

키워드 네트워크 분석 방법을 활용한 블록체인 트렌드 분석에 관한 연구

조성환*

A Study on Analysis of the Trend of Blockchain by Key Words Network Analysis

Seong-Hwan Cho*

요약 본 연구는 키워드 네트워크 분석에 사용되는 텍스트마이닝과 의미연결망 분석 방법을 활용하여 블록체인의 산업 활용 분야로 언론 및 정부 발표에서 언급되고 있는 '금융', '에너지', '물류'를 언급한 기사들을 비교 분석하였다. 블록체인 적용이 언급된 산업 분야별로 기사의 내용 및 키워드의 차이를 파악하고 비교 분석하는 것을 목적으로 하였다. 2017년 1월부터 2018년 7월까지 언론에서 보도한 총 43,093건의 기사를 Python BeautifulSoup을 이용하여 네이버 뉴스에서 수집하였고, 세 용어의 상호 중복을 제거하기 위한 정제 작업을 수행하였다. 이후 키워드 간 네트워크 분석을 위해 텍스톰(Textom)과 UCINET을 이용하여 세 용어에 대한 텍스트마이닝과 의미연결망 분석을 진행하였다. 분석 결과, 세 용어는 모두 '기술' 측면에서는 유사한 단어들이 있었으나, '정부 정책'이나 '산업'측면의 이슈 등에서 내용적 차이가 있었다. 또한 빈도 및 중심성에 있어서도 차이가 있음을 확인할 수 있었다.

Abstract This study aims to identify and compare contents and keywords used in articles related to blockchain applications to various industries. The text mining and Semantic Network Analysis, as methods of keyword network analysis, were used to analyze articles including terms of 'finance', 'energy' and 'logistics', which media and government frequently mentioned as areas that can apply blockchain technologies. For this study, data were collected from 43,093 articles from January, 2017 through July, 2018. Data crawling was carried out by using Python BeautifulSoup and data cleaning was performed in order to eliminate mutual redundancies of the three terms. After that, text mining and semantic network analysis were performed using Textom and UCINET for network analysis between keywords. The results showed that all the three terms were similar in terms of 'technology', but there were differences in the contents of 'government policy' or 'industry' issues. In addition, there were differences in frequencies and centralities of these terms.

Key Words : Blockchain, Finance, Logistics, Energy, Semantic Network Analysis

1. 서론

최근 블록체인 기술은 가상화폐와 같은 금융 분야에 국한된 기술에서 벗어나 다양한 산업에서 공공 및 민간 시범사업이 진행되고 있으며 2018년 6월에는 과

학기술정보통신부가 블록체인 글로벌 기술경쟁력 확보를 위한 「블록체인 기술 발전전략」을 발표하며 구체적인 추진계획을 세우고 있다.[1]

발표전략에는 현재는 물론 향후에도 블록체인 기술의 적용 가능성과 효과성이 높을 것으로 예상되는 금

This work was supported by the research grant of Geumgang University in 2017.

*Corresponding Author : Department of IT Software, Division of Global Convergence, Geumgang University (shcho@ggu.ac.kr)

Received August 31, 2018

Revised September 11, 2018

Accepted October 23, 2018

용, 에너지, 물류, 의료, 공공, 콘텐츠와 같은 6가지 산업 분야에 언급하며 그 중요성에 대해 기술하였고, 이러한 상황에서 분야 별 블록체인 연구동향에 대한 이해가 필요한 시점으로 판단되고 있다.

블록체인과 유사한 인공지능 등 다양한 과학기술동향에 대한 분석 방법으로는 키워드 네트워크 등의 데이터 분석 연구들이 해외는 물론 국내에서도 최근 활발히 연구되고 있다.[2][3] 그 중 블록체인과 관련된 연구들도 일부 수행되었지만 가상화폐 관련 분석 연구가 다수를 이루고 있는 상황이다.[4][5]

이에 본 논문은 블록체인 관련 연구 중에서 기존 연구들이 이루어진 가상화폐 관련이 아닌, 정부에서 발표한 6가지 산업에 대해 초점을 두고자 하였다. 그 중 현재 활용도가 가장 높은 ‘금융(finance)’, ‘물류(logistics)’, ‘에너지(energy)’ 분야라는 세 키워드에 대해 텍스트마이닝, 의미연결망분석을 이용하여 분류하고 키워드 별 어떠한 특성 차이가 있는지 비교분석을 시행하고 그 시사점을 찾아보고자 한다.

2. 연구 방법

2.1 분석대상과 방법

본 연구는 키워드 네트워크 분석에 사용되는 텍스트마이닝과 의미연결망 분석 기법을 활용하였다. 대상으로는 블록체인 기술의 활용 분야로 언급되는 ‘금융’, ‘물류’, ‘에너지’에 대한 최근 7개월간의 뉴스 기사들을 수집하였고, 이후 분석을 통해 키워드별로 분류하고 특성을 파악하였다.

2.1.1 네이버 뉴스 수집

뉴스 수집 도구는 Python BeautifulSoup을 이용하였고, 네이버에서 제공하는 API를 이용하여 기사를 수집하였다. BeautifulSoup은 HTML 코드를 Python이 이해하는 객체 구조로 변환하는 Parsing을 하고, 이 라이브러리를 이용해 의미 있는 정보를 추출해 낼 수 있다.[6] 수집되는 내용 정제를 위해서 2가지 작업을 하였고, 다음과 같다.[7]

첫째, 내용은 관련이 없지만 언론사 명에 ‘금융’, ‘물류’, ‘에너지’가 들어있기 때문에 본문이 수집되는 기

사들은 연관성을 검토하여 불필요한 기사는 제외하였고, 해당 언론사명은 <표 1>과 같다.

표 1. 수집 키워드와 관련하여 제외 검토한 언론사명
Table 1. Name of the news media to be excluded in relation to the collected keywords

키워드	금융	물류	에너지
언론사명	대한금융신문 한국금융신문 금융소비자뉴스 조세금융신문	물류신문	투데이에너지 에너지경제

둘째, ‘금융’, ‘물류’, ‘에너지’ 3가지 키워드 중 2가지 이상이 중복된 기사를 제거하였다. 이를 위해 네이버 포털의 3가지 키워드 상세검색(정확히 일치하는 단어/문장, 반드시 포함하는 단어, 제외하는 단어) 방법을 적용하여 데이터를 수집하였고, 중복을 제거한 결과는 아래 <표 2>과 같다.

표 2. 키워드별 뉴스 수집 수와 중복제거 수
Table 2. Number of news articles and elimination of redundancies by keywords

분류	2017~2018	중복제거	비고
블록체인 + 금융	39,664	34,952	금융- (물류+에너지)
블록체인 + 물류	5,281	2,482	물류- (금융+에너지)
블록체인 + 에너지	8,517	5,659	에너지- (물류+금융)
합	53,462	43,093	-

본 연구에서 수집한 블록체인 관련 ‘금융’ 기사 수는 39,664건이지만, 위 [표1]과 같이 물류, 에너지의 중복을 제거한 수는 34,952건이었다. ‘물류’와 ‘에너지’도 같은 방법으로 진행하였고 각각의 중복 제거한 결과는 2,482건, 5,659건이었다. 이러한 결과에 따라 본 연구에서는 위 <표 2>와 같이 총 43,093건의 데이터를 텍스트마이닝과 의미연결망 분석에 사용하였다.

2.1.2 텍스트마이닝 분석

텍스트마이닝(Text Mining)이란 자연어처리에 기반한 기술로, 비정형 텍스트 데이터에서 패턴 또는 관계를 추출하여 가치와 의미 있는 정보를 찾아내는 기법을 의미한다.[8] 텍스트마이닝 분석은 구조화되어 있지 않은 대용량의 텍스트 집합을 분석하기 위해 사용되며, 비정형 텍스트 자료를 자연어처리와 형태소 분석기술에 의거해 정제하고 유용한 키워드를 추출해 맥락 수준의 의미를 찾아내는 과정을 통해 수행된다.[9]

2.1.3 의미연결망 분석

‘의미연결망분석(Semantic Network)’은 일종의 ‘온라인 담론분석(Discourse Analysis)’으로 문장의 개념들과 상징들에 대한 구조를 밝히기 위해 키워드 네트워크를 구축해 분석을 진행하며, 텍스트에서 자주 활용하는 단어를 담론의 성격을 드러내는 하나의 지표로 간주하여 사용하는 분석 방법이다. [11]

의미연결망 분석은 행위자 대신 메시지의 구성요소인 단어들의 구조적 관계를 탐색하고 의미를 분석한다.[10]

2.1.4 텍스트(TEXTOM)과 UCINET

본 연구에서는 먼저 텍스트마이닝 분석을 위해 텍스트를 사용하여, 대규모 텍스트 데이터의 동시 출현 단어 빈도 계산, 텍스트의 핵심어 파악, 매트릭스를 생성하였고, 이후 의미연결망분석의 시각화 작업을 위해서 UCINET 6.0을 사용하였다.

네트워크 분석에 효과적인 분석 도구인 UCINET은 시맨틱 네트워크 분석을 위해 전 세계적으로 널리 이용되고 있으며, 본 연구에서는 핵심 키워드의 빈도 추출을 통한 텍스트의 맥락화와 키워드의 역할구조를 파악하기 위해 사용하였다. [5][9]

2.1.5 CONCOR 분석

네트워크 분석에서 블록 모델은 복잡한 관계형 구조를 단순하게 표현할 수 있다. 본 연구의 네트워크 분석을 위해 사용한 CONCOR(CONvergence of

iterated CORrelations) 분석은 단어 간의 동시출현 매트릭스의 피어슨 상관관계를 분석을 토대로 상관 행렬의 열 사이의 유사도를 측정한다. 이후 노드들의 블록을 찾고 블록들 간의 관계를 파악하여 유사성 집단(cluster)을 찾아내는 방법으로, 구조적 등위성 분석 방법 중에서 널리 활용되고 있다. 구조적 등위성(Structural Equivalence)이란 연결망 내의 노드들이 서로 어느 정도 같은 유형의 관계를 맺고 있는지를 나타내는 개념이다. 네트워크에서 노드들이 구조적 등위성을 보일 경우 비슷한 형태를 가질 확률이 높기 때문에 구조적 등위성 분석은 네트워크에서 구조적 특성이 비슷한 블록들의 등위적 계층을 찾아내는 것이다. 본 연구에서는 복잡한 네트워크에서 숨겨진 하위블록들의 관계에 대한 구조적 등위성을 확인하기 위해 CONCOR 분석을 사용하였다. [11]

3. 분석 결과

3.1 블록체인과 ‘금융’ 분석 결과

블록체인 키워드 중 ‘금융’과 관련하여 텍스트마이닝을 시행하고 상위 30건의 빈출 단어를 도출한 결과는 <표 3>과 같으며, 빈출단어와 관련된 주요 단어는 ‘핀테크’, ‘인공지능’, ‘빅데이터’ 등의 금융 기술 단어들과 ‘비트코인’, ‘거래소’, ‘규제’, ‘계좌’ 등 암호화폐 관련 단어들이 ‘금융’의 대표적 이미지로 검출되는 것을 확인할 수 있었다.

또한 상관관계 분석을 반복 수행하였고 이를 토대로 노드들의 블록을 찾고, 블록들 간의 관계를 파악하여 유사성 집단을 찾아내는 CONCOR 분석을 수행하였다.

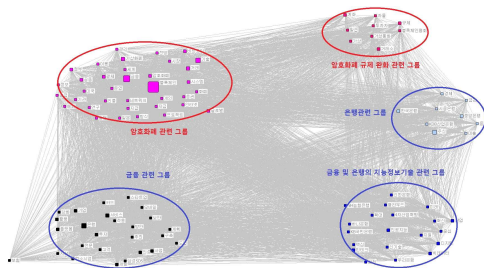


그림 1. '블록체인과 금융'의 CONCOR 분석 결과
Fig. 1. CONCOR analysis result of 'Blockchain and Finance'

표 3. '금융' 키워드 상위 주요 단어 빈도 30건
Table 3. Top 30 keyword frequency of 'Finance'

순위	키워드	빈도	순위	키워드	빈도
1	블록체인	23281	16	중앙은행	1835
2	금융	11005	17	투자	1763
3	기술	7129	18	규제	1707
4	은행	6952	19	거래소	1704
5	주식	4478	20	계좌	1647
6	가상화폐	3472	21	인증	1638
7	거래	3389	22	사업	1618
8	서비스	3179	23	인공지능	1559
9	활용	2300	24	공동	1554
10	디지털	2231	25	정부	1538
11	핀테크	2102	26	개발	1389
12	비트코인	2077	27	결제	1388
13	시스템	2075	28	산업	1318
14	기업	1914	29	한국은행	1308
15	암호화폐	1897	30	빅데이터	1260

'금융' 키워드 100건을 분석한 결과 5개의 클러스터가 생성됨을 확인할 수 있었고, 해당 클러스터들은 크게 '암호화폐' 관련 그룹과 '은행 및 금융' 그룹으로 나뉘어 나타나고 있으며, 이를 아래 <그림 1>과 같이 명명하였다.

3.2 블록체인과 '물류' 분석 결과

블록체인 키워드 중 '물류'와 관련하여 텍스트마이닝 분석을 시행하였고, 상위 30건의 빈출 단어를 도출한 결과는 <표 4>와 같다. 빈출단어와 관련된 주요

단어는 'IoT', '빅데이터', '인공지능' 등의 해운물류 기술 관련 단어들과 '삼성 SDS', '플랫폼', '컨소시엄', '업체' 등 블록체인을 활용한 물류 컨소시엄 관련 단어들인 '물류'의 대표적 이미지로 검출되는 것을 확인할 수 있었다.

또한 CONCOR 분석을 통해 '물류' 키워드 100건을 분석한 결과 6개의 클러스터가 생성됨을 확인하였다. 해당 클러스터들은 크게 '4차산업혁명 및 지능정보기술' 관련 그룹과 '해운물류 컨소시엄' 관련 그룹, '유통 이력관리' 관련 그룹 등으로 나뉘어 나타나고 있으며, 이를 위 <그림 2>와 같이 명명하였다.

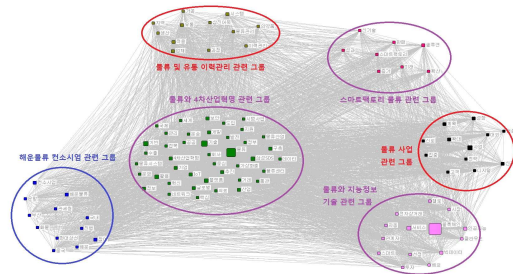


그림 2. '블록체인과 물류'의 CONCOR 분석 결과
Fig. 2. CONCOR Analysis Result of 'Blockchain and Logistics'

표 4. '물류' 키워드 상위 주요 단어 빈도 30건
Table 4. Top 30 keyword frequency of 'Logistics'

순위	키워드	빈도	순위	키워드	빈도
1	블록체인	3190	16	4차산업혁명	270
2	물류	2218	17	해운물류	270
3	기술	1208	18	개발	245
4	사업	760	19	시스템	243
5	기반	604	20	글로벌	240
6	서비스	550	21	컨소시엄	230
7	삼성SDS	489	22	인공지능	229
8	플랫폼	415	23	보안	216
9	활용	389	24	성장	212
10	산업	360	25	정보	197
11	적용	354	26	빅데이터	197
12	유통	324	27	업체	196
13	확대	280	28	추진	194
14	IoT	278	29	구축	191
15	기업	273	30	신규	184

3.3 블록체인과 ‘에너지’ 분석 결과

블록체인 키워드 중 ‘에너지’와 관련하여 텍스트마이닝 분석을 시행하고 상위 30건의 빈출 단어를 도출한 결과는 아래 <표 4>와 같다. 빈출 단어와 관련된 주요 단어는 ‘인공지능’, ‘IoT’, ‘빅데이터’, ‘4차산업혁명’ 등 지능정보기술 관련 단어들과 ‘KT’, ‘스마트에너지’, ‘스마트시티’, ‘신재생에너지’ 등 미래 에너지 활용 및 전망 등 블록체인을 활용한 에너지 사업 관련 단어들인 ‘에너지’의 대표적 이미지로 검출되는 것을 확인할 수 있었다.

또한 CONCOR 분석을 통해 ‘에너지’ 키워드 100건을 분석한 결과 5개의 클러스터가 생성됨을 확인하였다. 해당 클러스터들은 크게 ‘지능정보기술과 에너지’ 관련 그룹과 ‘가상화폐와 에너지’ 관련 그룹, ‘전력거래 및 신재생에너지’ 관련 그룹 등으로 나누어 나타나고 있으며, 이를 위 <그림 3>과 같이 명명하였다.

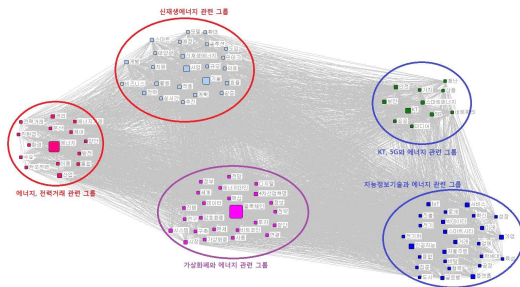


그림 3. ‘블록체인과 에너지’의 CONCOR 분석 결과
Fig. 3. CONCOR analysis result of ‘Blockchain and Energy’

표 5. ‘에너지’ 키워드 상위 주요 단어 빈도 30건
Table 5. Top 30 Keyword Frequency about ‘Energy’

순위	키워드	빈도	순위	키워드	빈도
1	블록체인	2760	16	스마트에너지	298
2	에너지	1970	17	시스템	255
3	기술	1215	18	스마트	252
4	사업	821	19	거래	239
5	KT	535	20	보안	212
6	인공지능	532	21	신재생에너지	208
7	플랫폼	433	22	안전	196
8	산업	420	23	효율	191
9	서비스	415	24	4차산업혁명	189
10	기업	387	25	빅데이터	189
11	관리	332	26	개발	182
12	적용	306	27	스마트시티	178
13	활용	303	28	글로벌	178
14	IoT	303	29	미래	175
15	전력	299	30	자율주행	167

4. 결론 및 시사점

본 연구는 2017년부터 2018년 7월까지 약 2년 동안의 블록체인과 관련된 금융, 물류, 에너지라는 세 가지 용어에 관한 관련 기사를 수집하고, 텍스트마이닝 분석과 의미연결망 분석을 통해 단어별 연관어와 관련 현상들을 파악하였으며, 분석 결과는 다음과 같다.

첫째, 블록체인과 금융, 물류, 에너지 키워드라는 각각의 분류에서 공통적으로 향후 발전 가능성이나 영향관계가 있을 ‘인공지능’, ‘빅데이터’, ‘IoT’와 같은 지능정보기술들의 빈도가 높게 나타났다. 이는 앞서 서론에서 언급한 과학기술정보통신부의 계획과 같이 해당 산업 분야에서 블록체인이 다른 기술들과 함께 주요 기술로 거론되고 이미 활용되고 있다는 점을 확인할 수 있다.

둘째, ‘암호화폐’, ‘가상화폐’ 등이 3가지 키워드에서 모두 언급되고 있는데, ‘금융’ 키워드에서는 일반적인 비트코인과 같은 거래소 중심의 투자나 투기 의미가 많지만, ‘물류’나 ‘에너지’ 키워드에서는 ‘전력거래’를 위해 새롭게 활용되는 수단으로, ‘물류 및 유통거래 이력’에서는 유통 과정 등을 확인하는 것에 주어지는 보상 수단 등으로 다르게 사용됨을 확인할 수 있다. 즉

같은 '가상화폐', '암호화폐'이지만 그 용도는 키워드별로 판이함을 확인할 수 있다.

셋째, 블록체인과 관련하여 3가지 분야에 대해 활발하게 사업을 추진하거나 연구를 진행하고 있는 은행 기관 및 민간기업(KT, 삼성SDS)을 파악할 수 있었다.

금융권에서는 다수의 은행이 블록체인 기술을 이용하여 금융 거래나 결제, 인증 등에서 연구를 진행하고 있음을 알 수 있다. 또한 물류에서는 삼성SDS를 필두로 해운물류에 블록체인을 적용하고 있고, 정부의 유통이력추적, 관리 등의 연구 및 사업이 진행됨을 알 수 있었다. 마지막으로 에너지 분야에서는 KT를 비롯한 기업들이 신재생에너지, 스마트에너지 등을 내세우며 연구 및 사업에 블록체인 기술을 활용하고 있음을 알 수 있었다. 이러한 기업 및 기관들은 블록체인과 자신들의 사업 관련 이슈들을 보도하여 용어에 대한 대중 인식과 영향력을 점차 키우고 있는 것으로 보인다.

본 연구는 2018년 6월, 과학기술정보통신부에서 발표한 「블록체인 기술 발전전략」에 기재된 블록체인이 활용될 6가지 산업 분야 중 '금융', '물류', '에너지' 키워드와 관련하여 2017년부터 보도된 기사 분석을 통해 각각의 용어에 함의되어 있는 의미를 파악해 보았다. 이러한 연구 결과를 바탕으로 향후 정부에서는 블록체인 산업진흥을 위한 심층적인 분석과 향후 주력해야 할 방향을 도출하여 그에 관한 적극적인 연구가 진행되어야 할 것으로 보인다. 또한 블록체인에 대한 가장 높은 관심을 보이고 있는 금융권, 물류회사, 에너지 산업계의 키워드 별 연관어에 대한 동향 분석을 하였다. 이는 향후 블록체인 산업의 방향을 정립하는 과정에서 참고자료로서 사용될 수 있을 것으로 판단된다.

REFERENCES

[1] Ministry of Science and ICT Press, "Blockchain Technology Development Strategies For Implementing Reliable Fourth Industrial Revolution", Ministry of Science and ICT, 2018.
 [2] Ju-Seop Park, Na-Rang Kim, Eun-Jung Han, "Analysis of Trends in Science and Technology using Keyword Network Analysis", Journal of the Korea Industrial Information Systems Research, v.23, no.2, pp.63-73, 2018.

[3] Hyun-Ju Jo, "Comparative Analysis of Topics of Research on Artificial Intelligence using the Topic Modeling", Master's Thesis, Graduate School of Yonsei Univ., 2018.
 [4] Eun-Bi Song, "Analysis of news on bitcoin using textmining", Master's Thesis, Graduate School of Ewha Womans Univ., 2018
 [5] Y. H. Cho, H. S. Son, "Heuristic Analysis on Virtual Money, Cryptocurrency, Virtual Currency by Using Text Mining through Naver News Articles", Textom, 2018.
 [6] Ryan Mitchell, Web Scraping with Python 2nd ed., pp.6-32, O'Reilly Media Inc. 2018.
 [7] Won-ha Lee, Big data Collecting, Analysis and Visualization with Python, BPAN Books, 2017.
 [8] Hyun-jung Kim, Nam-ok Jo, Kyung-shik Shin, "Text Mining-Based Emerging Trend Analysis for the Aviation Industry", Journal of Intelligence Information System, v.21, no.1, pp.65-82, 2015.
 [9] Keun-ha Chung, "A Study of foresight method based on textmining and complexity network analysis", Korea Institute of Science and Technology Evaluation and Planning, 2010.
 [10] Wang, W. & Rada, R, "Structured hypertext with domain semantics", ACM. Trans. Inform. Syst., 16, pp.372-412, 1998.
 [11] Yong-hak Kim, Social Network Analysis, 3rd ed., pp.70-85, Pakyongsa Press, 2016.

저자약력

조 성 환(Seongl-Hwan Cho)

[정회원]



- 1980년 성균관대학교 전자공학과(학사)
 - 1982년 성균관대학교 대학원 전자공학과(공학석사)
 - 1991년 성균관대학교 대학원 전자공학과(공학박사)
 - 1982년~1985년 해군사관학교 전기 및 전자공학과 전임강사
 - 1997년 미국 Columbia 대학 CATT Visiting Scholar
 - 1985년-2002년 동서울대학 컴퓨터공학과 부교수
 - 2003년-현재 금강대학교 글로벌 융합학부 IT 소프트웨어 교수
- 빅데이터, 저작권보호기술(DRM), 인공지능경망

<관심분야>