Journal of The Korea Convergence Society
Vol. 13. No. 5, pp. 261-266, 2022
e-ISSN 2713-6353
DOI : https://doi.org/10.15207/JKCS.2022.13.05.261

R과 LDA 기법을 활용한 클라우드 컴퓨팅 동향에 관한 연구: 해외 연구 동향을 중심으로

강태구
건양사이버대학교 IT비즈니스학과 교수

Study on Tendency of Cloud Computing Using R and LDA Technique : Focusing on Tendency of Overseas Studies

Tae-Gu Kang
Professor, Dept. of IT Business Administration, Konyang Cyber University

요 약 4차 산업혁명에서 비롯된 디지털 시대의 본격화와 COVID-19의 영향으로 기업뿐만 아니라 다양한 분야에서 변화가 일어나고 있는 상황이다. 디지털 서비스의 급격한 증가로 인해 클라우드 시장 또한 성장 가속화로 빠르게 변화하는 디지털 환경에서 중요성이 더욱더 강조되고 있는 실정이다. 이에 관련 산업뿐만 아니라 다양한 분야에서 클라우드는 지속적인 성장과 생존전략을 위한 대표적인 전략 중 하나라고 할 수 있다. 클라우드 관련 연구는 다양하게 진행되고 있지만 연구 동향은 미흡한 실정이다. 이에 본 논문은 R과 LDA 기법을 활용한 외국 학술 데이터베이스인 SCOPUS를 활용하여 클라우드 컴퓨팅의 연구 동향을 분석하였다. 분석 결과 클라우드 컴퓨팅 관련 높은 관심으로 수많은 연구가 진행되고 있는 것으로 나타났고, 키워드 분석 결과 클라우드 컴퓨팅이 가장 많이 도출되었다. 클라우드 컴퓨팅 주제 외에도 클라우드, 클라우드와 컴퓨팅, 데이터, 컴퓨팅의 다양한 주요 키워드들이 도출되었다. 본 연구 결과는 클라우드 컴퓨팅 동향 파악을 위한 실무적인 측면에서 관련 산업 활성화에 밑거름을 제공할 수 있는 기초자료로 활용될 것으로 기대된다.

주제어 : 클라우드 컴퓨팅, 클라우드, SCOPUS, R, LDA, 텍스트 마이닝

Abstract The full-fledged digital age derived from the fourth industrial revolution and the impact of COVID-19 lead to changes in various fields, including companies. In other words, the importance of cloud computing is being emphasized in the rapidly changing digital environment due to the rapid growth of the cloud market due to the rapid increase in digital services. The cloud may be one of the representative strategies for sustainable growth and survival in various fields as well as related industries. Although there have been a variety of studies on the cloud, the tendency of them has been not been adequately examined. This paper, therefore, analyzed the tendency of studies on the cloud computing. by using SCOPUS, the database of overseas academic journals using both R and LAD technique. The findings showed that many studies with high interest in the cloud computing have been conducted, the cloud computing were most often drawn from an analysis on key words. Moreover, various key words, including cloud, cloud and computing, data and computing were drawn, except for the theme of cloud computing. It is expected that could be used as a basic data, in that they provide the foundation for activating the related industries in terms of practice of the cloud computing.

Key Words : Cloud Computing, Cloud, SCOPUS, R, Text Mining

*Corresponding Author : Tae Gu Kang(tgkang@kycu.ac.ac.kr)
Received April 20, 2022 Revised May 10, 2022
Accepted May 20, 2022 Published May 28, 2022

1. 서론

4차 산업혁명 시대가 본격화됨에 따라 산업 전 분야와 ICT(정보통신기술) 융합으로 생겨난 혁명의 근간인 ICBM 플랫폼이 꼽히고 있다. ICBM은 사물인터넷(IoT), 클라우드(Cloud), 빅데이터(BigData), 모바일(Mobil)을 통합해서 서비스를 제공하는 플랫폼으로 4차 산업혁명 시대 기업들로 하여금 ICBM 플랫폼 구축 서비스 전략은 생존을 위한 필수 과제가 되었다[1]. 4차 산업혁명시대 디지털 트랜스포메이션의 본격화와 COVID-19의 영향으로 새로운 디지털 환경에 대한 변화는 예측할 수 없을 정도로 빠르게 변하고 있다. 가트너에 따르면 클라우드가 대중적인 수준을 넘어 확산되어 2025년까지 85% 이상의 조직이 클라우드 우선 원칙 채택으로 다양한 분야 및 기업에서 디지털 경험을 이끄는 중심이 될 것이라고 전망하고 있다[2]. 즉, COVID-19의 펜더믹과 디지털 서비스의 급증은 클라우드 시장 성장 가속화와 클라우드 관련 산업뿐만 아니라 기술의 선택은 더욱더 중요하게 되었다는 것을 의미한다. 결국 빠르게 변화하는 디지털 환경에서 기업뿐만 아니라 다양한 분야에서 지속적인 성장과 생존전략을 위한 대표적인 전략 중 하나가 클라우드라고 할 수 있다. 글로벌 기업과 대표적인 정보 기술 연구 및 자문 기관인 미국의 가트너 뿐만 아니라 국내의 정부, 공공기관, IT산업 분야 등에서도 클라우드 관련 조사를 통한 동향에 대하여 주기적으로 정보 제공을 하고 있다. 하지만 학술적인 차원에서의 클라우드 컴퓨팅 동향 관련 연구는 미흡한 실정이다. 기존 클라우드 컴퓨팅 동향에 대한 연구에서는 미래에 미치는 영향, 빠르게 변화하는 디지털 시대에 새로운 패러다임 기술들에 대하여 클라우드 컴퓨팅 진화에 미치는 영향을 중심으로 개념적 모델 및 사례를 통한 대응 방안 모색 등의 연구가 대부분이다. 본 연구는 기존 연구와 다르게 클라우드 컴퓨팅 동향에 관한 연구 동향 파악을 R과 LDA를 활용하여 토픽에 대하여 모델링하여 분석 수행한 점에서 기존 연구와의 차별성에 의의가 있다고 할 수 있다. 이에 본 연구에서는 클라우드 컴퓨팅 연구 동향 분석을 위해 SCOUPS 데이터베이스를 활용하여 클라우드에 대한 연구 동향에 대하여 파악해 보고자 한다. 또한 기존 클라우드 컴퓨팅에 대한 선행 연구 검토와 토픽의 중요 키워드를 R과 LDA(Latent Dirichlet Allocation)를 활용한 분석을 통해 클라우드의 주제와 방향성을 제시하고자 한다. 클라우드 컴퓨팅에 대한 성장 가속화는 다양한 분야에서의 지속 성장과 생존전략을 위한 대표전략 중 하나인 만큼 본 연구 결과는 정부, 공공기관, IT산업 분야 등 다양한 분야에 클라우드 컴퓨팅 동향을 파악하고 실무에 적용함으로써 실무적인 측면에서의 기초자료로 밑거름을 제공할 수 있을 것으로 기대된다.

2. 이론적 배경

2.1 클라우드 컴퓨팅

클라우드 컴퓨팅의 대표적인 정의로 미국 국립표준기술연구소(NIST: National Instiitue of Standards and Technology) 보고서에서는 “최소한의 관리 노력으로 서비스 제공자와 상호작용으로 유비쿼터스 환경에서 편리하면서도 쉽게 서버, 네트워크, 스토리지, 어플리케이션 등의 컴퓨터 자원 공유 풀에 네트워크를 통한 접속으로 자원을 빠르게 준비하여 관리 가능한 모델”이라고 정의하였다[3]. 클라우드란 마치 구름처럼 공중을 떠다니는 것처럼 특정 장소에 정지된 것이 아니라 언제 어디서든 접속 가능한 형태로 정교한 네트워크 형성으로 Fig. 1과 같이 네트워크 연결을 통해 가상의 서버가 고정된 서버를 대체함으로써 소프트웨어뿐만 아니라 데이터 저장이 가능한 형태이다[4]. 즉, 클라우드 컴퓨팅은 자원의 집합으로 사용자로 하여금 위치, 장치에 대하여 무관하고, 사용자의 요구에 따라 사용한 만큼의 사용량 측정으로 다양한 종류의 애플리케이션, 프로세스, 서비스들에 대하여 총칭하는 것으로 다양한 상황에 따라 확장성 능력을 가진 기술을 의미한다[5].

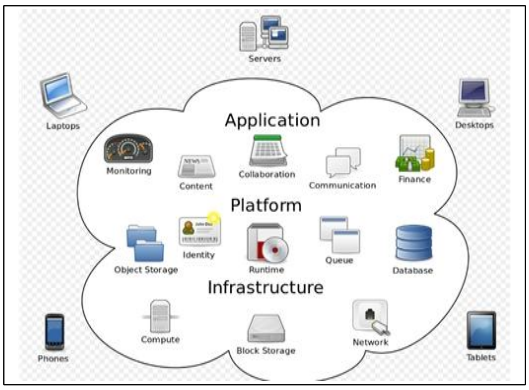

Fig. 1. Cloud Computing Schematic Diagram
source: Wikipdia[6], Tae Gu Kang[8]

클라우드 컴퓨팅 관련 정의는 Table 1과 같이 기업 및 기관에 따라 다양하게 정의하고 있다.

Table 1. Defining Cloud Computing

Division	Definition
Gartner	A form of computing providing services to customers, such as resources with high-quality scalability, by using technologies on internet.
IBM	Environment in which the distributed processing of mass DB is possible in the virtual space by using web-based application SW, and data can be brought and processed, regardless of terminals such as computers, cellphones, PDA, etc.
NIST	Computing for users to use IT resources as much as they need and accordingly pay the costs, where the real-time scalability is supported, depending on service load.
Google	Technology allowing us to use abundant computing resources, which a single computer cannot use, by connecting hundreds or thousands of user- and task-centered computers.
Oracle	A type of computing resiliently providing sources to users and allowing them to share the sources, with the metering method using resilient platforms virtualized for applications, self-service and web technology.
Wikipedia	Internet-based computing technology, in other words, web-based software service for putting programs in utility data servers on internet and bringing them to computers or cellphones, if needed.

source: KISA[7], Tae Gu Kang[8]

2.2 클라우드 컴퓨팅 연구 동향

대부분의 클라우드 컴퓨팅 동향 연구에서는 클라우드 컴퓨팅이 미래에 미치는 영향, 새로운 패러다임 기술이 클라우드 컴퓨팅 진화에 미치는 영향 등 개념적 모델, 사례에 대한 대응방안 모색 등의 연구가 대부분이다.

Gill et al.(2019)의 연구에서는 미래에 응용 가능한 핵심기술 식별에 대한 필요성으로 블록체인, IoT 및 인공지능의 세 가지 새로운 패러다임 기술이 클라우드 컴퓨팅 진화에 미치는 영향에 대한 탐색적 연구를 통해 미래 기술에 대한 개념적 모델을 제안하였다[9]. AlTwaijiry(2021)는 클라우드 컴팅에 대한 제한적인 제어, 공급업체의 종속성, 클라우드 보안성, 에너지 소비 등의 현재 클라우드의 한계에 대한 극복을 위해 엣지 컴퓨팅, 하이브리드 및 멀티 클라우드 솔루션, 그린 클라우드 컴퓨팅 등 동향 분석을 통한 클라우드 컴퓨팅의 한계에 대한 미래의 대안을 제시하였다[10]. Abbasi et al.(2019)의 사례연구에는 클라우드 컴퓨팅의 리소스 관리 문제에 대한 개선연구로 리소스 조정 및 애플리케이션 개발이라는 두 가지 영역에 대한 예측으로 향후 대응에 대안을 제시하였다[11].

2.3 텍스트 마이닝

텍스트 마이닝 활용은 비정형 데이터들 중에서 텍스트 데이터를 대상으로 비정형 및 반정형 데이터에 대하여 자연어 처리기법과 문서 처리 기법을 이용하여 대규모의 텍스트 데이터에서 유용한 정보를 추출 및 가공을 위한 목적으로 활용된다[12,13]. 즉, 인터넷 환경을 포함한 텍스트 중심의 온라인 기사, 정보, 사용자 댓글 등을 포함하며 그 외 신문, 여론조사, 이메일 등 다양한 텍스트를 중심으로 비정형 데이터를 DB(Data Base)화하는 방법으로 텍스트 자료처리 과정과 분석과정을 거친 후 추세와 패턴 분석으로 텍스트 데이터의 어휘, 의미 등을 추출하여 활용 및 분석하는 과정을 의미한다[13].

3. 연구 방법 및 절차

3.1 연구의 절차

본 연구의 절차는 Fig. 2와 같이 진행하였다.

논문 검색 데이터베이스를 우선 선정하고, 분석 대상 데이터를 수집하였다[14]. 분석 대상으로는 논문 제목 및 초록 수집이 가능한 해외 학술 제공 DB를 선정하였다. 수집된 데이터는 전처리 과정 후 데이터에 대한 분석을 수행하였다. 데이터 분석을 위해 텍스트 마이닝과 네트워크 분석 등을 수행하였다. 다음으로 결과에 대한 시각화를 진행하였다[15,16].

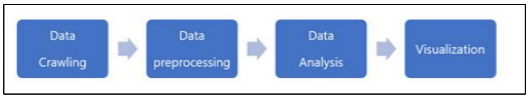

Fig. 2. Research Procedure

데이터 전처리를 위해 빅데이터 분석 도구인 R을 활용하여 R에서 제공되는 다양한 분석 패키지를 활용하여 분석하였다.

3.2 연구방법

본 연구 진행을 위해 해외논문을 제공하는 사이트 중에서 SCOPUS 데이터베이스에서 논문 제목을 'cloud computing'으로 한정하여 검색을 수행하였다[14].

우선 R을 활용하여 데이터를 수집하였다. SCOPUS에서는 23,116개까지 논문의 초록 및 제목 등 서지 제공을하고 있다. 수집된 데이터를 정제하고 전처리 후 데이터에 대한 분석 수행을 위해 R과 LDA를 활용하였다. 분석 수행 후 결과에 대하여 시각화를 진행하였다.

전처리 과정에서 데이터를 'article'로 한정을 한 후 8,750개, 'business', 'socia'l, 'material', 'decision science'로 한정하여 최종적으로 1,798개의 데이터를 분석에 사용하였다.

4. 분석 결과

Subject Area는 Table 2와 같다. 'computer science' 17,871편으로 가장 많은 분포를 보이고 있으며, 그다음으로는 'engineering', 'mathematics', 'decision science' 순이다.

Table 2. Subject Area

Subject Area	paper
computer science	17,871
engineering	8,382
mathematics	3,906
decision science	1,720
social science	1,458
business, management and accounting	1,247

Keyword를 살펴보면, Table 3에서 보는 결과와 같이 'Cloud Computing' 17,875편, 'Distributed Computer Systems' 2,166편, 'Cloud Computing Environments' 2,060편, 'Digital Storage' 1,889편, 'Network Security' 1,567편, 'Mobile Cloud Computing' 1,562편, 'Computer Systems' 1,397편 순이다.

Table 3. Number of Keyword papers

Keyword	paper
Cloud Computing	17,875
Distributed Computer Systems	2,166
Cloud Computing Environments	2,060
Digital Storage	1,889
Network Security	1,567
Mobile Cloud Computing	1,562
Computer Systems	1,397

4.2 분석결과

토픽 분석 결과는 다음과 같다. Fig. 3을 살펴보면, 서로 유사한 토픽은 없는 것으로 나타났다. 1번 토픽에서 중요 키워드로는 'cloud', 'computing', 'data', 'proposed', 'schedulin' 등이 있으며, 토픽의 주제는 'cloud' 임을 유추할 수 있다.

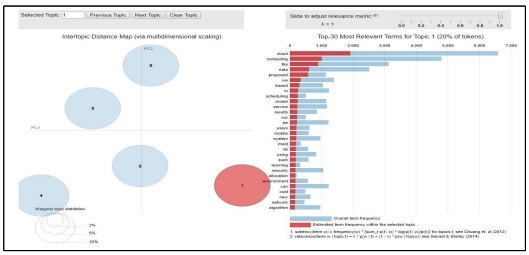

Fig. 3. LDA analysis results 1

Fig. 4를 분석한 결과 2번 토픽에서 중요 키워드로는 'cloud', 'data', 'computing', 'research', 'model' 등이 있으며, 토픽의 주제는 'cloud와 data' 임을 유추할 수 있다.

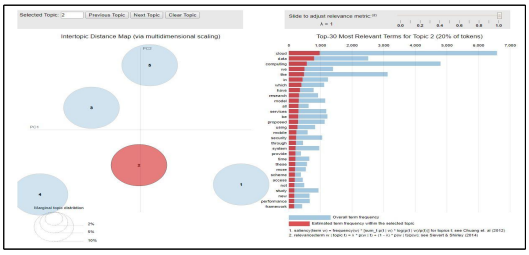

Fig. 4. LDA analysis results 2

Fig. 5의 분석 결과, 4번 토픽에서 중요 키워드로는 'computing', 'cloud, paper', 'resources', 'security', 'service' 등이 있으며, 토픽의 주제는 'computing' 임을 유추할 수 있다.

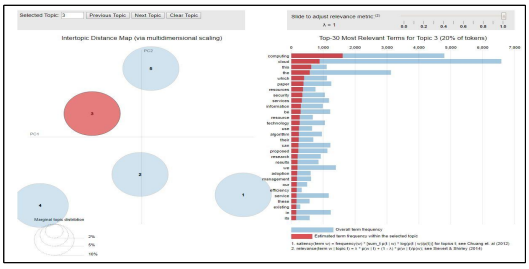

Fig. 5. LDA analysis results 3

Fig. 6을 분석해 본 결과 4번 토픽에서 중요 키워드로는 'cloud', 'computing', 'data', 'paper', 'srevice',

'information' 등이 있으며, 토픽의 주제는 'cloud' 임을 유추할 수 있다.

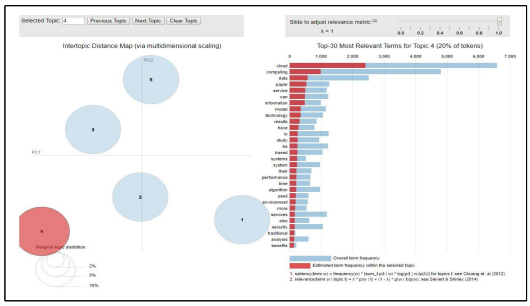

Fig. 6. LDA analysis results 4

Fig. 7에서의 분석 결과로는 6번 토픽에서 중요 키워드로 'computing', 'data', 'based', 'cloud', 'study' 등이 있으며, 토픽의 주제는 'computing' 임을 유추할 수 있다.

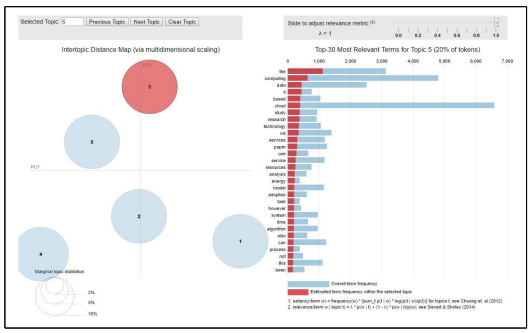

Fig. 7. LDA analysis results 5

5. 결론

본 연구에서는 'cloud computing'의 연구 동향을 살펴보고자 해외 논문 검색사이트인 SCOUPS 에서 논문 제목을 'cloud computing'으로 한정하고 검색하여 연구 동향을 파악하였다.

본 연구의 주요 분석 결과는 다음과 같다. 클라우드 컴퓨팅에 관한 연구로 'computer science' 17,871편으로 가장 많은 분포를 하고 있으며, Keyword를 살펴본 결과 'Cloud Computing' 17,875편으로 가장 많이 도출되었다. 또한 토픽 별 모델링 분석 결과로 'cloud', 'cloud'와 'data', 'computing'이 가장 중요한 토픽으로 도출되었다.

클라우드 컴퓨팅에 대한 높은 관심으로 수많은 선행 연구가 진행 되었지만 동향에 대한 연구에서는 미래에 미치는 영향, 빠르게 변화하는 디지털 시대에 새로운 패러다임 기술들에 대하여 클라우드 컴퓨팅 진화에 미치는 영향을 중심으로 개념적 모델 및 사례를 통한 대응 방안 모색 등의 연구가 대부분이다. 이에 본 연구는 기존 연구와 다르게 클라우드 컴퓨팅 동향에 관한 연구 동향 파악을 위해 R과 LDA를 활용하여 토픽에 대한 모델링으로 분석 수행한 점에서 기존 연구와의 차별성에 의의가 있다고 할 수 있다. 또한 클라우드 컴퓨팅에 대한 성장 가속화로 다양한 분야에서 지속 성장과 생존 전략을 위한 대표전략 중 하나인 만큼 본 연구 결과는 정부, 공공기관, IT산업 분야 등 다양한 분야에 클라우드 컴퓨팅 동향을 파악하고 실무에 적용함으로써 실무적인 측면에서의 밑거름을 제공할 수 있는 지침이 될 수 있을 것이다.

본 연구는 클라우드 컴퓨팅 연구 동향에 대한 미래 지향적으로 전략적 방안 수립을 위한 연구로써 의미를 가진다. 그러나 해외논문 검색사이트인 SCOUPS 논문으로 한정되어 진행된 만큼 일반화에 대한 제한적인 부분과 국내 동향을 파악하기는 어려운 점이 있다. 또한 클라우드 컴퓨팅 동향을 파악할 수 있지만, 미래 정책 수립에 대한 기초자료 활용에 한계가 있다. 국내뿐만 아니라 SCOUPS 논문 이외에 다양한 해외 학술 연구 DB의 논문을 추가하여 국내외 동향 분석과 더 나아가 뉴스, 신문기사, 소셜네트워크 등 폭넓은 데이터 분석으로 클라우드 컴퓨팅 동향에 대한 연구가 필요할 것으로 보인다.

본 논문이 제시하고 있는 연구의 한계점을 밑거름으로 향후 클라우드 컴퓨팅 동향에 대하여 보완되고 다양한 연구가 이루어진다면 클라우드 컴퓨팅 관련 산업들에 대한 활성화에 기여하여 더욱더 의미가 있는 연구 결과에 대하여 제시할 수 있을 것이다.

REFERENCES

[1] S. M. Yoo. (2017). *The era of the 4th industrial revolution is just around the corner. Success depends on the ICBM platform,* Dong-A Business Review. https://dbr.donga.com/article/view/1206/article_no/8015#header

[2] D. H. Kim. (2021). *"The Cloud Is The New Digital Experience Center" Gartner.* CIO KR. https://www.ciokorea.com/newsz

[3] P. Mell & T. Grance. (2011). The NIST definition of cloud computing., *NIST Special Publication,* 800-145.

[4] J. W. Park. (2011). *A Study on the Selection Method of Cloud Computing Service Using AHP*. Master's dissertation. MYONGJI University, Seoul.

[5] S. Marston, Z. Li, S. Bandyopadhyay, J. Zhang & A. Ghalsasi. (2011). Cloud Computing-The Business Perspective. *Decision Support Systems, 51(1),* 176-189.

[6] Wikipdia. (2022). *Cooud Computing*. Wikmedia project.

[7] KISA. (2013), *Recent Cloud Computing Service Trends*. Naju. KISA. All rights reserved.

[8] T. G. Kang, (2021). *A Study On RelativeImportance of Korea Cyber University Cloud Computing System Success Factors*. Master's dissertation. Chungbuk National University, Chungcheongbuk-do Cheongju.

[9] S. S. Gill et al. (2019). Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things, 8,* 100118. DOI : 10.1016/j.iot.2019.100118

[10] A. AlTwaijiry. (2021). *Cloud Computing Present Limitations and Future Trends*. ScienceOpen Preprints.

[11] A. A. Abbasi, A. Abbasi, S. Shamshirband, A. T. Chronopoulos, V. Persico & A. Pescapè. (2019). Software-defined cloud computing: A systematic review on latest trends and developments. *IEEE Access, 7,* 93294-93314.

[12] T. H. Jo. (2001). Concept and Application of Text Mining. *Journal of Scientific & Technological Knowledge Infrastructure, 5,* 76-85.

[13] J. H. Kim & S. S. Kim. (2021). A Study on the Analysis of Agricultural R&D Keywords Using Textmining Method. *Journal of the Korea Academia-Industrial cooperation Society, 22(2),* 721-732.

[14] E. M. Park & J. H. Seo. (2019). Effects of shared leadership, psychological empowerment and organizational justice on organizational commitment. *Journal of Digital Convergence, 17(6),* 177-184. DOI : 10.14400/JDC.2019.17.6.177

[15] S. T. Park & Y. K. Kim. (2019). A study on deriving an optimal route for foreign tourists through the analysis of big data. *Journal of Convergence for Information Technology, 9(10),* 56-63. DOI : 10.22156/CS4SMB.2019.9.10.056

[16] G. Li, J. S. Dai, E. M. Park & S. T. Park. (2017). A study on the service and trend of Fintech security based on text-mining: focused on the data of Korean online news. *Journal of Computer Virology and Hacking Techniques, 13(4),* 249-255. DOI : 10.1007/s11416-016-0288-9

강 태 구(Tae-Gu Kang) [정회원]

· 2007년 2월 : 충북대학교 경영정보학과(경영학석사)
· 2021년 2월 : 충북대학교 경영정보학과(경영학박사)
· 2017년 4월 ~ 현재 : 건양사이버대학교 IT비즈니스학과 교수
· 2018년 3월 ~ 현재 : 건양사이버대학교 정보통신 원장
· 관심분야 : 클라우드 컴퓨팅, 빅데이터, 경영정보, ERP
· E-Mail : tgkang@kycu.ac.kr