

Policy agenda proposals from text mining analysis of patents and news articles

Sae-Mi Lee¹, Soon-Goo Hong^{2*}

¹Full-time researcher, Smart Governance Research Center, Dong-A University

²Professor, Division of Management Information System, Dong-A University

특허 및 뉴스 기사 텍스트 마이닝을 활용한 정책의제 제안

이새미¹, 홍순구^{2*}

¹동아대학교 스마트 거버넌스 연구센터 전임연구원, ²동아대학교 경영정보학과 교수

Abstract The purpose of this study is to explore the trend of blockchain technology through analysis of patents and news articles using text mining, and to suggest the blockchain policy agenda by grasping social interests. For this purpose, 327 blockchain-related patent abstracts in Korea and 5,941 full-text online news articles were collected and preprocessed. 12 patent topics and 19 news topics were extracted with latent dirichlet allocation topic modeling. Analysis of patents showed that topics related to authentication and transaction accounted were largely predominant. Analysis of news articles showed that social interests are mainly concerned with cryptocurrency. Policy agendas were then derived for blockchain development. This study demonstrates the efficient and objective use of an automated technique for the analysis of large text documents. Additionally, specific policy agendas are proposed in this study which can inform future policy-making processes.

Key Words : Blockchain, Online news, Patent, Text mining, Topic modeling, Policy agenda

요약 본 연구의 목적은 텍스트 마이닝을 활용하여 특허와 뉴스 기사 분석을 통해 블록체인 기술 동향을 탐색하고 사회적 관심을 파악하여 블록체인 정책의제를 제안하는 것이다. 이를 위해 국내 블록체인 특허 요약문 327건과 온라인 뉴스기사 전문 5,941건을 수집하고 전처리 과정을 거쳐 LDA 토픽모델링 방법을 사용하여 특허 토픽 12개와 뉴스 토픽 19개를 추출하였다. 특허 분석을 통해 인증과 거래 관련 토픽이 높은 비중을 차지하였다. 뉴스 기사 분석 결과, 사회적 관심은 암호화폐에 치중되어 있는 것으로 나타났다. 이러한 분석 결과와 의제설정이론에 근거하여 블록체인 관련 정책의제를 도출하였다. 본 연구는 대용량 텍스트 문서 분석의 자동화된 기법을 활용하여 분석을 효율적·객관적으로 수행하였으며, 블록체인 기술 동향과 사회적 관심도를 파악한 실증된 기초 분석 자료를 기반으로 정책의제를 제안하였다. 본 연구에서 제시된 정책의제는 향후 정책 결정과정의 기초자료로 활용될 수 있을 것이다.

주제어 : 블록체인, 온라인 뉴스, 특허, 텍스트 마이닝, 토픽모델링, 정책의제

*This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A3A2075240)

*Corresponding Author : Soon-Goo Hong(shong@dau.ac.kr)

Received December 27, 2019

Revised February 10, 2020

Accepted March 20, 2020

Published March 28, 2020

1. Introduction

Blockchain has been gaining attention as a key technology in the realms of big data, artificial intelligence, the Internet of Things (IoT), drones, and 3D printing in the 4th industrial revolution era. Blockchain technology is a general-purpose technology that can support information exchange and transaction; its application potential continues to expand not only into the economic field, but also into the political, social and scientific fields[1]. Although many investigations regarding the potential applications of blockchain technology have been underway in various fields, it is still in its infancy. Thus, most research surrounding blockchain technology pertains to its development and application to solving time inefficiencies and business process costs[2,3]. In addition, government support and policy development are still essential to manage the competitiveness of blockchain technology.

The purpose of this study is to propose a blockchain policy agenda by analyzing blockchain patent abstracts and online news articles using the LDA that automatically analyzes large text documents.

To establish sufficient and relevant blockchain policy, investigation into social issues and the trends in blockchain technology development are first needed. In particular, the following aspects require careful investigation: 1) the areas currently being studied by blockchain technology and 2) areas emerging as social issues. However, there are not many studies that investigate these aspects. Therefore, here, blockchain policy agendas for news articles and patents related to blockchain technology are derived using a text mining technique.

News articles and patent data were chosen as analysis targets for the following three reasons. First, news articles and patent data involve content that is easy to access. With the rapid spread of digital journalism and social media,

access to the massive amounts of digital content generated every day, across the world, can be easily obtained.

Second, the media contributes to an agenda-setting function, which refers to the ability, e.g. of the media, to highlight the importance of various issues in people's minds through repeated news reports[4]. It is important to analyze the media content between the public and the government since most policies are formed and implemented based on public opinion[5].

Third, a patent provides important information related to technology[6]. A country with research and development plans can identify the technology trends of competing countries by analyzing patent data. In doing so, a potentially valuable research area, that is, vacant technology, can be identified to guide the development of new products or new technologies, thus highlighting the importance of patent analysis[7].

The next chapter presents the concept of blockchain and related previous studies, as well as the theoretical background of the methodology used in the analysis. Section 3 explains the research method, Section 4 presents the findings of the research, and lastly Section 5 describes the conclusions and implications of this research.

2. Literature review

2.1 Blockchain

Blockchain is defined as a P2P network, wherein each network component autonomously communicates and cooperates with the others, and common values are pursued by breaking away from the centralized system[8]. Blockchain has been applied to and used in various fields such as finance, public use, and security together with cryptocurrency because it ensures a reliable contract without the role of a mediator acting as

a connection point[9].

The blockchain revolution is divided into four categories: blockchain 1.0, 2.0, 3.0, and 4.0[10]. Blockchain 1.0 refers to applications related to cryptocurrency used in digital payment systems such as Bitcoin. Blockchain 2.0 is a smart contract, and includes the entire economic market using blockchain to extend existing transactions such as stocks, bonds, and spot contracts. Blockchain 3.0 includes applications used in industrial services such as vote counting, digital health records, and digital arts[11]. Blockchain 4.0 focuses on creating various industrial ecosystems by fusing these blockchain technologies.

2.2 Topic modeling

Topic modeling refers to a text analysis method that extracts meaningful information from documents written in languages, which are unstructured data. Topic modeling, one of several text mining techniques, has recently attracted attention for its ability to perform in-depth analysis of text data by extracting meaning from words contained in a document. Topic modeling is an unsupervised machine learning technique that automatically extracts meaningful topics from a set of text documents[12]. In other words, it learns the structure of each topic from a huge amount of document collections without human supervision[13]. Furthermore, topic modeling reveals topics within the text corpus by recognizing a document as a mixture of topics and capturing the patterns of word co-occurrences in the document[14].

LDA estimates the ratio of topics that are latent in a document through the distribution of words making up the document[15]. Thereby, a topic with the largest ratio becomes the representative topic of the corresponding document. Based on the assumption that the

topic distribution follows LDA, the occurrence possibility of a word by topic is estimated indirectly through the number of presumed topics and the distribution of words in the entire document. LDA has a feature that separates or combines the same words if they have different meanings in the context. That is, it can process synonyms (words with different spellings but the same meaning, e.g., picture and photo that refer to camera characteristics, and blue and green that refer to color types) and polysemic words (words with different meanings in different situations, words with more than one meaning, e.g., mouse) by grouping words of the same topic into a single group. Therefore, it is possible to take a semantic approach in topic classification of documents[16,17]. This study adopts LDA topic modeling among various text mining techniques for analysis.

2.3 Agenda-setting using text mining

This study approaches policy agenda-setting based on the agenda-setting theory. The agenda-setting theory is a typical theory in the field of mass communication that has been further developed since McCombs & Shaw[18] presented it. They argued that as people, events, and issues were exposed more frequently and covered more thoroughly by the media, public interests in such highly exposed topics naturally increased regardless of the topics' real importance. As an extension to the agenda-setting theory, the shaping of media and public interests are considered to likely lead to executions of political actions.

Since it is practically impossible for an audience to directly collect and evaluate the vast amount of information contributing to the media's effects on the public, it is assumed that people have no choice but to rely on the information provided by the media. The study by Iyenga et al.[4] found that issues stressed by the

media were perceived as important by the audience, whereas issues not taken seriously by the media were not perceived as important by the audience, indicating that the media agenda affected the public agenda. Given that most policies are formed and implemented based on the agenda aroused in public opinion, it is critical to identify the coverage given by the media, as it may essentially mediate government and civil society[5].

Text mining is a technique for analyzing unstructured data such as images, voice, video, and text, and can be used to derive policy tasks. The unstructured data accounts for over 80 percent and it is becoming more and more important to analyze it[19]. Studies that derive agendas by analyzing text data and suggesting policy implications across various fields are underway. Advantageously, text mining enables policy makers to more objectively derive the interests and sentiments of policy stakeholders including the public, the media, and interest groups, while minimizing the subjective intentions of the researcher.

Through Twitter analysis using text mining, Bae, Son & Song[20] claimed that social media content has been a major research subject in information search and text mining, in that social media content not only provides the vast amount of text data generated in real time, but also is used as an influential channel for agenda-setting and public opinion formation. Sung & You[21] analyzed the public opinion on civil affairs and policies through text mining,

The research using the patent and news data is as follows. Kam et al.[22] Collects Korean and US patents in the biotechnology field using technology classification based on IPC code and analyzes them using text mining to analyze the trends of technology development through patents. Choi[23] analyzed the trend of big data patents using patent keywords. Jeong et al.[5] performs text mining analysis for 2,569 articles

on the disabled reported in 10 top daily newspapers for 15 years, and conducts research on what agendas were set and reported regarding the labor of the disabled by major media outlets in each period.

Unlike previous researches aforementioned, this study is to propose has originality in that it derives policy agenda for blockchain development from various perspectives by analyzing both patent data reflecting technology trends and news articles representing social interests. By analyzing patents and news articles together, we propose a more realistic policy agenda by simultaneously examining the social interests and the public trends on blockchain policies.

3. Methodology

3.1 Data collection and preprocessing

Fig. 1 shows the overall analysis process of this research. Since we targeted online news articles and patent data for analysis, full-text articles relating to blockchain were first collected from the two most read daily newspapers specializing in IT (etnews and Digital Times) among online news provided by the portal site Naver using the Python BeautifulSoup library. The collection period ranged from January 1, 2016-June 30, 2019 during which a total of 5,941 news articles were collected. Additionally, as of June 30, 2019, a total of 327 published patents were collected that included the word 'blockchain' in the patent abstracts and titles, and which were to be registered among domestic patent data provided by the Korea Intellectual Property Rights Information Service(KIPRIS), run by the Korea Institute of Patent Information.

Next, preprocessing on the collected data was performed. Nouns from patent abstracts and news articles were extracted and compared using various Korean morphological analyzers such as OKT of Korean Natural Language Processing in

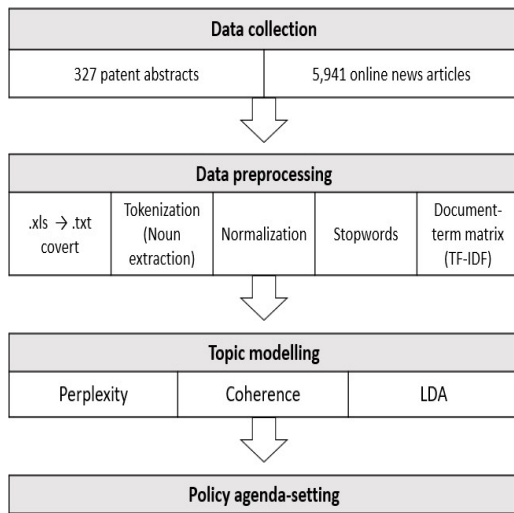


Fig. 1. Analysis process

Python(KoNLPy), Hannanum, Komoran, and Kkma, which transformed words into an easy-to-analyze form through the use of natural language processing techniques. Of the analyzers used, OKT produced the best results.

For data refinement, news articles and patent data collected in Excel files were first transformed into text files in which only nouns were extracted through morphological analysis. Second, nouns unnecessary for analysis were removed and synonyms among the extracted nouns normalized. The phrases "reporter," "reporter's email address," and "reporter's name," were removed from the news data, and "invention" was removed from the patent data. Third, the specific word 'blockchain' was removed; the word "blockchain" does not provide discriminatory information to extract topics as it is commonly included in all documents.

A matrix consisting of words and frequencies that appeared in a document was generated by transforming the refined text data into a document-term matrix. When generating the matrix, the word frequency was calculated using the Term Frequency-Inverse Document Frequency(TF-IDF) weighting value. The TF-IDF weighting value is a measurement value that

indicates how important a word is in a particular document when the target text to be analyzed consists of multiple documents. Thus, the TF-IDF is calculated based on the probability of occurrence of a word rather than on the number of its occurrences. A high absolute frequency does not necessarily mean that the word is important; i.e., in texts such as news articles and patent abstracts, the frequency of a word that does not represent the document can still be high.

3.2 LDA topic modeling

LDA topic modeling models which keywords should be assigned to a topic based on the words in a document as well as the probability of how many keywords of each document should be included in every topics[12]. To classify online news and patent data by topic, topics were extracted using LDA of the Python Genism package for text files that had been preprocessed.

To perform topic modeling, the number of topics should be determined. Generally, the researcher sets the number of topics by deciding the number of topics that are best categorized; alternatively, the number of topics can be set by an algorithm to determine the number of topics using the complexity value of probability distribution. Here, coherence score was used to more objectively determine the appropriate number of topics.

Coherence is a measurement of the consistency of a topic. If topic modeling is done well, semantically similar words are close to each other in a topic. That is, the similarity of words in a topic can be calculated to identify whether the topic consists of semantically identical words. The higher the coherence score, the higher the semantic consistency.

In this study, iteration value is decided based on previous studies which recommend an iteration value between 1,000 and 2,000[24]. Thus the iteration value was set to 1,000 and the default value was used for α in this research.

4. Results

4.1 Patent topic analysis

By analyzing the 327 patent abstracts related to blockchain, technology trends were identified and agendas derived. As described above, topics were extracted through data preprocessing and LDA(details are shown in Table 2). To calculate the coherence score of the patent data, the number of topics was set to 5 to 25. The calculation results of coherence score showed that the coherence score was the highest at 12 topics(as shown Fig 2). High coherence means that each topic is determined by a single obvious factor; thus, it is desirable to set the number of topics as the final number of topics with the highest coherence value[25]. Based on previous studies, the number of patent data topics was set to 12.

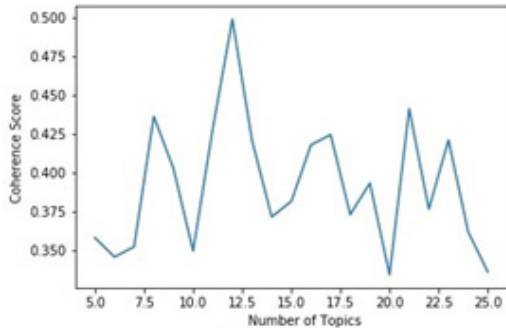


Fig. 2. Coherence score graphs for patents

The meaning of the topic extracted by LDA can be identified by using a set of keywords. Keywords are determined based on the co-occurrence probability of words in the document according to the LDA topic modeling algorithm, and the keyword with the highest probability is placed at the top of each topic. In Table 1 and 2, the keywords that are listed first for each topic represent the topic most. Therefore, each topic label was specified by reviewing the keywords that belonged to the 12 topics(Table 1). It should be noted that when

specifying the topic label, the same keyword can belong to several topics. In such cases, the relative importance of the keyword can vary by each topic; therefore, the meaning of a topic must be determined by combining it with other keywords in the topic. Here, three experts decided each topic label by examining the keywords in the topic. Among the patent topics, 5 topics were extracted that were related to payment and transactions(Topic-1, 2, 8, 10, 11). Seven topics were extracted that were related to security and authentication(Topic-3, 4, 5, 6, 7, 9, 12). These topics are further outlined in Table 2.

Table 1. Patent topics

Topic: Keywords	%
Certificate(Topic-4): key, public, remittance, certificate, user, terminal	36.54
Location authentication(Topic-6): area, mobile, marking, location, database, command	16.83
Online transaction(Topic-11): transaction, virtual, value, currency, server, authentication	10.92
Electronic wallet(Topic-2): wallet, issue, progress, currency exchange, game, buy	5.34
Internet certificate(Topic-7): Internet, rating, decision, district, posting, object	4.45
Copyright management(Topic-1): self, software, copyright, copyrighted work, transfer, segment	4.27
File forgery and tampering(Topic-9): file, check, tampering, integrity, metadata, contract	4.14
Self-learning module(Topic-8): hash, value, self, outgoing, learning, sensor	4.09
Digital content encryption(Topic-3): digital, document, currency, version, password, capacity	3.83
Electronic document security(Topic-12): electronic document, digital signature, user, individual, digital, platform	3.48
Payment method(Topic-10): object, uniqueness, type, rule, sound wave, payment	3.35
Open source software(SW) license(Topic-5): deposits and withdrawal, license, transaction, software, compliance, open source	2.76

To identify technical trends, the analysis results of the proportion ratio of each topic for patent abstracts are as follows. The proportion ratio represents the weight of the topic in the

entire document set. Topic-4 accounts for 36.64%, which is the highest, followed by location authentication Topic-6 with 16.83%, and Topic-11 with 10.92%. Accordingly, technologies related to authentication, security, payment, and forgery/tampering constitutes the majority of patents, wherein authentication technology constitutes the largest proportion.

4.2 News topic analysis

To identify social issues and thus derive agendas, 5,941 online news articles were analyzed. Coherence score was calculated by setting the number of topics to a maximum of 35 with the same method that was used for deriving agendas when analyzing the patent data(as shown Fig. 3). Resultingly, LDA was performed by setting the optimal number of topics to 19. The 19 news topics extracted from the LDA analysis are shown in Table 3.

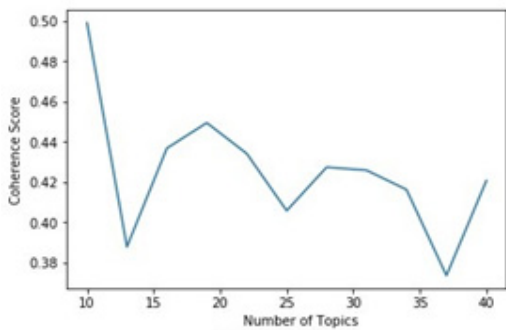


Fig. 3. Coherence score graphs for news articles

Among news topics, two topics were extracted related to cryptocurrency(Topic-5,6); three topics were extracted related to security: security(Topic-3,4,11); four topics were extracted related to finance(Topic-8,14,15,19). Other extracted topics included the 4th industrial revolution (Topic-10), business platform (Topic-16), applied technology (Topic-1), forum (Topic-9), job creation (Topic-7), reorganization (Topic-12), Samsung SDS (Topic-13), blockchain specialized company (Topic-18), and patent (Topic17). These

are further outlined in Table 3.

The analysis results of the proportion ratio of each topic are as follows. Cryptocurrency

Table 2. News article topics

Topic: Keywords	%
Cryptocurrency transaction(Topic-6): service, password, exchange, platform, payment, currency	55.43
Cryptocurrency policy(Topic-5): innovation, overseas, cryptocurrency, ICT, selection, government	18.31
Security(Topic-4): product, protection, sales, commodity, partnership, patent,	7.82
4th industrial revolution(Topic-10): 4th industrial revolution, medicine, confirmation, city, effort	5.46
Business platform(Topic-16): price, real estate, estimation, creation, IP (Intellectual Property), capability	4.22
Applied technology(Topic-1): AI, wallet, collection, VR, field, account	2.80
Forum(Topic-9): discuss, international, promising, planning, trend, debate	2.48
Job creation(Topic-7): job, small and medium sized business, 10000, 5 year, hyper-connection, core technology	0.47
Reorganization(Topic-12): Nonghyup Bank, management, small and medium size, Securities Depository, reorganization, subsidiary	0.46
Financial certification(Topic-2): Bank of Korea, iris, body, Bank of Korea, international, Fast Identity Online (FIDO)	0.35
Over-The-Counter(OTC) stock trading(Topic-14): Nasdaq, Initial Public Offering(IPD), Proof of Concept(PoC), initial public offering(IPD), OTC market, Korea Over-The-Counter (K-OTC)	0.30
Cyber security(Topic-11): attack, pilot project, ransomware, target, training, Institute for Information and Communications Technology Promotion (IITP)	0.28
Joint blockchain network for the securities industry(Topic-8): issue, maintenance, member company, Woori Bank, Mirae Asset Daewoo, Yuanta Securities	0.27
Samsung SDS(Topic-13): Samsung SDS, Yu-sung Chung, forum, director, bulk, medicine	0.24
Financial transaction platform(Topic-15): financial investment industry, CID, Tae-lyong Kim, warranty, startup, Korea Financial Investment Association	0.23
Company specialized in blockchain(Topic-18): The loop, specialized company, concert, Kyung-joon Lee, the newest, Yello Financial Group	0.23
Financial Security Institute(Topic-3): Financial Security Institute, Chang-eon Heo, bio authentication, director, pilot, financial company	0.23

(Topic-6) accounts for more than half of the total of 5,923 news articles, at 55.43%, followed by cryptocurrency policy (Topic-5) at 18.31%; Thus, cryptocurrency related topics are overwhelmingly dominant.

4.3 Policy agenda-setting

In Sections 4.1-4.2, the results of topic extractions were presented and the distribution of topics and keywords within the topics examined by analyzing patent abstracts and online news articles related to blockchain technology. In the analysis of patents and news data, it is notable that topics related to authentication and security constitute the majority of topics for patents, yet topics related to cryptocurrency are overwhelming present among news articles. Based on these results, the following blockchain-related policy agendas are proposed.

First, there is a need for governmental regulatory innovation pertaining to blockchain involving cryptocurrency-based blockchain technology. Given that the topics related to cryptocurrency (Topic-5 and 6) account for approximately 73% of the total, there is a huge social interest and demand related to cryptocurrency. Regarding topics related to patents, cryptocurrency-based blockchain technology has not advanced enough to satisfy public interests and or to be given the time needed for further development, to a degree that cryptocurrency is hard to find. Initially, blockchain was represented as a cryptocurrency and regarded with speculation in Korea. This contributed to the negative perceptions towards blockchain and regulations in Korea, resulting in fewer patents in Korea than in the US or China; a single Korean company was not included in the 10 leading patent application companies[25]. To overcome this relative shortcoming, the government designated Pusan as a blockchain regulatory free zone with the aim of regional innovation and growth by drastically reducing

regulations that prevented the development of new growth industries and new technologies. Despite the government's willingness to implement innovation policies by lifting off the regulations on blockchain, there are some arguments that such innovation policies are not effective, since cryptocurrency utilization services are not allowed at all. Thus, technology regulation on cryptocurrency-based blockchain still remains a problem to be solved. In response to regulatory issues, domestic blockchain companies are running blockchain services by establishing subsidiaries in overseas countries such as Singapore, Japan, and Switzerland. Thus, service regulations on cryptocurrency-based blockchain should be reviewed.

Second, to vitalize the blockchain business, Korea's blockchain service policy should focus on strengthening consumer security needs. As can be seen from the analysis results of the topics, blockchain technology is mainly used in finance related fields such as payment, online transactions, electronic wallets, encryption, financial certification, the Financial Security Institute, financial transaction platforms, OTC stock trading, and joint blockchain networks for the securities industry. For such financial transactions, personal information leakage results in serious social issues by increasing the possibility of breaching the information held by individuals. Therefore, the need for security technology based on blockchain is increasingly important to defend against such threats to personal information and information identity. At the government level, clear rules should be established and an institutional framework provided surrounding information protection and security for online financial transactions based on blockchain.

Third, to build a blockchain-based ecosystem, government-led development strategy is needed for value creation using blockchain, through the pilot projects of the central and local

governments. As can be seen from the 4th industrial revolution (Topic-10), which is ranked 3rd among the news article topics, blockchain is one of the technologies representing the 4th industrial revolution; additionally, blockchain is included in the current government's promotion of 4th industrial revolution strategic technology. In the 4th industrial revolution era, the convergence of multidisciplinary technologies no longer an optional choice, but a necessity for survival. However, as indicated in the extraction results of the blockchain-related patent topics, blockchain technology is mostly made up of technologies related to authentication and security. The news article analysis results also show that news articles are mostly related to financial topics. That is, there are almost no topics on the application of blockchain technology to various fields.

Fourth, there is a need to expand blockchain education and to foster talented individuals. In examining the news article topics presented in Table 3, cybersecurity (Topic-11) includes the keywords 'training' and 'talent,' and job creation (Topic-7) ranks 8th among the topics. As the adoption of blockchain moves rapidly amongst startups, there is a high demand for talented individuals who have received reliable blockchain training from reputable institutions; the demand for blockchain experts increases by 517% on a year-to-year basis[26]. Samsung announced in August 2018 that it would support 500 startups and train 10,000 software experts over the next five years to foster an open innovation ecosystem and create jobs for youth[27]. The experts training program for blockchain technology was offered as one of the '2019 Innovation Growth and Intensive Training Project for Youth Talents' hosted by the Ministry of Science and ICT and held by the Institute for Information and Communications Technology Promotion. The Banking industry is also actively recruiting digital professionals including

blockchain specialists, through on-demand employment, and is internally training professionals to gain a competitive edge in the digital banking market. In this way, not only major companies and central government departments, but also schools, local governments, and public institutions should provide more educational opportunities for the majority and make an effort to train and discover high-quality talents by expanding and conducting digital professional training that facilitates blockchain proficiency.

5. Conclusion

5.1 Implications

This study identified the sources of interest in blockchain from the perspective of society members including companies, the public, and the market, according to news from the media. Additionally, it analyzed patent data to explore the major technology trends in blockchain. LDA topic modeling was performed to analyze patents and news documents as unstructured text data. Based on the extracted topics, media agendas and technology trends were identified, and policy implications in the fields of blockchain science and technology were proposed. Ultimately, theoretical implication, practical implications, and limitations of this work are derived as follows.

Theoretical implications are two-fold: First, the analysis performed demonstrates an efficient and objective analysis of large text documents via an automated technique. Through complementary use of coherence score to calculate the optimal number of topics, and the LDA topic modeling algorithm to extract topics, the subjectivity of the researcher is excluded and thus more objective research results provided. In the previous research, the researcher conducted several tests to determine the number of topics, involved the subject of the researcher, and it was

difficult to provide a basis for the appropriate number of topics. This method can be applied to the analysis of unstructured text data in a variety of fields such as medical records, speeches, meeting minutes, and product and service reviews as well as online news articles and patent abstracts used in this analysis. Therefore, it can be used in various fields including medicine, political science, business administration, hotel & tourism industry, and so on.

Second, previous studies were examined with the agenda-setting theory as the theoretical framework; based on this, an empirical analysis was conducted, preparing the theoretical foundation for deriving the science and technology policy agendas for continued blockchain development. This paper attempted to derive the agenda by analyzing news articles based on the agenda-setting theory, which implies that the media agenda moves to the public agenda, which naturally leads to political action. In this study, we derived the agenda based on empirical analysis using text mining techniques and proposed a policy agenda based on this.

Practical implications of this research are also two-fold: First, the agendas needed for blockchain policy were derived by including both technical and social perspectives. There are hardly any previous studies surrounding the convergence of technological trends and social interests. The analysis of patent abstracts approached blockchain from a technical perspective, while the analysis of online news articles incorporated the perspective of society members, markets, and policy. The technique used in this study can be utilized to derive policy agendas in other fields.

Second, the policy agendas presented in this study can be used as basic data in the policy making process. By verifying technologies that have gained a competitive edge in the field of blockchain technology, a plan to continuously promote them can be actuated. Furthermore, it is

expected that the establishment of policy support plans for the research and development of related fields can be further developed by the identification of areas in which blockchain technology is not thoroughly utilized.

5.2 Limitations and future research

Despite the above implications, there are several limitations of this research. First, the data is limited both temporally and quantitatively. For patent data, which represents the technical perspective in this research, there is a considerable amount of time lag between patent applications and publication. Thus, the number of patents presented herein from year to year may not necessarily reflect the amount of technological innovation; there is the possibility that patents are under examination, or that technologies have been applied but not yet registered; it is important to note that the process from examination to registration takes one year. Additionally, because blockchain is a state-of-the-art technology, the number of related patent abstracts is significantly less than that of other documents, resulting in insufficient analysis from technical aspects. To overcome these limitations, should the number of patents increase in the future, the technical aspects should be investigated in more detail through reanalysis.

Second, it is possible to propose technology policy that is limited to Korea by analyzing only domestic patents; yet this is limited in that it ignores identification and comparative analysis of technology trends in the global market. To gain a global competitive advantage by comparing the technology trends of other countries wherein the development of blockchain technology has been actively underway, which technologies should be promoted and which sector areas have been overlooked should be investigated, and the results used to inform policy decisions.

REFERENCES

- DOI : 10.1109/TKDE.2014.2313872
- [1] M. Swan. (2015). *Blockchain: Blueprint for a new economy*. Sebastopol, CA, USA: O'Reilly Media, Inc.
 - [2] A. Dorri, S. S. Kanhere & R. Jurdak. (2017). Towards an optimized blockchain for IoT. *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*, Pittsburg, PA, USA, ACM, 173-178.
 - [3] K. Biswas & V. Muthukkumarasam. (2016). Securing smart cities using blockchain technology. *Proceedings of the 2016 IEEE 18th International Conference on High Performance Computing and Communications (HPCC)*, Sydney Australia, 1392-1393.
 - [4] S. Iyengar, M. D. Peters & D. R. Kinder. (1982). Experimental demonstrations of the "not-so-minimal" consequences of television news programs. *The American Political Science Review*, 76, 848-858.
 - [5] J. J. Jeong, J. M. Lee & S. Y. Choi. (2018). Analysis of news regarding the disabled labor using text mining techniques. *Reinterpretation of Disability*, 48-100.
 - [6] C. Jeong & K. Kim. (2014). Creating patents on the new technology using analogy-based patent mining. *Expert systems with applications*, 41(8), 3605-3614.
 - [7] G. J. Kim, S. S. Park & D. S. Jang. (2015). Technology forecasting using topic-based patent analysis. *Journal of Scientific and Industrial Research*, 74(5), 265-270.
 - [8] S. B. Lim, H. R. Choi & N. R. Kim. (2018). *Smart governance system based on blockchain..* Pusan, South Korea: Sejong Publisher. DOI : 9791159792267.
 - [9] C. Park & K. Park. (2018). Trend analysis of application fields of blockchain technology using patent data. *Journal of Korean Institute of Next Generation Computing*, 14(2), 72-81.
 - [10] A. Rahaman. (Aug 20, 2018) *InterValue The World's first practical blockchain 4.0*. Medium. <https://medium.com/@aminurrahaman/intervalue-the-worlds-first-practical-blockchain-4-0-b9324878c262>
 - [11] S. Miao & J. M. Yang. (2018). Bibliometrics-based evaluation of the blockchain research trend: 2008-March 2017. *Technology Analysis & Strategic Management*, 30(9), 1029-1045.
 - [12] D. M. Blei, A. Y. Ng & M. Jordan. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
 - [13] S. Arora et al. (2013). A practical algorithm for topic modeling with provable guarantees. *Proceedings of the 30th International Conference on International Conference on Machine Learning*, Atlanta, GA, USA, v. 28.
 - [14] X. Cheng, X. Yan, Y. Lan & J. Guo. (2014). BTM: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering*, 26(12), 2928-2941.
 - [15] D. M. Blei. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
 - [16] Z. Zhai, B. Liu, H. Xu & P. Jia. (2011). Constrained LDA for grouping product features in opinion mining. *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Ho Chi Minh City, Vietnam; Springer, Berlin, Heidelberg, 448-459.
 - [17] J. Born, E. Scheihing, J. Guerra & L. Cárcamo. (2014). Analysing microblogs of middle and high school students. *Proceedings of the European Conference on Technology Enhanced Learning*, Lyon, France, Springer: Cham, 15-28.
 - [18] M. E. McCombs & D. L. Shaw. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176-187.
 - [19] E. M. Park & J. H. Seo. (2019). A Study on Leadership Typology in Sports Leaders Based on Big Data Analysis. *Journal of the Korea Convergence Society*, 10(7), 191-198.
 - [20] J. Bae, J. Son & M. Song. (2013). Analysis of twitter for 2012 South Korea presidential election by text mining techniques. *Journal of Intelligence and Information Systems*, 19(3), 141-156.
 - [21] B. K. Sung & Y. Y. You. (2018). Analysis of the complaints and policy of the Ministry of Employment and Labor using the R program. *Journal of the Korea Convergence Society*, 9(7), 41-46.
 - [22] J. S. Kam, M. Y. Kim & B. H. Hyun. (2013). A study on analysis of patent information based biotechnology research trend and promising research themes. *Journal of Technology Innovation*, 21(2), 25-56.
 - [23] J. C. Choi. (2018). Big Data Patent Analysis Using Social Network Analysis. *Journal of the Korea Convergence Society*, 9(2), 251-257.
 - [24] J. An, K. Ahn & M. Song. (2016). Text mining driven content analysis of Ebola on news media and scientific publications: using bio research papers and news text data. *Journal of the Korean Society for Library and Information Science*, 50(2), 289-307.
 - [25] T. Jeong. (June 17, 2019). *There are no Korean companies in the global top 10 blockchain patents*. The Herald Business. <http://news.heraldcorp.com/view.php?ud=20190617000149>
 - [26] S. Kim. (June 6, 2019). *Blockchain Developers and Professional Construction Talents Shortage... Government and industry "employ and educate*. Financial News. <http://www.fnnews.com/news/201906061710510492>
 - [27] J. Park. (Aug 9, 2018). *'Innovation Ecosystem's Way to Live' Large companies have started to support startups*. Digital Times. http://www.dt.co.kr/contents.html?article_no=2018080902109932781011&ref=naver

이 새 미(Sae-Mi Lee)

[정회원]



- 2017년 2월 : 세종대학교 경영학과 경영학박사
- 2010년 1월 : Northeastern University, Master of Science in Leadership
- 2018년 11월 ~ 현재 : 동아대학교 스마트거버넌스연구센터 전임연구원
- 관심분야 : 빅데이터, 경영전략, 서비스

경영

· E-Mail : emailme6@naver.com

홍 순 구(Soon-Goo Hong)

[정회원]



- 2000년 8월 : University of Nebraska at Lincoln 경영학박사
- 1995년 5월 : University of Nebraska at Lincoln 경영학 석사
- 2001년 9월 ~ 현재 : 동아대학교 경영정보학과 교수
- 관심분야 : MIIS, 빅데이터, 정보시스

템평가

· E-Mail : shong@dau.ac.kr