

‘블록체인 활용’ 관련 빅데이터를 활용한 토픽 분석: 신문 기사를 중심으로

김성애¹, 전수진^{2*}

¹운암중학교 기술교사, ²호서대학교 혁신융합학부 교수

Topic Analysis Using Big Data Related to 'Blockchain usage': Focused on Newspaper Articles

Sungae Kim¹, Soojin Jun^{2*}

¹Technology Teacher, Div. of Technology Education, Woonam Middle school

²Assistant Professor, Dept. of Innovation and Convergence, Hoseo University

요약 이 연구에서는 블록체인 기술의 활용과 관련된 주요 토픽을 분석하기 위해 신문기사에 나타난 ‘블록체인 기술 활용’ 빅데이터를 토픽 모델링기법을 적용하였다. 이를 위해 2013년부터 2019년까지, 21개의 신문사로부터 15,617건을 대상으로 토픽을 추출하고 주요 트렌트를 시기별로 구분하여 분석하였다. 분석결과 블록체인 기술 활용과 관련된 기사는 2015년부터 기하급수적으로 증가하였으며 IT_과학 분야와 경제 분야에 집중되었다. 기간에 따라 차이는 있지만 암호화폐, 비트코인, 가상화폐와 관련된 키워드의 가중치가 높았다. 금융거래에 집중되었던 블록체인기술은 빅데이터, 사물인터넷, 인공지능으로 점차 확대되었다. 이에 따라 기업의 토픽 변화도 함께 이루어져 금융거래를 위한 은행에서 다양한 분야로 확대되면서 대기업과 글로벌기업으로 집중되었다. 이 연구를 통해 블록체인기술의 활용과 관련한 신문기사의 주요 토픽과 함께 이러한 토픽들이 어떠한 변화 추이를 보이고 있는지에 대해 확인할 수 있었다.

키워드 : 블록체인기술 활용, 빅데이터, 토픽분석, 신문기사, 토픽 모델링 기법

Abstract To analyze the main topics related to the use of blockchain technology, the Topic Modeling Technique was applied to the ‘Blockchain Technology Utilization’ big data shown in newspaper articles. To this end, from 2013 to 2019, when newspaper articles on the use of blockchain technology first appeared, the topics were extracted from 21 newspapers and analyzed by time to 15,537 articles. As a result of the analysis, articles related to the utilization of blockchain technology have increased exponentially since 2015 and focused on IT_science and economics. Key words related to cryptocurrency, bitcoin and virtual currency were weighted high, although they differed depending on time. Blockchain technology, which had focused on financial transactions, gradually expanded to big data, Internet of Things and artificial intelligence. As a result, changes in corporate topics were also made together to expand into various fields at banks for financial transactions, focusing on large and global companies. The study showed how these topics were changing, along with the main topics in newspaper articles related to the use of blockchain technology.

Key Words : Blockchain technology utilization, big data, topic analysis, newspaper articles, topic modeling techniques

1. 서론

세계 경제 포럼에서도 2027년 세계 GDP의 10%가

블록체인 기술을 기반으로 생산되고 2023년에는 정부가 블록체인(Block chain)으로 세금을 거두기 시작할 것으로 예측하고 있다[1]. 이러한 세계적인 흐름에 발

*Corresponding Author : Soojin Jun(soojin3587@gmail.com)

Received January 25, 2020

Revised February 18, 2020

Accepted February 20, 2020

Published February 29, 2020

맞추어 우리나라에서도 블록체인 기술과 관련된 정책적인 지원이 확대되고 있다. 과학기술정보통신부는 2020년도 업무계획에서 블록체인 기술을 집중 투자 대상으로 정하고 있음을 발표하였다. 즉, 제 4차 산업 혁명에 대응하는 디지털 국가가 되기 위한 선도 사업으로 블록체인 기술을 선정하였다고 할 수 있다. 과학기술정보통신부가 추진하는 블록체인기술의 주요 키워드는 대중적 서비스이다. 이는 블록체인기술을 다양한 행정 서비스에 활용함으로써 국민들이 블록체인 기술을 쉽게 체감할 수 있도록 하기 위함이다.

모든 사물이 연결되는 제 4차 산업혁명 시대에는 다양하고 방대한 양의 데이터에 대한 높은 보안성과 안정성이 매우 중요해지면서 블록체인기술이 주목받고 있다[2,3]. 블록체인기술은 P2P 네트워크를 기반으로 한 분산 장부 시스템이며[4] 비 대칭키 암호 체계를 사용하기 때문에[5] 보안성이 매우 뛰어나다. 특히 데이터를 위, 변조하기 위해서는 블록체인 참여자들 전체의 합의가 있어야 하기 때문에 위, 변조가 현실적으로 불가능하다[6]. 이러한 이유로 제 4차 산업 혁명을 대비하면서 블록체인 기술이 인공지능기술과 함께 두 축을 이루고 있다. 가상화폐 비트코인의 열풍으로 일반인 사이에 큰 관심을 불러일으킨 블록체인 기술은 현재는 금융 거래를 넘어 다양한 분야에서 연구가 수행되고 있다. 특히 이러한 토픽들은 다양한 매체를 통해 다루어지는데 그 중 신문은 대중생활과 밀접한 사건과 사고를 오랜 기간 동안 다루고 있기 때문에 특정 주제에 대해 오랜 기간에 걸쳐 트렌트를 분석하는데 매우 효과적이라고 알려져 있다[6].

이에 본 연구에서는 정부, 민간기업, 학계에 이르기까지 다양하게 활용되고 주목하고 있는 블록체인기술 활용과 관련된 빅데이터(Big Data)를 신문기사를 통해 수집하고 이를 토픽을 중심으로 분석하여 블록체인 기술의 활용에 대한 트렌트를 분석하고자 한다.

2. 선행연구

블록체인과 관련된 빅데이터를 활용한 토픽 분석에 대한 연구를 살펴본 결과는 다음과 같다.

강원지역을 대상으로 하여 블록체인과 관련된 주요 일간지와 지역신문을 대상으로 기사를 분석한 연구[7]에서는 블록체인 기술 도입에 대한 강원 지역의 주요 산업에 대한 핵심 토픽을 분석하였다. LDA기반의 토픽

픽모델링 등의 방법으로 분석한 결과 블록체인 기술 도입으로 인한 산업별 핵심 토픽의 연계성이 상이하여 산업별로 상이한 기대효과가 나타남을 확인하였다.

미국, 중국, 한국을 중심으로 블록체인과 관련된 연구 동향을 분석한 연구[8]에서는 LDA기반의 토픽모델링을 활용하였다. 그 결과, 미국은 경제와 금융, 중국은 기술과 시스템, 한국은 정책과 규제를 토픽으로 한 연구동향의 차이를 확인하였다.

블록체인 기술 동향을 탐색하되 특허와 관련된 토픽을 분석한 연구[9]에서는 LDA기반의 토픽모델링이 활용되었다. 전문가에 의해 결제 및 거래 관련 분야, 보안과 인증 관련 분야 12개의 특허 토픽을 결정하고 특허 요약문을 대상으로 분석한 결과 인증, 보안, 결제, 위변조 등과 관련된 기술이 주를 이루고 있었으며 그 중 인증 기술이 가장 큰 비중을 차지하고 있었다.

블록체인기술과 관련된 연구동향을 분석한 연구[10]에서는 전 세계의 논문을 대상으로 토픽분석을 실시하였다. 연구진에 의해 개발된 W2V-LSA를 활용하여 토픽분석을 하였으며 기존에 활용되고 있는 PLSA기반의 토픽분석결과에서 보이지 않았던 세부적인 토픽을 추출할 수 있었다. 기존의 분석방법과의 비교를 통해 연구진에 의해 개발된 분석방법에 대한 유용성을 입증하였다.

선행 연구를 분석한 결과 블록체인과 관련된 토픽 분석은 주로 LDA 기반의 토픽 모델링 기법이 활용되었다. 토픽 모델링은 대용량의 텍스트에 사용된 단어들을 이용하여 잠재되어 있는 토픽을 추출하고 그 토픽들이 서로 어떻게 관련되어 있고 어떻게 변화해 가는지를 분석할 수 있는 빅데이터 분석 방법이다[11]. 따라서, 본연구와 같이 신문기사와 같은 빅데이터의 토픽을 추출하고 이들의 연관성과 변화를 분석하는데 가장 효과적인 방법이라고 할 수 있다. 기존의 연구에서는 특정 지역에 국한되어 신문기사를 분석한 만큼 이를 전국으로 확대하여 국내의 경향을 확인하는 본 연구는 매우 의미 있는 연구라고 할 수 있다.

3. 연구 방법

3.1 연구 대상

본 연구에 활용된 분석 데이터는 블록체인 기술 활용과 관련된 기사가 처음 등장한 2013년부터 2019년까지의 신문기사이다. 이 연구에 포함된 신문사는 조

선일보, 중앙일보, 동아일보 등 중앙지, 파이낸셜뉴스, 한국경제, 매일경제 등 경제지, 디지털 타임스, 전자신문 등 전문지로서 총 21곳이다.

블록체인기술 활용과 관련된 신문기사를 수집하기 위해 한국언론재단 뉴스 검색 사이트인 빅카인즈(www.kinds.or.kr)를 이용하였다. 빅카인즈에서 ‘블록체인’과 ‘활용’이 모두 포함된 경우를 대상으로 신문 기사를 검색하였으며 최종적으로 총 15,617건이 수집되었다. 이 중 중복된 기사 323건을 제외한 15,294건을 분석대상으로 하였다. 연구 대상 신문사와 수집된 신문 기사 수는 Table 1과 같다.

Table 1. Newspapers Analyzed

Division	The number of newspapers	The number of collection
Central	11	2,327
Economic	8	10,085
Professional	2	3,205
total	21	15,617

3.2 연구 방법

본 연구 절차는 Fig. 1과 같이 분석데이터 수집, 데이터 전 처리, 토픽분석(LDA기반 토픽모델링, 워드클라우드), 결과 해석의 4단계로 나뉜다.

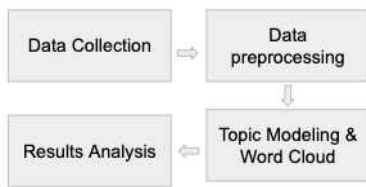


Fig. 1. Research Process

빈도수 분석은 어떠한 주제가 중요하게 다루어졌는지를 효과적으로 나타낼 수 있지만 단어의 빈도수만으로는 여러 분야의 이슈를 보여주는 데는 한계가 있다[12] 따라서 단어들의 빈도수를 통계적으로 분석하여 전체 데이터에서 토픽들을 자동으로 추출하여 분류할 수 있다는 점에서 토픽모델링 기법은 주제 분석시 매우 유용하다고 할 수 있다[13]. 또한 유의미한 데이터의 시간의 흐름에 따른 동향 파악에 효율적인 분석방법 중 하나로 알려져 있다[14]. 본 연구에서는 토픽모델링 기법 중 다른 기법에 비해 결과 해석이 쉽고 다양한 토픽들을 도출하는데 유리한 LDA기법

[15,16]을 활용하였다.

4. 연구 결과

4.1 신문기사 빈도 분석

Fig. 2는 연도별 신문기사 수의 변화를 그래프로 나타낸 것이다. 2013년 처음으로 관련기사가 등장한 이후 2015년부터 꾸준히 증가하다가 2018년 급격한 증가 추이를 나타내고 있다.

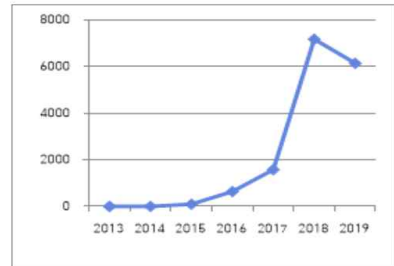


Fig. 2. Change in the Number of newspapers articles by year

키워드 빈도수 상위 35개를 분석한 결과 암호화폐, 네트워크, KT, 비트코인, 가상화폐의 빈도수가 높게 나타난 것을 알 수 있다. 이를 대상으로 워드클라우드 로 시각화한 결과는 Fig. 2와 같다.

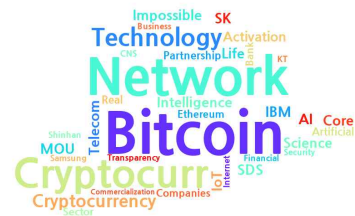


Fig. 3. Word Cloud by total keyword

연도별로 빈도수가 높은 상위 10개의 키워드를 정리한 결과는 Table 2와 같다.

2013년과 2014년은 관련 기사수가 각각 2개기 때문에 상위 10개의 키워드가 분석되지 않았다. 따라서 Table 2에는 2015년부터 분석되었다. 2013년과 2014년에 빈도수가 2이상인 키워드는 2013년에는 암호화, 마운틴폭스, WSJ, 코인테스크였으며 2014년에는 부테린, 가상화폐 이시리움, 생태계 순이었다.

Table 2에서 나타나듯이 2019년 블록체인 활용과 관련된 기사는 금융과 관련된 기사 중심에서 벗어나는 양상을 보이고 있다고 할 수 있다. 즉, 블록체인의

활용이 금융권에 국한되지 않고 비금융권으로 확대되고 있다고 할 수 있다[6]. 2019년의 상위 10개 키워드에 등장하고 있는 클레이튼, 그라운드 X, 디앱은 이전에는 등장하지 않았던 키워드로서 국내에서 개발된 서비스 중심 블록체인 플랫폼과 관련된 키워드라는 점에서 비금융권으로 확대되고 있음을 뒷받침하고 있음을 알 수 있다.

Table 2. Top 10 Keywords by Year

2015	2016	2017	2018	2019
Fintech	Fintech	Virtual currency	Cryptocurrency	Cryptocurrency
Bitcoin	Bitcoin	Bitcoin	Virtual currency	Network
Virtual currency	Capital markets	Fintech	KT	Companies
Network	Financial sector	A.I	Network	KT
Banks	Network	encryption	Companies	Clayton
Coin plug	Big data	Network	Bitcoin	LG
P2P	IoT	Big data	China	Ethereum
Financial sector	Virtual currency	Samsung SDS	A.I	AI
Uri Bank	Financial company	Financial sector	ICO	ground X
Financial transactions	Financial services	New technology	Ethereum	D-App

4.2 토픽 분석

블록체인 기술의 활용과 관련된 신문기사를 LDA 기반 토픽 모델링 한 결과는 Table 3과 같다.

Topic 1은 은행, 인증, 인공지능, 보안, 빅데이터, 모바일, 클라우드와 같이 블록체인의 활용 분야에 대한 단어들이 나타났다.

Table 3. Results of Topic Modelling

Topic	Proportion	Topic words
Topic 1	0.297	service, Fintech, finance, platform, data, Bank, AI, certification, security, Big data, cloud, IoT, insurance
Topic 2	0.160	representation, service, platform, market, business, Enterprise, invest, global, Kakao, exchange, ecosystem

Topic 3	0.178	money, Bitcoin, deal, Virtual exchange, password, invest market, ICO, publish, price currency, Coin, dollar, Bank
Topic 4	0.216	4th, Industrial Revolution, industry, future, innovation, government, support, World, A.I, change, Policy, education
Topic 5	0.149	regulation, government, deal uses, service, industry, finance, Information, apply, Overseas, system, Policy

이는 금융권 뿐 아니라 비금융권을 포괄하는 블록체인의 활용과 관련된 주제임을 유추할 수 있다. 이는 Table 2에서도 볼 수 있듯이 금융권에 국한되어 있던 블록체인의 활용이 비금융권으로 확대되면서 은행과 같은 금융관련 토픽 뿐 아니라 보다 다양한 관련 토픽들이 나타났음을 알 수 있다.

Topic 2는 시장, 투자, 생태계, 스타트업, 회사, 출시 등과 같이 블록체인 관련 시장에 대한 이슈들이 나타났다. 시작, 출시, 진출, 설립, 구축, 참여 등과 같은 단어를 통해 블록체인 시장이 활발하게 이루어지고 있음을 알 수 있었다.

Topic 3은 통화, 코인, ICO, 달러, 은행, 송금 등과 같이 금융, 거래, 경제와 관련된 단어들로 이루어졌다. 이는 Table 2에서도 알 수 있듯이 2018년도까지도 금융권과 관련된 키워드들이 빈도수가 높게 나타났던 것을 비추어보았을 때 토픽으로 다루어질 만큼 많은 신문기사에서 다루고 있음을 알 수 있다.

Topic 4는 미래, 지원, 교육, 변화, 경제, 일자리 등과 같이 블록체인과 관련된 사회적 이슈들로 이루어져 있었다. 특히 제 4차 산업혁명이 가장 비중이 높은 키워드로 나타나는 등 블록체인과 관련된 이슈들이 다양하게 다루어지고 있음을 알 수 있다

Topic 5는 정부, 규제, 거래, 정책 등 블록체인과 정책에 대한 키워드로 이루어져 있다. 우리나라의 블록체인과 관련된 정책이 규제에 집중되었던 것을 반영하듯이 정책과 관련된 토픽에서 가장 비중이 높았던 단어는 '규제'였다. 블록체인 활용과 관련된 신문기사에서 추출된 토픽들의 비중은 Table 3에 제시되었다. 토픽별 비중은 각각의 토픽에 대한 중요도를 나타낸 것이다. 분석결과 활용분야, 관련이슈, 금융거래, 관련시장, 정책 순이다.

블록체인 활용과 관련된 토픽의 워드클라우드 분석

결과는 Fig. 4이다. 각 토픽의 연도별 흐름을 분석하기 위해 토픽을 연도별로 시각화하면 Fig. 5와 같다.



Fig. 4. Example of Word Cloud by Topics

활용분야와 정책은 2015년부터 2019년까지 꾸준히 증가하고 있음을 알 수 있다. 이는 블록체인의 활용이 산업적인 측면뿐만 아니라 국가정책적인 측면까지 확산되었음을 알 수 있다. 금융 거래의 경우 2017년 이후로 점차 감소하고 있음을 알 수 있다. 즉, 금융권과 관련된 블록체인의 활용에 대한 토픽이 점차 낮아지고 있는 것으로 해석된다. 관련 시장과 관련된 토픽 또한 꾸준히 증가되고 있는 것을 알 수 있다.

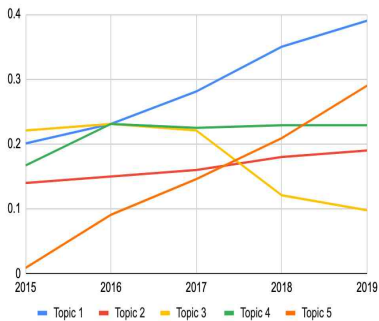


Fig. 5. Topic Trends by years

5. 결론

본 연구에서는 ‘블록체인의 활용’과 관련한 신문기사의 경향을 살펴보고자 2013년부터 2019년까지의 신문기사 15,617건을 수집하여 분석하였다. 논문이나 특허와 관련된 분석은 이루어졌으나 이와 관련된 국내 신문기사 전체를 분석한 연구는 전무하였다. 이에 6여년간의 신문 기사의 경향을 분석함으로써 블록체인의 활용에 대한 변화 추이를 확인하고 이와 관련된 연구의 흐름도 추정할 수 있었다. 본 연구의 결과와 다음과 같다.

첫째, 빈도수 분석을 한 결과 2015년 이후 꾸준히

게 증가한 블록체인 활용과 관련된 기사는 2018년 기하급수적으로 증가한 것을 확인할 수 있었다. 또한, 연도별 상위 10개 키워드를 분석한 결과 2015년 이후 꾸준히 상위 키워드로 자리 잡고 있던 금융과 관련된 키워드들이 2019년 축소된 것을 확인할 수 있었다. 이는 블록체인 기술이 금융권과 관련된 활용에 머무르지 않고 2019년을 기점으로 비금융권으로의 변화가 이루어졌음을 확인할 수 있었다. 또한, KT, IBM 등 기존 기업이 아닌 클레이튼, 그라운드 X와 같은 블록체인 플랫폼 개발 기업이 상위 키워드로 등장하고 있는 것은 블록체인 시장의 변화가 가속화되고 있다는 것을 입증하는 것이라고 할 수 있다.

둘째, LDA 기반의 토픽 분석을 한 결과 활용분야, 관련이슈, 금융거래, 관련시장, 정책 5개의 토픽이 추출되었다. 이 중 활용 분야와 관련된 토픽이 가장 높은 비중을 나타냈다. 이는 빈도수 분석에서와 마찬가지로 금융권에서 시작된 블록체인의 활용이 비금융권까지 확대되는 것과 연관지어 분석할 수 있다. 또한 블록체인의 활용이 4차 산업혁명을 이끌어가는 핵심 기술로 꼽히는 인공지능, 빅데이터, 사물인터넷과 같은 ICT기술과 연관성을 가지고 있음도 확인할 수 있었다. 뿐만 아니라 연도별 토픽 분포를 확인함으로써 블록체인 활용에 대한 신문기사의 트렌트를 분석할 수 있었다.

이와 같이 신문기사에서 다양한 토픽으로 블록체인의 활용을 다루고 있음을 확인할 수 있었다. 은행과 같은 금융권에서 시작한 블록체인이라는 아이디어는 이제 우리의 삶 속에서 흔하게 접할 수 있는 기술로 자리매김할 것이다. 제 4차 산업혁명을 이끌어갈 주요 기술인 블록체인 기술이 어떤 주제를 통해 신문기사에서 다루어지고 있는지 블록체인 기술의 활용과 관련한 과거를 분석한 것은 블록체인 기술의 미래를 예측하는데 중요한 기초자료가 될 것이다.

REFERENCES

[1] WEF, (2015). Deep Shift Technology Tipping Points and Societal Impact, Retrieved from http://www3.weforum.org/docs/WEF_GAC15_Technological_Tipping_Points_report_2015.pdf

[2] Y. Lee, & S. H. Woo. (2018). Research for the convergence of IoT and Blockchain, in

- Proceeding of the of the Korean Institute of Information and Communication Sciences Conference 2018*, Daejeon, 507-509,
- [3] Ministry of Science and ICT, and Korea Institute of Science and Technology Planning and Evaluation, (2018). The future of blockchain,
- [4] S. Nakamoto, (2008). Bitcoin: a peer-to-peer electronic cash system. Retrieved from <http://bitcoin.org/bitcoin.pdf>
- [5] Y. K. Yang, S. B. Cho, & S. H. Chun. (2019). A Study on the Utilization Status and Development Plan for Blockchain Technology: Focusing on Cryptocurrency Policies of Foreign Countries. *The Journal of Business Education*, 33(2), 47-70. DOI : 10.34274/krabe.2019.33.2.003
- [6] S. Kim & S. J. Jeon. (2019). Analysis of Keywords 'Using Blockchain Technology' through Text Mining: Focused on Newspaper Articles. In *Proceeding of Korean Institute of Information and Telecommunication Women's ICT Conference*, 73-76.
- [7] O. S. Yang & J. H. Han. (2019). The Main Topics of Applying Blockchain Technology to the Core Industries in Gangwon Economy : Big Data Machine Learning-based Topic Modeling & Network Analysis. *The Journal of Professional Management*, 22(4), 307-334
- [8] H. I. Jo, J. W. Kim, & B. G. Lee. (2019). A Study on Research Trends of Blockchain Using LDA Topic Modeling : Focusing on United States, China, and South Korea. *Journal of Digital Contents Society*, 20(7), 1453-1460.
- [9] S. M. Lee & S. G. Hong. (2019). Analysis of Blockchain Trends Using Topic Modelling Technique. *Korean Institute of Information and Telecommunication Women's ICT Conference*, 44-47.
- [10] S. Kim, H. Park. & J. Lee. (2018). Word2vec based Latent Semantic Analysis (W2V-LSA): New Topic Modeling Method for Trend Analysis on Blockchain Technology Research, In *Proceedings of the Korean Institute of Industrial Engineers 2018*. 2018(11), 2296-2304
- [11] D. M. Blei. (2012). Probabilistic topic models, *Communication of the ACM*, 55(4), 77-84.
- [12] J. Park, E. Park. & D. Jo. (2015). Automated Text Analysis of North Korean New Year Addresses : 1946-2015. *Korean Political Science Review*, 49(2), 27-61.
- [13] M. R. Jin & H. K. Go. (2019). Analysis of trends in mathematics education research using text mining. *Communications of Mathematical Education*, 33(3), 275 - 294
- [14] H. Kim & H. Rhee. (2016). Trend Analysis of Data Mining Research Using Topic Network Analysis, *Journal of the Korea Society of Computer and Information*, 21(5), 141-148. DOI : 10.9708/jksci.2016.21.5.141
- [15] D. M. Blei, (2012). Probabilistic topic models, *Communications of the ACM*, 55(4), 77-84.
- [16] T. L. Griffiths & M. Steyvers, (2004). Finding scientific topics, *Proceedings of the National academy of Sciences*, 5228-5235.

김성애(Sungae Kim)

[중신회원]



- 2000년 2월 : 성균관대학교 바이오패카트로닉스학과(공학사)
- 2002년 2월 : 한국고원대학교 기술교육과(교육학석사)
- 2015년 8월 : 한국고원대학교 기술교육과(교육학박사)
- 2002년 3월 ~ 현재 : 운암중학교 교사

- 관심분야 : 공학기술교육, 로봇교육, IT융합교육
- E-Mail : ksys21@g.skku.edu

전수진(Soojin Jun)

[중신회원]



- 2000년 2월 : 경인교육대학교 초등교육과(교육학사)
- 2002년 2월 : 경인교육대학교 컴퓨터교육과(교육학석사)
- 2002년 2월 : 고려대학교 컴퓨터교육과(이학석사)
- 2019년 3월 ~ 현재 : 호서대학교 혁신융합학부 교수

- 관심분야 : 정보교육, 컴퓨팅사고력, 교수학습방법 및 평가
- E-Mail : soojin3587@gmail.com