22	Linked data	2	QoS	14	Energy efficiency	130	Internet	147
23	MTC	2	Cloud	13	Middleware	130	Energy efficiency	144
24	Recommender systems	2	Energy efficiency	13	Bluetooth	115	Bluetooth	135
25	Research challenges	2	Business model	12	Ontology	115	Architecture	132
26	Resource management	2	ipv6	11	Architecture	112	Ontology	123
27	SDN	2	Logistics	11	Wireless network	107	QoS	120
28	Security	2	Smart home	11	QoS	105	Wireless network	117

(Continues)

(Continued)

	2000-2005 (Number of keywords: 516)		2006-2011 (Number of keywords: 5,458)		2012-2016 (Number of keywords: 48,263)		2000-2016 (Number of keywords: 54,237)	
Rank	Keywords	Number of appearances	Keywords	Number of appearances	Keywords	Number of appearances	Keywords	Number of appearances
29	Service composition	2	Smart objects	11	Pervasive computing	101	Pervasive computing	109
30	Service-oriented architecture	2	Internet	10	Interoperability	97	Embedded system	106