

뉴스 빅데이터를 활용한 수소 이슈 탐색

최일영¹ · 김혜경^{2†}

¹경희대학교 경영대학원, ²울산연구원

A Study on Social Issues for Hydrogen Industry Using News Big Data

ILYOUNG CHOI¹, HYEK-KYEONG KIM^{2†}

¹Graduate School of Business Administration, Kyunghee University, 26 Kyungheedaero, Dongdaemun-gu, Seoul 02447, Korea

²Ulsan Research Institute, 6 Samsanjung-ro, Nam-gu, Ulsan 44720, Korea

†Corresponding author :
hkkim@uri.re.kr

Received 13 January, 2022

Revised 18 March, 2022

Accepted 9 April, 2022

Abstract >> With the advent of the post-2020 climate regime, the hydrogen industry is growing rapidly around the world. In order to build the hydrogen economy, it is important to identify social issues related to hydrogen and prepare countermeasures for them. Accordingly, this study conducted a semantic network analysis on hydrogen news from NAVER. As a result of the analysis, the number of hydrogen news in 2020 increased by 4.5 times compared to 2016, and as of 2018, the hydrogen issue has shifted from an environmental aspect to an economic aspect. In addition, although the initial government-led hydrogen industry is expanding to the mobility field such as privately-led fuel cell electric vehicles and hydrogen fuel, terms showing concerns about the safety such as explosions are constantly being exposed. Thus, it is necessary not only to expand the hydrogen ecosystem through the participation of private companies, but also to promote hydrogen safety.

Key words : Hydrogen industry(수소산업), Text mining(텍스트 마이닝), Semantic network(의미연결망), News big data(뉴스빅데이터), Term frequency-inverse document frequency(단어빈도-역문서빈도), CONCOR analysis(반복상관관계수렴 분석)

1. 서론

기후변화 및 화석연료의 고갈 문제가 국제사회의 중요한 의제로 대두되고 온실가스 배출에 대한 환경 규제가 강화되면서 수소는 화석연료를 대체할 유력

한 대안으로 부상하였다. 수소가 신기후체제에 적합한 에너지로 평가받는 것은 지역 편중이 없고, 사용 과정에서 탄소를 배출하지 않아 친환경성이 높다는 데 기인한다. 더불어, 국제사회에서 130개국 이상이 2050 탄소중립을 선언하면서 대용량 재생에너지의

안정적 저장과 유통에 적합한 에너지 캐리어의 역할 또한 주목받고 있다.

글로벌 수소경제는 2050년 2.940조 원의 시장과 3천만 개 일자리를 창출할 것으로 전망된다¹⁾. 수소경제는 수소를 주요 에너지원으로 사용하는 경제산업 구조로 수소에너지의 안정적 생산부터 활용에 필요한 모든 분야의 산업과 시장을 포함한다. 수소산업은 수소전기차와 연료전지 분야 전방산업의 가치 창출 기대가 높으며, 밸류체인 전반에 걸쳐 다양한 산업과 연계되어 있어 후방산업 파급효과 또한 높은 특징이 있다. 이같이 수소산업은 경제적 파급효과가 큰 혁신 산업이라는 점에서 시장 선점을 위한 주요국들의 경쟁이 치열하다.

한국 정부는 2018년 8월 「혁신성장 전략투자 방향」에서 수소경제를 전략투자 분야로 선정하였다. 선정 이유는 온실가스 저감, 신재생에너지 보급 확대, 신산업 육성으로 환경과 경제 정책을 동시에 추구한다. 이후, 2019년 1월 국가차원의 「수소경제 활성화 로드맵」을 수립하였다. 로드맵에서 한국의 수소산업 양대 축인 수소차와 연료전지가 제시되었고 2040년까지 수소차 620만 대 생산, 수소충전소 1,200개 구축, 발전용 연료전지 15 GW 보급 목표를 설정하였으며, 목표 달성을 통해 2040년에는 연간 43조 원의 부가가치와 42만 개의 새로운 일자리 창출을 기대하고 있다.

정부는 수소경제 활성화 로드맵 후속대책으로 수소경제 표준화 전략 로드맵(2019년 4월), 수소 인프라 및 충전소 구축방안(2019년 10월), 미래자동차 산업 발전전략(2019년 10월), 수소 기술개발 로드맵(2019년 10월), 수소시범도시 추진전략(2019년 10월), 수소안전관리 종합대책(2019년 12월)을 순차적으로 발표하였고, 2020년 2월 세계 최초로 수소 경제 육성 및 수소 안전관리에 관한 법률을 제정하여 수소경제 선도국가 조기 진입과 정착을 위한 기틀을 마련하였다.

국가주도의 정책 수립에 이어, 최근에는 글로벌 기업의 RE100 참여 및 ESG 투자 확대 등 환경을 고려한 민간부문의 경영활동과 지방정부의 수소산업 육성 정책 발표가 경쟁적으로 이어지고 있다. 또한,

다양한 이해관계자들이 수소경제 진입을 위해 정책 추진 과정의 문제점, 해결책 등의 의견을 개진하고 있다. 일반적으로 이들 정보와 소식이 뉴스를 통해 생산되고 확산된다. 따라서 뉴스를 수집하여 분석하는 것은 다양한 관점에서 수소 분야 이슈의 내용과 강도를 파악하는 데 유용하다²⁾.

그러나 기존의 수소산업 관련 연구는 수소충전소의 경제성 분석^{3,6)}, 수소 인프라 전략^{7,8)}, 해외사례분석⁹⁾, 수소충전소의 안전성 연구^{10,11)} 등으로 기술적 또는 전략적 관점의 연구가 대다수이다.

수소산업은 친환경이라는 기대와 수소폭발이라는 우려와 불안이 공존한다. 수소경제로의 진입과 안착을 위해서는 수소에 대한 사회적 이슈를 파악하여 적정 대응책을 마련하는 것이 중요하다.

본 연구에서는 온라인 뉴스의 키워드를 활용하여 의미연결망을 구축한 후 convergence of iterated correlations (CONCOR) 분석 등을 통해 수소분야 이슈를 도출하고자 한다. 이를 기반으로 이슈의 흐름과 함의를 파악하고 지속가능한 수소경제를 위한 시사점을 제시하고자 한다.

2. 이론적 배경

2.1 TF-IDF

뉴스 등의 텍스트 데이터 분석 및 정보검색을 위해 의미 있는 단어를 추출하는 것이 중요하다. 이를 위해 일반적으로 많이 사용되는 기법인 term frequency-inverse document frequency (TF-IDF) 기법으로, 단어의 상대적 중요도를 추출한다. TF-IDF 기법은 식 (1)과 같이 한 문서 내에 자주 사용된 단어는 중요한 단어이며, 많은 문서에서 자주 사용된 단어는 보편적인 단어라는 것을 가정한다.

$$TF-IDF = tf(t, d) \times \log\left(\frac{D}{df(t)}\right) \quad (1)$$

여기서 $tf(t, d)$ 와 $df(t)$ 는 각각 문서 d 에서 단어 t 의 빈도와 단어 t 를 포함하고 있는 문서의 빈도를 나

타내며, D 는 문서의 개수이다.

이러한 TF-IDF는 문서분류¹²⁻¹⁴⁾, 추천시스템¹⁵⁻¹⁷⁾, 감정분석^{18,19)}, 의미연결망^{20,21)} 등 다양한 분야에서 활용되고 있다. 따라서 본 연구에서는 의미연결망을 구축하기 위하여 TF-IDF를 통해 뉴스 빅데이터를 자연어 처리하여 상위 100개의 중요 단어를 추출하였다.

2.2 CONCOR 분석

의미연결망(semantic network)은 문서 내에 동시에 사용된 단어를 링크로 연결하는 네트워크로, 단어 들 간의 구조적 관계를 분석하기 위해 사용되며, 특히 특허데이터 분석^{22,23)}, 문서분석²⁴⁾ 등에서 많이 활용되고 있다.

일반적으로 의미연결망에서 이슈를 도출하기 위해 CONCOR 분석이 많이 활용되고 있다. CONCOR 분석은 단어 간의 상관관계를 이용하여 동일한 패턴을 가지는 단어를 군집화 하는 구조적 등위성을 구하는 방법이다. CONCOR 분석은 지급결제²⁵⁾, 국방정책²⁶⁾, 산림치유²⁾, 관광지²⁷⁾, 블록체인²⁸⁾ 등 다양한 분야에서 이슈를 도출하기 위해 사용되고 있다. 따라서 본 연구에서는 CONCOR 분석을 수행하여 수소 관련 뉴스의 이슈를 도출하였다.

3. 연구 방법

본 연구의 목적은 뉴스 빅데이터로부터 수소 관련 이슈를 발견하는 것이다. 이러한 목적을 달성하기 위한 연구 프로세스는 Fig. 1과 같이 4단계로 구성되어 있다.

첫 번째 단계는 뉴스를 수집하는 단계이다. 뉴스 수집 기간은 국내의 수소 시장이 급격히 성장한 기점부터 주요국 정책 발표가 이어진 2020년까지로 한정하였다. 2015년 12월 한국 정부는 「수소차 보급 및 시장 활성화 계획」을 발표하면서 2016년부터 수소시장의 성장을 견인하였으며, 2019년 1월 수소경제 활성화 로드맵을 시작으로 2020년까지 정부의 후속대책 발표를 통해 수소산업 육성에 대한 강력한 의지

를 표명하였다. 본 연구에서는 국내 대표 포털사이트인 네이버(<https://www.naver.com>)에서 2016년 1월 1일부터 2020년 12월 31일까지 5년 동안 수소 단어를 포함하고 있는 뉴스를 수집하였다.

두 번째 단계는 뉴스를 전 처리하는 단계로, 뉴스에서 수소가 들어가 있는 문장만 추출한 후 R 프로그램의 tm 패키지를 이용하여 이음동어어 및 불용어 처리, 숫자, 구두점 및 특수문자 등을 제거하였다. 그리고 연구 범위 내 이슈와 관련성이 낮은 북한 수소 폭탄 등의 단어를 제거한 후, KoNLP 패키지와 한글 형태소 사전 NIADic을 이용하여 3글자 이상의 명사만을 추출하였다.

세 번째 단계는 TF-IDF를 통해 중요 단어를 추출하는 단계로, 100회 이상 뉴스에 노출된 단어 중 TF-IDF 값이 큰 100개의 단어를 도출하였다.

마지막 단계는 수소 관련 뉴스의 특징을 살펴보기 위해 구조적 등위성 분석을 수행하는 단계로, 세 번째 단계에서 추출된 100개 단어에 대해 Ucinet을 이용하여 의미연결망을 구축한 후 CONCOR 분석을 수행하였다.

4. 분석 결과

4.1 데이터 수집

네이버에서 수소 단어가 포함된 뉴스를 수집한 후,

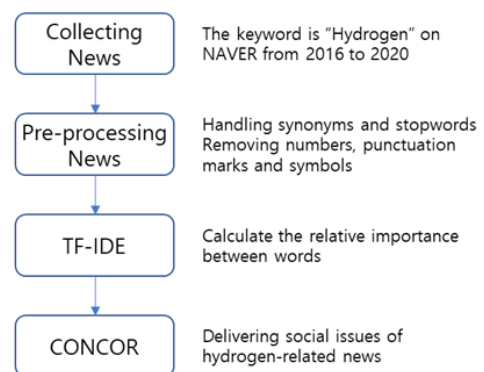


Fig. 1. Framework of the study

속보, 동일 뉴스 등을 제거한 결과, 분석에 사용된 뉴스는 총 9,501건이고 연차별 건수는 Fig. 2와 같다. 2016년 이후 수소경제활성화 로드맵이 발표된 2019년까지 수소관련 뉴스량은 연평균 97.3%의 폭발적 증가세를 나타낸 이후, 2019년을 정점으로 2020년에는 크게 감소하였다. 그러나 2020년 뉴스생산량은 2016년 대비 4.5배 증가한 수치로 수소분야에 대한 경제사

회적 관심이 증폭되었음을 분명히 보여준다.

4.2 연차별 빈도분석

수소 관련 이슈의 시계열적 특징과 흐름을 파악하기 위하여 연차별 단어의 빈도분석한 결과는 Fig. 3과 같다. 2016년은 수소차 및 연료전지의 보급을 정책 수단으로 저탄소로의 전환정책이 추진된 시점이다. 「수소차 보급 및 시장 활성화 계획」에 이어 「미세먼지 관리 특별대책(2016년 6월)」과 「제1차 기후변화 대응 기본계획(2016년 12월)」을 통해 수소차(~2030년, 64만 대)와 수소충전소를 보급하여 온실가스를 감축(~2030년, 0.8백만 톤)하려는 중장기 목표를 설정하고 세부 이행계획을 공표한 시기이다. 당시는 경제적 관점에서의 수소 이슈보다는 수소차 보급을 통한 국가의 온실가스 감축 목표 달성과 기후변화 대응 등 환경적 관점이 두드러진 특징이 있다. 뉴스에 노출된 단어의 빈도는 이를 반영하여 수소전기자동차(1,812회),

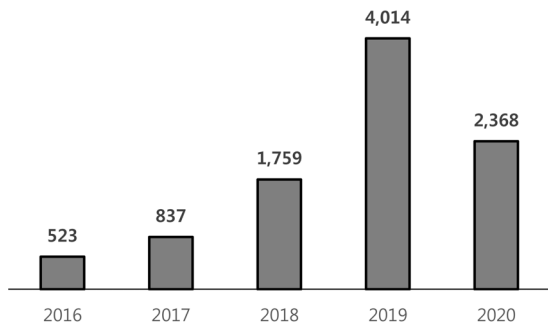


Fig. 2. Number of news related on hydrogen by year from 2016 to 2020



Fig. 3. Major news keyword related to hydrogen by year from 2016 to 2020

자동차(1,099회), 수소연료전지(267회), 수소충전소(235회), 에너지(213회), 울산광역시(186회), 연료전지(182회), 친환경자동차(164회), 광주광역시(134회), 스테이션(117회) 등의 순으로 나타났다.

자동차와 충전소 외에 빈도 높은 단어는 울산광역시와 광주광역시로 현대자동차그룹의 생산 공장이 위치한 지역이라는 공통점이 있다. 이 지역은 수소차 관련 민관합동 시범사업과 인프라 구축이 조기 시행된 지역이다. 특히 울산은 세계 최초 수소전기차가 양산된 도시로 2016년 전국 최초 수소택시가 운행되면서 언론 노출 빈도가 상대적으로 높았다.

2017년 수소 관련 뉴스는 전년 대비 160% 증가하였으나, 내용적으로는 유사한 흐름으로 전개되었다. 다만, 2017년에는 에너지전환정책 추진을 위한 「제3차 에너지기본계획」과 「재생에너지 3020 이행계획(안)」이 발표되면서 에너지와 연료전지 등의 단어 노출이 증가하였다. 2017년도 뉴스에 노출된 단어의 빈도는 이를 반영하여 수소전기자동차(3,735회), 자동차(2,220회), 에너지(475회), 수소전기버스(395회), 친환경자동차(385회), 수소연료전지(359회), 울산광역시(354회), 수소충전소(353회), 충전소(352회), 수소에너지(307회) 등의 순으로 나타났다.

2018년은 신정부의 「혁신성장 전략투자 방향」에서 수소경제를 3대 투자 분야 중 하나로 선정한 해로, 경제적 관점에서의 수소 관련 뉴스 생산이 급증한 시기이며, 양적으로도 전년 대비 210% 이상 증가한 것으로 나타났다. 키워드 빈도는 수소전기자동차(7,060회), 자동차(3,696회), 수소충전소(1,689회), 수소전기버스(1,358회), 울산광역시(804회), 충전소(758회), 에너지(695회), 수소산업(692회), 수소연료전지(587회), 수소에너지(555회) 등의 순으로 나타났다.

울산광역시는 자동차 뿐 아니라 석유화학산업단지 보유하고 있어 석유화학 공정에서 발생하는 부생수소를 기반으로 전국 수소생산량의 50% 이상을 생산·공급하는 지역으로 부각되면서 지속적으로 뉴스에 노출된 것으로 나타났다.

2019년은 수소차와 연료전지를 두 축으로 한 「수소경제 활성화 로드맵」이 발표되면서 국내 수소산업

육성이 본격화된 시기이다. 이를 반영하여 ‘수소경제’와 ‘수소산업’ 등의 키워드가 새롭게 등장하거나 빈도를 높였으며, 수소차 보급이 전국적으로 확대되면서 수소충전소 수요 관련 이슈가 증가한 것으로 나타났다. 2019년도 뉴스에 노출된 단어의 빈도는 수소전기자동차(9,571회), 수소충전소(5,567회), 자동차(4,963회), 수소경제(4,905회), 에너지(2,894회), 수소연료전지(1,968회), 충전소(1,879회), 수소전기버스(1,834회), 수소산업(1,832회), 연료전지(1,827회) 등의 순이다.

마지막으로 2020년의 뉴스 키워드 빈도는 수소전기자동차(5,059번), 자동차(4,076번), 수소충전소(2,790번), 수소연료전지(1,871번), 수소경제(1,821번), 에너지(1,331번), 수소전기트럭(1,255번), 수소전기버스(1,171번), 수소생산(979번), 충전소(911번) 등의 순으로 나타났다. 2020년 수소관련 뉴스의 특징은 수소의 생산방식과 활용 범위의 확대라 할 수 있다. 수소 생산에 있어서는 화학반응에 의해 부수적으로 생산되는 부생수소보다 친환경성이 높은 개질수소, 수전해수소에 대한 관심과 요구가 높아졌으며, 활용 분야에 있어서도 수소자동차 이외에 수소연료를 사용하는 다양한 모빌리티가 부상하면서 관련 단어의 빈도가 증가한 것으로 나타났다.

2016-2020년에 노출 빈도가 높은 수소 관련 키워드는 수소전기자동차(27,237회), 자동차(16,054회), 수소충전소(10,634회), 수소경제(7,222회), 에너지(5,608회), 수소연료전지(5,052회), 수소전기버스(4,828회), 충전소(3,998회), 울산광역시(3,681회), 연료전지(3,660회) 등의 순이다. 이를 연차별로 분석해 본 결과, 수소 관련 이슈는 환경적 관점과 경제적 관점이 공존하는 분야이지만 강도의 측면에서 보면 2018년을 기준으로 수소는 환경적 관점의 이슈에서 경제적 관점의 이슈로 전환된 것으로 나타났다. 다만, 최근 탄소중립 논의의 확산과 함께 수소경제 또한 친환경성을 높이는 방향으로의 모색이 두드러지면서 부생수소 및 관련 지역에 대한 키워드 노출 빈도가 낮아진 것을 확인하였다. 이를 반영하여 정부부처의 보도주제도 환경부에서 경제산업 및 기술개발 관련 부처로

확대된 것을 확인할 수 있었다.

한편, 시간이 경과하면서 수소차 보급과 같은 단순 정책에서 수소의 생산-유통-활용을 포괄하는 생태계 조성 측면에서의 뉴스가 증가하였으며, 정부 정책 외에 민간 기업의 기술개발 및 경영활동에 관한 뉴스가 증가한 것으로 나타났다.

4.3 TF-IDF 분석

분석 대상 단어의 뉴스 노출 빈도와 TF-IDF 간의 상관관계를 살펴보면 Table 1과 같다. 상관계수는

Table 1. Correlation between frequency and TF-IDF

	Frequency	TF-IDF
Frequency	1	
TF-IDF	0.842	1

Table 2. Top 20 keyword result of the TF-IDF analysis

Keyword	TF-IDF
Hydrogen electric vehicle	660.22
Hydrogen refueling station	545.63
Automobile	523.88
Hydrogen electric bus	501.48
Hydrogen fuel cell	430.31
Hydrogen economy	400.08
Ulsan Metropolitan City	363.91
Energy	327.45
Hydrogen fluoride	296.11
Hydrogen tank	285.17
Charging station	276.56
Fuel cell	268.92
Hydrogen production	255.04
Hydrogen industry	247.91
Hydrogen energy	247.45
Hydrogen sulfide	191.58
Eco	185.58
Eco-friendly car	182.82
President	180.21
Gangwon-do	174.30

0.842로 강한 양의 상관관계를 보여 노출 빈도가 높은 단어는 중요한 단어라고 판단할 수 있다.

TF-IDF 값이 높은 상위 20개의 단어를 살펴보면 Table 2와 같이 수소전기자동차, 수소충전소, 자동차, 수소전기버스, 수소연료전지, 수소경제, 울산광역시, 에너지, 불화수소, 수소탱크, 충전소, 연료전지, 수소 생산, 수소산업, 수소에너지, 황화수소, 친환경, 친환경 자동차, 대통령, 강원도 순이다.

이같은 수소 관련 중요 단어는 한국의 수소 이슈의 세 가지 특성을 드러낸다. 첫째는 수소를 에너지 원으로 사용하는 수소모빌리티 중심의 산업육성이며, 둘째는 기업보다는 중앙·지방정부 중심의 관주도적 특성이며, 마지막으로 폭발사고 등에 대한 안전 우려의 상존이다.

향후, 주기적인 TF-IDF 분석을 통해 수소산업의 활성화 흐름과 국민인식 및 수용성 수준을 파악하는 것은 정책평가와 대응방안 마련에 필요한 정보를 산출하는 데 유용한 방법이 될 것이다.

4.4 구조적 등위성 분석

TF-IDF 분석을 통해 추출된 상위 100개의 단어를 이용하여 수소 이슈에 대한 구조적 이해를 위해 의미연결망을 통한 CONCOR 분석을 수행하였다. 분석 결과는 100개의 단어는 Fig. 4와 같이 4개의 군집을 형성하였다. 군집 1의 주요 단어는 수소전기자동차, 수소충전소, 자동차, 수소전기버스, 수소연료전지, 수

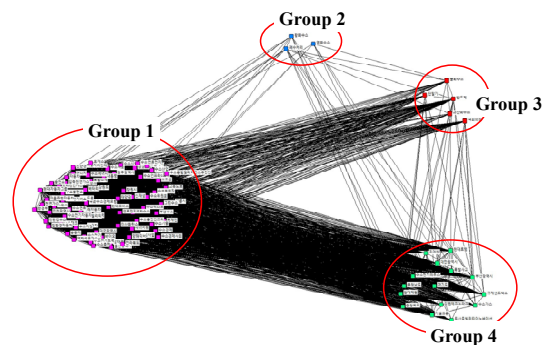


Fig. 4. Result of the CONCOR analysis

소경제, 울산광역시, 에너지, 수소탱크, 충전소 등으로 수소 비즈니스와 수소산업을 육성계획을 수립하고 있는 지방자치단체와 관련되어 있음을 알 수 있다. 군집 2의 주요단어는 황화수소, 폐수처리, 염화수소 등으로 폐기물 합성가스와 관련되어 있다. 군집 3의 주요단어는 불화수소, 과산화수소, 반도체, 국회 의장, 단말기 등으로 반도체 공정과 관련되어 있다. 마지막으로 군집 4의 주요 단어는 부산광역시, 폭발 사고, 충청남도, 경기도, 강원테크노파크, 두산모빌리티이노베이션, 대전광역시, 현대로템, 수소전기하우 스, 수소가스 등으로 지방자치단체 및 가스폭발과 관련되어 있다.

5. 결론

본 연구는 뉴스에 나타난 수소 관련 단어를 분석하여, 상위 빈도 및 TF-IDF 단어를 파악하고, CONCOR 분석을 통하여 이슈를 도출하였다. 이를 위하여 네이버 뉴스에서 2016년 1월 1일부터 2020년 12월 31일까지 제목 또는 본문에 수소 단어를 포함하고 있는 기사를 수집 및 전 처리하였다.

분석 결과는 다음과 같다. 첫째, 수소 관련 뉴스생산량은 5년간 연평균 46% 증가하여 다양한 이슈의 생산과 장기에 걸쳐 관심의 증가 추이를 보여주었다.

둘째, 수소 관련 이슈는 환경적 관점과 경제적 관점이 공존하는 분야이지만 강도의 측면에서 2018년을 기준으로 환경적 관점의 이슈에서 경제적 관점의 이슈가 강화되는 방향으로 전개된 것으로 나타났다. 그러나 최근 국제사회의 탄소중립 의제화와 함께 수소경제 또한 친환경성을 높이는 방향으로의 모색이 강화되면서 재생에너지를 활용한 수소생산에 관한 논의가 증가하였고 부생수소 및 관련 지역에 대한 단어의 노출 빈도가 감소추세인 것을 확인하였다.

셋째, 시간이 경과하면서 수소차 보급과 같은 단순 정책에서 수소의 생산-유통-활용을 포괄하는 생태계 조성 측면에서의 뉴스가 증가하였으며, 수소자동차 이외에 수소연료를 사용하는 다양한 모빌리티 관련 단어가 증가하여 활용 분야의 확대가 모색됨을

확인할 수 있었다.

넷째, 기업보다는 중앙·지방정부 중심의 관주도적 특성이 두드러졌다. 그러나 시계열적으로 보면 민간 기업의 기술개발 및 경영활동에 관한 뉴스가 증가한 것으로 나타났다.

마지막으로 폭발사고 등의 안전에 대한 우려를 보여주는 단어가 꾸준히 노출되고 있어 국민 안전성 우려를 해소할 수 있는 대책이 필요한 것으로 나타났다. 수소의 사용과 관련 시설의 확대에 앞서 안전 기준 정비 및 관리체계 마련이 긴요하며, 국민인식 및 주민 수용성 제고를 위한 정책이 마련되어야 할 것이다.

향후, 뉴스 빅데이터의 주기적인 분석은 앞서 제기한 수소산업 생태계 조성을 위한 밸류체인 별 이슈의 점검과 민간기업의 참여 확대, 국민의 수용성 제고 등에 대한 흐름을 이해하고 대응책을 마련하는데 있어 유용한 방법이 될 것이다.

그러나 본 연구는 다음과 같은 한계점이 있다. 첫째, 네이버에서 제공하는 뉴스를 기반으로 빅데이터를 수집하였다. 그러나 네이버가 제공하는 뉴스 이외에도 많은 언론사가 존재하기 때문에 이를 고려하지 못한 한계점이 있다. 둘째, 수소 관련 뉴스의 이슈를 파악하기 위하여 의미연결망의 구조적 등위성 분석을 수행하였다. 그러나 의미연결망에서 이슈를 도출할 수 있는 중심성 등의 측정지표가 있음에도 고려하지 못하였다. 향후 이러한 한계를 고려한 후속 연구를 통해 급변하는 수소 분야 환경과 트렌드를 분석하는 데 좋은 연구가 될 것으로 판단된다.

References

1. Hydrogen Council, "Hydrogen scaling up: a sustainable pathway for the global energy transition", Hydrogen Council, 2017. Retrieved from <https://hydrogencouncil.com/wp-content/uploads/2017/11/Hydrogen-scaling-up-Hydrogen-Council.pdf>.
2. J. S. Park and P. S. Yeoun, "A network analysis on the forest healing issues using big data - focused on Korean web news from 2005 to 2019 -", The Journal of Korean Institute of Forest Recreation, Vol. 24, No. 2, 2020, pp. 63-71, doi:

- <https://doi.org/10.34272/forest.2020.24.2.006>.
3. B. W. Kang, T. H. Kim, and T. H. Lee, "Analysis of costs for a hydrogen refueling station in Korea", *Trans Korean Hydrogen New Energy Soc*, Vol. 27, No. 3, 2016, pp. 256-263, doi: <http://dx.doi.org/10.7316/KHNES.2016.27.3.256>.
 4. B. J. Kim, "An economic analysis of the hydrogen station enterprise considering dynamic utilization", *Trans Korean Hydrogen New Energy Soc*, Vol. 28, No. 1, 2017, pp. 47-55, doi: <https://doi.org/10.7316/KHNES.2017.28.1.47>.
 5. B. J. Kim and J. W. Kim, "An analysis of the economy of scale for domestic on-site hydrogen fueling stations", *Journal of Energy Engineering*, Vol. 16, No. 4, 2007, pp. 170-180. Retrieved from <http://www.koreascience.or.kr/article/JAKO200707341597329.pub>.
 6. B. J. Kim, W. L. Yoon, and D. J. Seo, "Analysis of the economy of scale for domestic steam methane reforming hydrogen refueling stations utilizing the scale factor", *Trans Korean Hydrogen New Energy Soc*, Vol. 30, No. 3, 2019, pp. 251-259, doi: <https://doi.org/10.7316/KHNES.2019.30.3.251>.
 7. B. I. Choe, "A Korean strategy for the hydrogen society infrastructure based on the liquefied hydrogen", *Superconductivity and Cryogenics*, Vol. 22, No. 1, 2020, pp. 9-12. Retrieved from <https://www.earticle.net/Article/A370584>.
 8. J. H. Han, S. J. Kim, and C. B. Kim, "A strategy development of hydrogen energy industrial infrastructure by using SWOT/AHP method", *Journal of Korea Technology Innovation Society*, Vol. 19, No. 4, 2016, pp. 822-847. Retrieved from <https://www.koreascience.or.kr/article/JAKO201610364778833.j>.
 9. Y. Y. Kim, J. S. Park, and J. S. Jeong, "A study on overseas case analysis and empirical analysis for the development of the hydrogen industry in Chungcheongbuk-do", *International Business Review*, Vol. 24, No. 2, 2020, pp. 77-90. Retrieved from <http://www.kaibm.or.kr/html/sub0303.html?pageNm=article&journal=1&code=378914&issue=28464&Page=2&year=2020&searchType=title&searchValue=>.
 10. S. K. Kang, Y. S. Huh, and J. S. Moon, "A study on safety improvement for packaged hydrogen refueling station by risk assessment", *Trans Korean Hydrogen New Energy Soc*, Vol. 28, No. 6, 2017, pp. 635-641, doi: <https://doi.org/10.7316/KHNES.2017.28.6.635>.
 11. K. W. Rhie, T. H. Kim, and T. H. Lee, "Quantitative safety assessment for hydrogen station dispenser", *Trans Korean Hydrogen New Energy Soc*, Vol. 17, No. 3, 2006, pp. 272-278. Retrieved from <https://www.hydrogen.or.kr/upload/papers/KHNES.Vol.17,No.03-05.pdf>.
 12. J. S. Jeong, M. K. Jee, M. H. Go, H. D. Kim, H. Y. Lim, Y. R. Lee, and W. I. Kim, "Related documents classification system by similarity between documents", *Journal of Broadcast Engineering*, Vol. 24, No. 1, 2019, pp. 77-86, doi: <https://doi.org/10.5909/JBE.2019.24.1.77>.
 13. P. Bafna, D. Pramod, and A. Vaidya, "Document clustering: TF-IDF approach", 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016, pp. 61-66, doi: <https://doi.org/10.1109/ICEEOT.2016.7754750>.
 14. D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec", *Information Sciences*, Vol. 477, 2019, pp. 15-29, doi: <https://doi.org/10.1016/j.ins.2018.10.006>.
 15. B. K. Jeon and H. C. Ahn, "A collaborative filtering system combined with users' review mining: application to the recommendation of smartphone apps", *J. Intell. Inform. Syst.*, Vol. 21, No. 2, 2015, pp. 1-18, doi: <https://doi.org/10.13088/jiis.2015.21.2.01>.
 16. R. Iwabuchi, Y. Nakajima, H. Honma, H. Aoshima, A. Kobayashi, T. Akiba, and S. Masuyama, "Proposal of recommender system based on user evaluation and cosmetic ingredients", 2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA), 2017, pp. 1-6, doi: <https://doi.org/10.1109/ICAICTA.2017.8090967>.
 17. Z. K. Zhang, T. Zhou, and Y. C. Zhang, "Tag-aware recommender systems: a state-of-the-art survey". *Journal of Computer Science and Technology*, Vol. 26, 2011, pp. 767, doi: <https://doi.org/10.1007/s11390-011-0176-1>.
 18. Z. H. Deng, K. H. Luo, and H. L. Yu, "A study of supervised term weighting scheme for sentiment analysis", *Expert Systems with Applications*, Vol. 41, No. 7, 2014, pp. 3506-3513, doi: <https://doi.org/10.1016/j.eswa.2013.10.056>.
 19. M. Rathi, A. Malik, D. Varshney, R. Sharma, and S. Mendiratta, "Sentiment analysis of tweets using machine learning approach", 2018 Eleventh International Conference on Contemporary Computing (IC3), 2018, pp. 1-3, doi: <https://doi.org/10.1109/IC3.2018.8530517>.
 20. S. S. Lee, "A content analysis of journal articles using the language network analysis methods", *Journal of the Korean Society for Information Management*, Vol. 31, No. 4, 2014, pp. 49-68, doi: <https://doi.org/10.3743/KOSIM.2014.31.4.049>.
 21. M. Yoo, S. Lee, and T. Ha, "Semantic network analysis for understanding user experiences of bipolar and depressive disorders on Reddit", *Information Processing & Management*, Vol. 56, No. 4, 2019, pp. 1565-1575, doi: <https://doi.org/10.1016/j.ipm.2018.10.001>.
 22. J. Choi and Y. S. Hwang, "Patent keyword network analysis for improving technology development efficiency", *Technological Forecasting and Social Change*, Vol. 83, 2014, pp. 170-182, doi: <https://doi.org/10.1016/j.techfore.2013.07.004>.

23. B. Yoon and Y. Park, "A text-mining-based patent network: analytical tool for high-technology trend", *The Journal of High Technology Management Research*, Vol. 15, No. 1, 2004, pp. 37-50, doi: <https://doi.org/10.1016/j.hitech.2003.09.003>.
24. S. Duari and V. Bhatnagar, "Complex network based supervised keyword extractor", *Expert Systems with Applications*, Vol. 140, 2020, pp. 112876, doi: <https://doi.org/10.1016/j.eswa.2019.112876>.
25. H. J. Seo, "The issues deduction of the payment system by the text network analysis", *Journal of Payment and Settlement* Vol. 10, No. 2, 2018, pp. 37-64. Retrieved from <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART002419261>.
26. H. J. Seo, "Identifying key issues in Korea's defense policy - application of text network analysis to the 2018 defense white paper -", *Korean Journal of Military Affairs*, Vol. 6, 2019, pp. 39-70, doi: <https://doi.org/10.33528/kjma.2019.12.6.39>.
27. H. J. Lee and Y. O. Kang, "Understanding tourist's region of attraction and image of city through social network data analysis", *Journal of the Korean Urban Geographical Society*, Vol. 23, No. 1, 2020, pp. 101-114, doi: <http://dx.doi.org/10.21189/JKUGS.23.1.8>.
28. S. H. Cho, "A study on analysis of the trend of blockchain by key words network analysis", *Journal of Korea Institute of Information, Electronics, and Communication Technology*, Vol. 11, No. 5, 2018, pp. 550-555, doi: <https://doi.org/10.17661/jkiict.2018.11.5.550>.