

빅데이터, IoT, 인공지능 키워드 네트워크 분석

구영덕*

Analysis on Big data, IoT, Artificial intelligence using Keyword Network

Young-Duk Koo*

요약

본 논문에서는 빅데이터, IoT, 인공지능 관련 네트워크 분석을 통해 국내 연구동향을 파악하고 관련 시사점 도출을 목적으로 한다. 이를 위해, 2018년 국가연구개발정보를 활용하여 분석을 수행하였으며, 주요 기초 통계 분석과 언어 네트워크 분석을 수행하였다. 분석 결과, 빅데이터, IoT, 인공지능 관련 연구개발은 기초단계, 개발단계를 중심으로 연구가 진행 중이며, 대학과 중소기업의 비중이 높은 것으로 나타났다. 또한 언어 네트워크 분석 결과, 관련 분야는 스마트팜, 헬스케어 분야에 활용하기 위한 연구를 중심으로 이루어지고 있는 것으로 판단된다. 이러한 연구결과를 바탕으로 본 연구에서는 인공지능을 활용하기 위해서는 빅데이터가 반드시 필요하며, 개인 식별화 연구가 더욱 활발히 진행되어야 한다는 점과 단순 R&D 활동이 아닌 기술 사업화가 이루어지기 위한 전 주기 지원이 필요하며, 적용 분야를 확대할 필요가 있다는 점을 주장하였다.

ABSTRACT

This paper aims to provide strategic suggestions by analyzing technology trends related to big data, IoT, and artificial intelligence. To this end, analysis was performed using the 2018 national R&D information, and major basic analysis and language network analysis were performed. As a result of the analysis, research and development related to big data, IoT, and artificial intelligence are being conducted by focusing on the basic and development stages, and it was found that universities and SMEs have a high proportion. In addition, as a result of the language network analysis, it is judged that the related fields are mainly research for use in the smart farm and healthcare fields. Based on these research results, first, big data is essential to use artificial intelligence, and personal identification research should be conducted more actively. Second, they argued that full-cycle support is needed for technology commercialization, not simple R&D activities, and the need to expand application fields.

키워드

AI, Bigdata, IoT, Network
인공지능, 빅데이터, 사물인터넷, 네트워크

1. 서론

최근 4차 산업혁명을 필두로 하여 차세대 산업 동력을 찾는 데 범국가적으로 몰두하고 있다. 이 4차 산

업혁명의 핵심은 결국 IoT, 빅데이터 그리고 인공지능 기술의 융합으로 산업계 데이터를 이용하여 어떻게 하면 더 효율적이고 자동화된 생산 공정을 만들 수 있는지에 중점을 두고 있다. IoT를 활용한 데이터 수

* 교신저자: 한국과학기술정보연구원
• 접수일 : 2020. 10. 20
• 수정완료일 : 2020. 11. 18
• 게재확정일 : 2020. 12. 15

• Received : Oct. 20, 2020, Revised : Nov. 18, 2020, Accepted : Dec. 15, 2020
• Corresponding Author : Young-Duk Koo
Korea Institute of Science and Technology Information
Email : ydkoo@kisti.re.kr

집, 빅데이터 기술을 이용한 실시간 데이터 저장 그리고 인공지능 기술을 활용한 분석, 분류, 예측 기반의 지능형 시스템을 구축하는 것이다[1]. 최근 정부는 이러한 빅데이터, 인공지능, IoT 분야를 디지털 뉴딜 정책의 한 축으로 삼고 육성 전략을 수립하고 있다. 특히 인공지능 학습 데이터 구축, 빅데이터 플랫폼 센터 구축, AI 데이터 가공 바우처, IoT AI 기반 신 데이터 구축 등을 적극 추진하고 있다. 학술적으로도 신경망 인식, 센서 네트워크 등 다양한 분야에서 연구가 진행되고 있다[2-3].

본 논문에서는 국가연구개발사업 리스트를 중심으로 주요 키워드를 분석함으로써, 국내 빅데이터, IoT, 인공지능 관련 공유된 의미를 파악하였다. 과학기술 정책이란 언어로 형성되는 것으로서, 실제 연구개발 추진에 있어 다양한 연구주제를 주요 키워드로서 자신들의 주요 연구내용을 공유한 것이라 할 수 있다. 연구의 키워드는 네트워크의 형태로서 시각화가 가능하고, 네트워크 분석을 통하여 표면에 드러난 의미 파악 뿐만 아니라 내면에 숨겨져 있는 의미 파악이 가능하다. 언어구조를 공간적으로 표시함에 따라, 키워드에 나타난 주요 개념과 다른 개념과의 관계를 파악 가능한 것이다.

이를 위해, 본 논문 II장에서는 주요 분석방법론으로 네트워크 분석 방법과 VOSVIEWER에 대한 소개, III장에서는 주요 분석 결과, IV장 및 V장에서는 분석 결과 검토 및 결론을 도출하였다.

II. 분석 방법론

2.1 네트워크 분석

본 연구에서는 빅데이터, IoT, 인공지능 분야 연구 트렌드를 분석하기 위해 언어네트워크 방법을 활용하였다. 일반적으로 네트워크 분석은 각 노드들 사이에 맺고 있는 ‘관계’를 통하여 네트워크를 검토한다[4]. 사회네트워크적 시각에서 메시지의 내용분석을 함에 있어서 중요한 요소는 핵심 단어간 의미론적 연관이다. 최근 들어 텍스트에 기반한 양적 내용분석의 새로운 유형으로 언어네트워크 분석에 대한 연구가 진행되고 있다. 언어네트워크 분석이란 텍스트에 출현하는 단어와 단어 사이의 관계를 링크로 구성된 네트워크

를 해석하는 분석기법이다[5]. 이는 언어와 지식은 단어사이의 관계를 통해 형성된 네트워크라는 인식에 근거한다. 즉, 언어네트워크 분석은 사회네트워크 분석 기법과 텍스트 분석을 복합적으로 연계함으로써 특정연구 분야 및 현상에 대한 지식과 이해를 증진하고자 하는 시도로 이해할 수 있다[6].

언어네트워크 분석을 활용하여 정책을 연구한 국내 연구 사례는 다수 존재한다. 대표적인 연구로는 이명박 정부와 노무현 정부의 정권 초기 대통령 연설문을 대상으로 언어네트워크 분석을 수행하였다. 또한, 새마을운동 정책을 분석하였고, 여성 정책을 분석한 바 있다. 또한, 뉴스사이트에서 방통 융합환경에서의 주파수 정책 이슈를 분석한 사례도 있다[7]. 그러나, 우리나라에서 언어 네트워크 분석한 연구 중 과학기술 분야의 정책연구에 적용한 사례가 거의 전무하므로 과학기술 정책 분석에 새로운 방법론을 적용한 것에 의의가 존재한다.

본 연구에서는 네트워크 지표 중에서 융합 지수를 활용하여 얼마나 다양한 키워드가 융합되고 있는지에 대해 분석을 수행하였다. 융합과 관련된 지표는 많은 연구가 진행 중이며, 특히 네트워크 지표(연계 중심성 기반), 저널 지표(Shannon Entropy, Jini Index), 그리고 최근 Rao-Stirling 지수가 제시되고 있다. 특히 Shannon Entropy와 Jini Index는 저널의 인용 통계 및 색인 인용에 대해 동시발생 행렬을 사용하여 측정하게 된다. 본 연구에서 사용된 Jini Index는 각 노드에서의 불순도(impurity) 또는 다양도(diversity)를 측정하는 척도 중 하나로서 식(1)과 같이 정의한다 [8-9].

$$G-1 = \sum_{n=1}^n \left(\frac{n_j}{n} \right)^2 \quad (1)$$

여기서 n 은 전체노드개수 n_j 는 해당 노드 개수이다.

2.2 VOSVIEWER

VOSviewer는 Eck and Waltman(2010)에 의해 개발된 소프트웨어로 주로 네트워크를 맵형태로 구현하기 위해 개발되었다. VOSviewer는 다양한 관계에 대해 동시 발생행렬표를 구현하고, 이를 통해 맵을 구현한다. VOSviewer 구현은 3가지 단계를 거친다. 첫 번째 단계는 동시 발생행렬표를 기반으로 카테고리간

유사도를 계산한다. 유사도 계산에 사용된 Association strength 식은 식 (2)와 다음과 같다.

$$Association\ strength = \frac{m C_{ij}}{C_{ii} C_{jj}} \quad (2)$$

여기서 C_{ij} 는 i 와 j 간 발생 회수를 의미하며, C_{ii} 와 C_{jj} 는 각각 i 와 j 의 발생회수를 의미한다. 또한 m 은 전체 모수를 의미한다.

두 번째 단계는 첫 번째 단계에서 계산된 유사도를 기반으로 맵을 구현한다. 즉 유사도가 높은 관계는 맵 상에서 인접하며, 유사도가 낮은 관계는 멀리 떨어져 있다. 마지막 단계는 노드들을 클러스터링하고 밀도를 구분하여 표시한다[10].

III. 국가 연구개발 분석 결과

3.1 국가연구개발사업 기초 통계 분석결과

빅데이터, IoT, 인공지능 분야 2018년 국가연구개발사업에 대한 연구수행주체는 주로 중소기업과 대학을 중심으로 연구가 진행되고 있는 것으로 나타났다.

[Unit : Number, %]

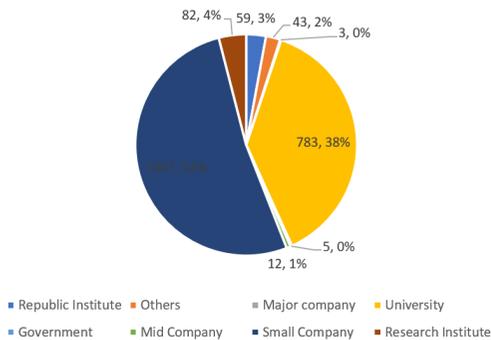


그림 1. 빅데이터, IoT, 인공지능분야 국가연구개발사업 연구수행 주제 비중
Fig. 1 Ratio of subjects performing national R&D projects in the fields of big data, IoT, and artificial intelligence

빅데이터, IoT, 인공지능 분야 연구개발 단계는 기초연구와 개발연구를 중심으로 진행되는 것으로 나타났다.

[Unit : Number, %]

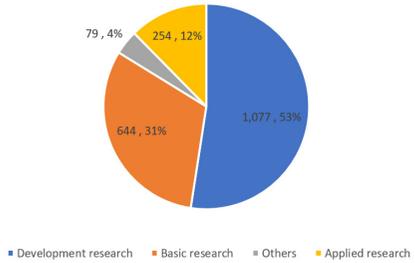


그림 2. 빅데이터, IoT, 인공지능분야 국가연구개발사업 연구개발단계 비중
Fig. 2 Ratio research and development stages of national R&D projects in the fields of big data, IoT, and artificial intelligence

빅데이터, IoT, 인공지능 분야 경제사회목적별로는 산업생산 및 기술에 적용하기 위한 연구가 가장 활발히 진행되는 것으로 나타났다.

[Unit : Number]

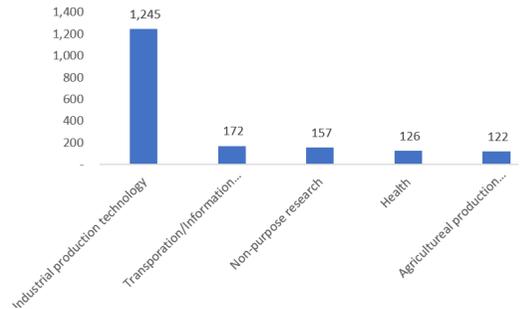


그림 3. 빅데이터, IoT, 인공지능분야 국가연구개발사업 경제사회목적별 과제 수
Fig. 3 Number of projects by economic and social purpose of national research and development projects in the fields of big data, IoT, and artificial intelligence

빅데이터, IoT, 인공지능 분야 연구수행 기관 상위 5개 지역을 살펴보면, 한국과학기술원, 한국전자통신연구원, 서울대학교가 가장 많은 연구를 진행하고 있는 것으로 나타났다.

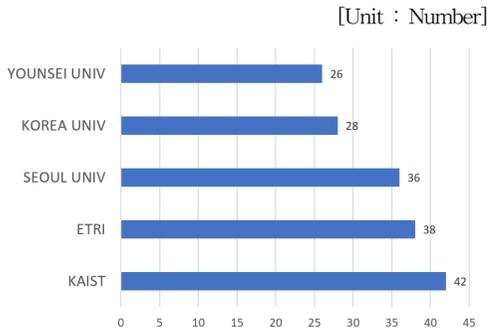


그림 4. 빅데이터, IoT, 인공지능분야 국가연구개발사업 연구수행 기관 상위 5개 기관
Fig. 4 Top 5 institutions for national R&D projects in the fields of big data, IoT, and artificial intelligence

빅데이터, IoT, 인공지능 분야 수행 형식은 대부분 자유 공모 형식으로 정부 사업을 발주하여 관련 연구를 진행하고 있는 것으로 나타났다.

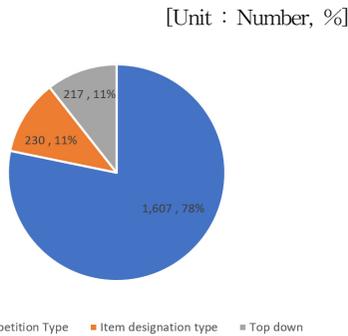


그림 5. 빅데이터, IoT, 인공지능분야 국가연구개발사업 수행 방식 비중
Fig. 5 Ratio of national R&D projects in the fields of big data, IoT, and artificial intelligence

빅데이터, IoT, 인공지능 분야 연구기관 유형별 공동연구는 대학과 공동연구를 가장 활발히 진행하고 있으며, 기업과도 공동연구를 상대적으로 활발히 진행하고 있는 것으로 나타났다.

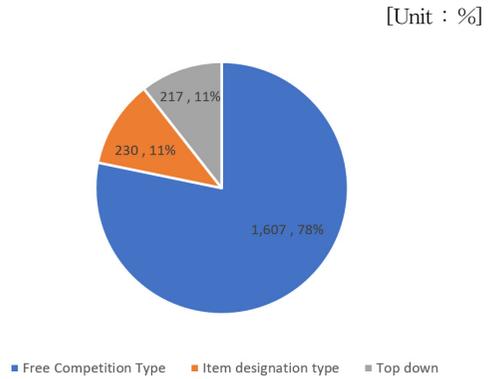


그림 6. 빅데이터, IoT, 인공지능분야 국가연구개발사업 공동연구 비중
Fig. 6 Share of joint research on national R&D projects in the fields of big data, IoT, and artificial intelligence

3.2 빅데이터, IoT, 인공지능 네트워크 분석 결과

빅데이터, IoT, 인공지능 관련 국가연구개발사업 키워드 빈도 수 분석 결과, 빅데이터, 인공지능 키워드가 가장 많은 빈도 수를 기록하고 있으며, 이를 중심으로 한 딥러닝, 기계학습, 머신러닝이 가장 높은 것으로 나타났다.

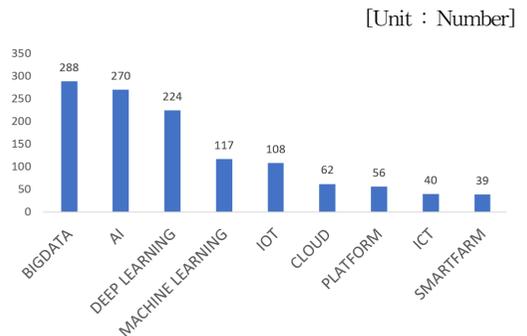


그림 7. 빅데이터, IoT, 인공지능분야 국가연구개발사업 상위 10대 키워드 빈도 수
Fig. 7 Frequency of the top 10 keywords for national R&D projects in the fields of big data, IoT, and artificial intelligence

키워드 네트워크 상에서 빈도수가 높은 상위 5개 키워드에 대해 jini index값을 분석한 결과, 빅데이터, 인공지능 관련 키워드가 가장 높은 것으로 나타났다. 즉, 관련 키워드가 빅데이터와 인공지능 분야에 있어

가장 다양한 키워드를 가지고 연구를 진행하는 것으로 분석되었다.

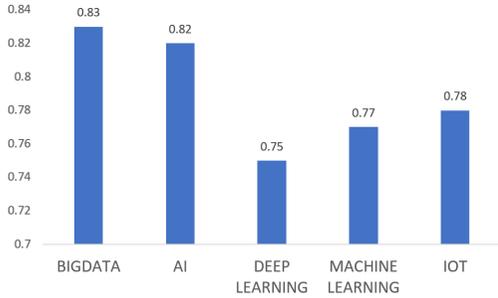


그림 8. 빅데이터, IoT, 인공지능분야 상위 5개 키워드에 대한 jini index 값

Fig. 8 Jini index value for the top 5 keywords in the fields of big data, IoT, and artificial intelligence

연관맵 상에서 빅데이터 키워드와 유사도가 높은 상위 5개 키워드를 살펴보면, 인공지능, 스마트팜이 유사도가 높은 것으로 나타났다. 인공지능과 유사도가 높은 키워드 역시, 빅데이터이며, 이외에 딥러닝, 챗봇 등 의료 서비스 분야에 대한 유사도가 높은 것으로 나타났다.

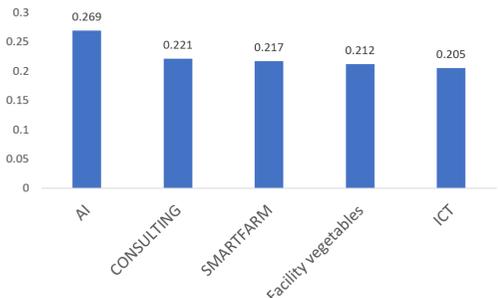


그림 9. 빅데이터 관련 키워드 상위 5대 Pearson 상관계수 값

Fig. 9 Top 5 Pearson correlation coefficient values for big data related keywords

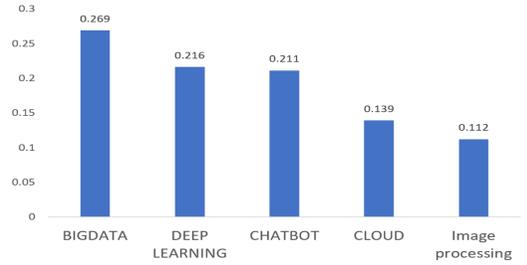


그림 10. 인공지능 관련 키워드 상위 5대 Pearson 상관계수 값

Fig. 10 Top 5 Pearson correlation coefficient values for artificial intelligence related keywords

3.3 VOSVIEWER 분석 결과

그림 11은 빅데이터, IoT, 인공지능 분야 국가연구개발사업 발사업 키워드간 연관맵을 구현한 것이다. 그림 11에서와 같이, 빅데이터 인공지능 관련 연구에서는 빅데이터, 인공지능을 연계하기 위한 연구가 가장 활발히 진행 중이며, 세부적으로 딥러닝, 인공지능경망, 데이터 베이스 등 빅데이터 기반 인공지능 상용화를 위한 연구개발이 가장 활발한 것으로 분석된다. 또한, 이를 활용하는 분야에 있어서는 스마트팜에 인공지능과 빅데이터를 적용하기 위한 연구와 반려동물 케어를 위해 인공지능과 빅데이터를 결합하기 위한 연구 역시 활발히 진행 중인 것으로 나타났다. 사물인터넷 분야의 경우 헬스케어 분야에 적용하기 위한 연구가 가장 활발히 진행 중이며, 이를 위한 빅데이터 분석 관련 연구도 활발히 진행 중인 것으로 나타났다. 이외에도, 인공지능과 빅데이터 분야를 활성화하기 위해서는 데이터 보안 문제가 중요한 이슈이며, 이를 위한 비식별화 관련 연구도 진행되고 있는 것으로 분석되었다.

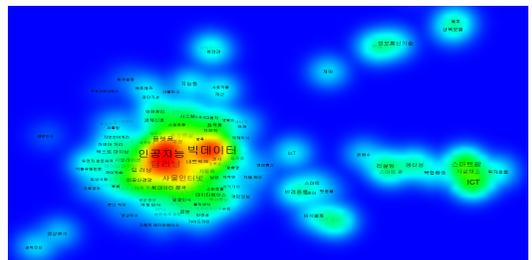


그림 11. 빅데이터, IoT, 인공지능 분야 국가연구개발사업 키워드 연관맵

Fig. 11 Big data, IoT, artificial intelligence field national R&D project keyword related map

IV. 분석결과 검토

본 연구에서는 2018년에 추진한 국가연구개발사업에서 빅데이터, 인공지능, IoT 관련 추진 사업에 대해 기초분석 및 네트워크 분석을 통해 관련 동향을 분석하였다.

분석 결과, 빅데이터, 인공지능, IoT 관련 기술은 정부에서 자유 공모 형식으로 주로 기초연구와 개발 연구를 중심으로 연구개발을 추진 중에 있으며, 기초연구의 경우 대학을 중심으로, 개발연구의 경우 중소기업을 중심으로 연구개발을 진행하고 있는 것으로 분석된다. 주요 연구기관으로는 한국과학기술원, 한국전자통신연구원 등 출연연에서 많은 연구를 진행하고 있으나, 규모 면에서는 중소기업이 활발히 연구를 진행하고 있는 것으로 나타났다. 또한 국가연구개발 상에서 공동연구는 중소기업과 대학을 중심으로 가장 활발히 진행되고 있는 것으로 분석되었다.

빅데이터, 인공지능, IoT 관련 연구개발 키워드를 분석한 결과, 가장 빈도 수가 높은 키워드는 빅데이터, 인공지능을 제외하면 딥러닝, 클라우드, ICT, 스마트팜 등이며, 이는 관련 연구가 가장 활발히 적용되는 분야는 스마트 팜이다. 키워드 간 연관법에 대해 분석한 결과, 인공지능과 빅데이터를 융합한 연구와 스마트팜 등 농업 분야에 이를 적용하기 위한 연구와 반려동물 관련 산업에 있어 IoT 기술을 접목한 연구 그리고 헬스케어 중심으로 연구가 진행되고 있는 것으로 분석된다. 특히 인공지능 분야의 경우에는 스마트팜 관련 기술과 연관성이 높으며, 빅데이터 분야의 경우, 챗봇 등 헬스케어 분야와 연계성이 높은 것으로 판단된다. 특히 챗봇 분야의 경우 인공지능과 빅데이터 기술을 결합한 연구성과가 가장 필요한 분야로 판단된다.

키워드 간 유사도를 분석한 결과에 있어 가장 큰 특징은 인공지능, 빅데이터 관련 연구에서 인공지능에서는 빅데이터가, 빅데이터에서는 인공지능 키워드의 연관성이 가장 높은 것으로 나타났다. 이는 빅데이터가 없는 인공지능은 무의미하다는 것을 의미한다. 인공지능은 수십년 동안 존재해 왔지만, 인공지능의 원재료인 데이터가 폭발적으로 증가함에 따라 엄청난 속도로 발전할 수 있게 되었다는 것이다.

V. 결론

이러한 분석 연구결과를 바탕으로, 본 연구에서 제안하는 사항은 다음과 같다. 첫 번째, 인공지능 관련 연구 추진에 있어 빅데이터 분석은 반드시 필요한 사항으로, 이를 추진하기 위해서는 개인 식별화에 대한 연구가 더욱 활발히 진행될 필요가 있으며, 특히 정부에서 최근 통과된 데이터 3법을 기점으로 이를 더욱 가속화하여 연구를 집중할 필요가 있다. 두 번째, 관련 연구에 있어 기초연구는 대학에서, 개발연구 부분은 중소기업에서 집중하고 있는 상황에서 대학과 중소기업 간 공동연구를 통해 단순히 R&D에 그치는 것이 아니라 기술사업화까지 이루어 질 수 있도록 전 주기 지원방안이 필요하다. 세 번째로, 현재 추진되고 있는 스마트 팜 등 농업분야 이외에도 헬스케어 분야 등 공공 분야에서 적용이 가능한 분야로 과제 지원을 확대할 필요가 있다. 네 번째로, 엄청난 데이터를 모으기 위해서는 정부에서 추진하는 신 데이터댐 구축을 활용하여 해당 지역을 중심으로 관련 산업을 직접화할 필요가 있다.

본 연구는 국가연구개발사업 자료를 바탕으로, 빅데이터, 인공지능 및 IoT 분야 연구 동향을 파악하였다는 점과 데이터 기반 관련 R&D 육성방향을 제시하였다는 점에서 의의가 있다. 그러나, 2020년 본격적으로 추진되기 시작한 디지털 뉴딜 정책에 따른 정부정책을 반영하지 못한다는 한계가 있다. 따라서, 향후 연구에서는 2020년 데이터 구축이 완료된 이후, 이를 추가적으로 분석하여, 관련 분야 동향을 확인하고 해당 전략을 조정하는 것이 필요하다.

감사의 글

본 논문은 2020년도 한국과학기술정보연구원의 지역특화 산학연정 협력 생태계 운영 (K-20-L03-C06-S01) 사업의 지원으로 수행되었습니다.

References

- [1] Y. Ji, J. Yu, and S. Lee, "IoT, Bigdata and AI," *Communications of the Korean Institute of Information Scientists and Engineers*, vol. 35, no. 7, 2017. pp. 43-50.
- [2] S. Jang, "Design and Implementation of an Efficient Communication System for Collecting Sensor Data in Large Scale Sensors Networks " *The Journal of the Korea Institute of Electronic Communication Sciences*, vol. 24, no. 1, 2020, pp. 113-119.
- [3] C. H. Yoon, "The study of Authorized / Unauthorized Vehicle Recognition System using Image Recognition with Neural Network " *The Journal of the Korea Institute of Electronic Communication Sciences*, vol. 15, no. 2, 2020, pp. 299-306.
- [4] S. Wasserman, and K. Faust, *Social networks analysis: methods and applications*. Oxford, UK: Cambridge University Press, 1994.
- [5] R. Popping, *Computer-assisted Text Analysis*. London: Saga, 2000.
- [6] Y. Choi and S. Park, "Research Trends in the Field of Local Government and Administration - An Analysis of Articles Published in Korean Society and Public Administration, 1991-2012 - ," *Korean Journal of Public Administration*, vol. 45, no. 1, 2011, pp. 123-139.
- [7] E. Kim, "A Comparative Study on the Spectrum Policy Issues in Convergence Environment Semantic Network Analysis of News Coverage in the U.S. and S. Korea," *International Telecommunications Policy Review*, vol. 17, no. 4, 2010. pp. 107-139.
- [8] E. Kim, "An Efficient Hierarchical Multi-Hop Clustering Scheme in Non-uniform Large Wireless Sensor Networks," *Computer Science and its Applications*, vol. 39, no. 3, 2012. pp. 248-257.
- [9] A. Clauset, M.E.J.NEWMAN, Cristopher Moore, "Finding community structure in very large networks," *Physical Review*, 2004. pp. 69-70.
- [10] N. J. Van Eck, L. Waltman, R. Dekker, J. van den Berg, "A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS," *Journal of the American Society for Information Science and Technology*, vol. 6, no. 1, 2010, pp. 2405-2416.

저자 소개

**구영덕(Young-Duk Koo)**

1988년 2월 인천대학교 기계공학과 (공학사)

1990년 2월 인천대학교 기계공학과 (공학석사)

1998년 2월 인천대학교 기계공학과(공학박사)

1991년~현재 : 한국과학기술정보연구원 책임연구원

※ 관심분야 : 미래 유망기술, 계량정보분석, 기술시장분석

