



정책자료 2021-07

# 2021년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

정영철  
천미경



## 연구진

연구책임자 **정영철** 한국보건사회연구원 연구위원  
공동연구진 **천미경** 한국보건사회연구원 연구원

정책자료 2021-07

## 2021년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

발행일 2021년 12월  
발행인 이태수  
발행처 한국보건사회연구원  
주소 [30147]세종특별자치시 시청대로 370  
세종국책연구단지 사회정책동(1~5층)  
전화 대표전화: 044)287-8000  
홈페이지 <http://www.kihasa.re.kr>  
등록 1999년 4월 27일(제2015-000007호)  
인쇄처 고려씨엔피

---

© 한국보건사회연구원 2021  
ISBN 978-89-6827-797-9 93510  
<https://doi.org/10.23060/kihasa.d.2021.07>

## 발|간|사

스마트 사회에서 빅데이터가 넘쳐나고 있다. 이러한 빅데이터는 우리의 일상적인 삶 속에서 그 가치를 더해가고 있으며 최근에는 정부 정책분야에 있어서도 그 효용성을 인정받아 더욱 적극적으로 활용하게 되었다.

그러나 이러한 빅데이터를 활용함에 있어 초기에는 많은 비용을 지불하기도 하고 단일데이터를 사용함에 있어 제한점이 문제로 지적되기도 하였다.

이에 정부에서는 데이터 축적·유통을 활성화하고, 데이터의 유통·거래 기반을 마련하기 위해 2019년 3월 각종 데이터를 수집·분석·유통하는 ‘빅데이터 플랫폼 및 센터 구축’사업을 3개년간 추진하여 2021년 10월 현재 16개 영역별 빅데이터 플랫폼이 운영 중이다.

그러므로 이번 연도 사업에서는 영역별 빅데이터 플랫폼을 살펴봄과 더불어 연구원에서 활용 중인 소셜빅데이터 분석 플랫폼을 활용하여 보건·복지 분야에서의 주요 정책인 ‘장애인’, ‘노인’, ‘아동’, ‘건강보장’과 관련된 문서를 수집하고 다양한 분석 결과를 살펴보았다.

정책 연구 추진에 있어 이러한 자원을 활용한 분석은 기존의 조사데이터 및 행정데이터 분석과 더불어 많은 시사점을 줄 것이므로 앞으로 다양한 빅데이터 플랫폼, 소셜빅데이터 플랫폼에 대해 더욱 적극적인 활용이 필요하다 하겠다.

본 보고서에 수록된 모든 내용은 우리 연구원의 공식적인 견해가 아니며 연구에 참여한 연구진의 의견임을 밝힌다.

2021년 12월

한국보건사회연구원 원장

이 태 수



# 목 차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



Abstract .....	1
요 약 .....	3
<b>제1장 서론 .....</b>	<b>7</b>
<b>제2장 영역별 빅데이터 플랫폼 .....</b>	<b>15</b>
1. 통합데이터지도 .....	17
2. 금융 빅데이터 플랫폼 .....	20
3. 환경 비즈니스 빅데이터 플랫폼 .....	22
4. 문화 빅데이터 플랫폼 .....	23
5. 교통빅데이터 플랫폼: 국가교통데이터 오픈마켓 .....	25
6. 헬스케어 빅데이터 플랫폼: 국립암센터CONNECT .....	26
7. 유통 빅데이터 플랫폼: 한국데이터거래소 .....	28
8. 통신 빅데이터 플랫폼 .....	29
9. 지역경제 빅데이터 플랫폼: 경기지역경제포털 .....	31
10. 중소중견기업 빅데이터 플랫폼 .....	32
11. 산림빅데이터 플랫폼: 산림빅데이터 거래소 .....	33
12. 해양수산 빅데이터 플랫폼: 해양수산 빅데이터 거래소 .....	34
13. 소방안전 빅데이터 플랫폼 .....	36
14. 스마트치안 빅데이터 플랫폼 .....	37
15. 라이프로그 빅데이터 플랫폼 .....	39
16. 디지털산업혁신 빅데이터 플랫폼 .....	40

---

17. 농식품 빅데이터 플랫폼: 농식품 빅데이터 거래소 .....	42
<b>제3장 소셜빅데이터 분석 .....</b>	<b>45</b>
1. 자료원 .....	47
2. 장애인 정책관련 키워드 분석 .....	51
3. 노인 정책관련 키워드 분석 .....	60
4. 아동 정책관련 키워드 분석 .....	68
5. 건강 보장 관련 키워드 분석 .....	75
6. 보건복지데이터포털에서의 LUCY 2.0 활용 사례 .....	90
7. 마치며 .....	92
<b>참고문헌 .....</b>	<b>93</b>

# 표 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



〈표 1-1〉 소셜 미디어의 특성 .....	12
〈표 2-1〉 빅데이터 플랫폼 및 센터 현황(2021년 10월 기준) .....	18
〈표 2-2〉 통합데이터지도 구성 메뉴 .....	20
〈표 2-3〉 금융 빅데이터 플랫폼 구성 메뉴 .....	21
〈표 2-4〉 환경 비즈니스 빅데이터 플랫폼 구성 메뉴 .....	23
〈표 2-5〉 문화 빅데이터 플랫폼 구성 메뉴 .....	24
〈표 2-6〉 국가교통데이터 오픈마켓 구성 메뉴 .....	26
〈표 2-7〉 국립암센터CONNECT 구성 메뉴 .....	27
〈표 2-8〉 한국데이터거래소 구성 메뉴 .....	29
〈표 2-9〉 통신 빅데이터 플랫폼 구성 메뉴 .....	30
〈표 2-10〉 경기지역경제포털 구성 메뉴 .....	31
〈표 2-11〉 중소중견기업 빅데이터 플랫폼 구성 메뉴 .....	33
〈표 2-12〉 산림 빅데이터거래소 구성 메뉴 .....	34
〈표 2-13〉 해양수산 빅데이터거래소 구성 메뉴 .....	35
〈표 2-14〉 소방안전 빅데이터 플랫폼 구성 메뉴 .....	37
〈표 2-15〉 스마트치안 빅데이터 플랫폼 구성 메뉴 .....	39
〈표 2-16〉 라이프로그 빅데이터 플랫폼 구성 메뉴 .....	40
〈표 2-17〉 디지털산업혁신 빅데이터 플랫폼 구성 메뉴 .....	42
〈표 2-18〉 농식품 빅데이터 거래소 구성 메뉴 .....	43
〈표 3-1〉 LUCY 2.0 수집 채널 개요(2021년 12월 9일 기준) .....	48
〈표 3-2〉 장애인 정책과 관련하여 LUCY 2.0에서 설정된 최종 검색어 .....	52
〈표 3-3〉 2021년 4월 20일 '장애인의 날' 뉴스 예 .....	54
〈표 3-4〉 2021년 월별 '장애인 정책' 정보량 상위 20개 연관 키워드 .....	57
〈표 3-5〉 노인 정책과 관련하여 LUCY 2.0에서 설정된 최종 검색어 .....	60
〈표 3-6〉 2021년 6월 사회적거리두기 완화와 노인학대 관련 뉴스 예 .....	62
〈표 3-7〉 2021년 월별 '노인 정책' 정보량 상위 20개 연관 키워드 .....	65
〈표 3-8〉 아동 정책과 관련하여 LUCY 2.0에서 설정된 최종 검색어 .....	68

---

〈표 3-9〉 2021년 1월 아동학대 관련 뉴스 예 .....	70
〈표 3-10〉 2021년 월별 '아동 정책' 정보량 상위 20개 연관 키워드 .....	73
〈표 3-11〉 건강 보장과 관련하여 LUCY 2.0에서 설정된 최종 검색어 .....	75
〈표 3-12〉 건강보험 보장성 강화 정책 발표 기사 .....	76
〈표 3-13〉 건강보험 보장성 강화 반대 관련 기사1 .....	77
〈표 3-14〉 건강보험 보장성 강화 반대 관련 기사2 .....	77
〈표 3-15〉 건강보험 보장성 강화 관련 국정감사 이슈 기사 .....	78
〈표 3-16〉 건강보험 보장성 강화 성과 관련 기사 .....	78
〈표 3-17〉 건강보험 보장률 상승 관련 기사 .....	79
〈표 3-18〉 건강보험 보장성 강화관련 성과대회 관련 기사 .....	80
〈표 3-19〉 2017~2021년 '건강 보장' 정보량 상위 20개 연관 키워드 .....	84
〈표 3-20〉 2021년 월별 '건강 보장' 정보량 상위 20개 연관 키워드 .....	87



# 그림 목차

KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



[그림 1-1] 2016-2020 데이터산업 직접매출 시장규모 .....	10
[그림 1-2] 글로벌 빅데이터 시장의 산업 분야별 시장 규모 및 전망 .....	10
[그림 2-1] 통합데이터지도 구성도 .....	18
[그림 2-2] 금융 빅데이터 플랫폼 구성도 .....	21
[그림 2-3] 환경 비즈니스 빅데이터 플랫폼 개념도 .....	22
[그림 2-4] 문화 빅데이터 플랫폼 구성도 .....	24
[그림 2-5] 국가교통데이터 오픈마켓 개념도 .....	25
[그림 2-6] 국립암센터CONNECT 구성도 .....	27
[그림 2-7] 한국데이터거래소 데이터 보유현황 .....	28
[그림 2-8] 통신 빅데이터 플랫폼 개념도 .....	30
[그림 2-9] 중소중견기업 빅데이터 플랫폼 구성도 .....	32
[그림 2-10] 해양수산 빅데이터 거래소 구성도 .....	35
[그림 2-11] 소방안전 빅데이터 플랫폼 구성도 .....	36
[그림 2-12] 스마트치안 빅데이터 플랫폼 서비스 구성도 .....	38
[그림 2-13] 라이프로그 빅데이터 플랫폼 구성도 .....	39
[그림 2-14] 디지털산업혁신 빅데이터 활용사례 .....	41
[그림 3-1] LUCY 2.0 주요 기능 .....	50
[그림 3-2] LUCY 2.0 트렌드 검색 - 고급검색 화면 예시 .....	50
[그림 3-3] 2021년 월별 '장애인 정책' 문서량 .....	53
[그림 3-4] 2021년 채널별 '장애인 정책' 문서 비율 .....	54
[그림 3-5] 2021년 '장애인 정책' 정보량 상위 100개 연관 키워드 .....	55
[그림 3-6] 2021년 '장애인 정책' 토픽 매칭 현황 .....	58
[그림 3-7] 2021년 '장애인' 정보량 상위 30 해시태그 .....	59
[그림 3-8] 2021년 월별 '노인 정책' 문서량 .....	61
[그림 3-9] 2021년 채널별 '노인 정책' 문서 비율 .....	62
[그림 3-10] 2021년 '노인 정책' 정보량 상위 100개 연관 키워드 .....	63
[그림 3-11] 2021년 '노인 정책' 토픽 매칭 현황 .....	66

---

[그림 3-12] 2021년 '노인' 정보량 상위 30 해시태그 .....	67
[그림 3-13] 2021년 월별 '아동 정책' 문서량 .....	69
[그림 3-14] 2021년 채널별 '아동 정책' 문서 비율 .....	70
[그림 3-15] 2021년 '아동 정책' 정보량 상위 100개 연관 키워드 .....	71
[그림 3-16] 2021년 '아동 정책' 토픽 매칭 현황 .....	74
[그림 3-17] 2017년~2021년 월별 '건강 보장' 문서량 .....	81
[그림 3-18] 2021년 전체기간의 채널별 '건강 보장' 문서 비율 .....	82
[그림 3-19] 2017년~2021년 '건강보장' 정보량 상위 100개 연관 키워드 .....	83
[그림 3-20] 2017년~2021년 '건강 보장' 감성분석 .....	85
[그림 3-21] 2021년 '건강 보장' 토픽 매칭 현황 .....	88
[그림 3-22] 2021년 '건강 보장' 정보량 상위 30 해시태그 .....	89
[그림 3-23] 보건복지데이터포털 메인화면 .....	90
[그림 3-24] 보건복지데이터포털 소셜데이터 메뉴 화면 .....	91



## Abstract

### **Social Big Data trend analysis based on health and welfare issues in 2021**

Project Head: Chung, Youngchul

In the digital society, big data, which focuses on the size and value of data, is taking its place as a core resource in various parts of society, and many advantages are being emphasized in policy making.

In a smart environment, as public and private information produced in social media are exchanged, social data also becomes big data, leading to social big data.

This study briefly examines big data platforms in various fields, and analyzes the total amount of documents related to the disabled, the elderly, children, and health insurance, changes in the monthly document volume, negative/positive, etc. using a social big data analysis platform. Keywords and hashtag search results were reviewed

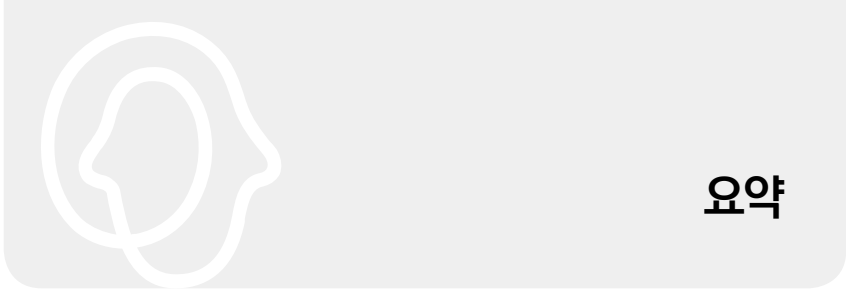
In the future, it is expected that big data analysis platforms and social big data analysis platforms that are widely available will be used more actively in policy establishment, evaluation, and policy research.

Keyword : evidence-based policy research, bigdata analysis platform, social bigdata analysis platform

---

Co-Researchers: Cheon, Mikyung





## 1. 연구의 배경 및 목적

디지털사회에서 데이터의 규모와 가치에 주안점을 둔 빅데이터는 사회 곳곳에서 핵심자원으로서 자리를 잡아가고 있으며 사회과학 분야에서도 새로운 연구 설계가 가능하여 정책 수립과 운용상 효율성을 높일 수 있는 장점이 부각되고 있다. 또한 스마트 환경에서 소셜 미디어에서 생산되는 공적, 사적 정보들이 교류됨으로써 데이터 또한 빅데이터화되어 소셜 빅데이터에 이르게 되었다.

이러한 소셜 빅데이터는 데이터 품질, 데이터의 대표성, 데이터의 신뢰성, 특히 개인의 사적정보가 포함되어 있어 개인정보침해 등에 대한 문제 등 한계를 드러내기도 하나 점차 다양한 정책분야, 근거기반 정책연구에서 기존의 정형화된 조사데이터와 행정데이터를 보완하는 수단으로 활용 가치를 더해가고 있다.

이번 연구에서는 인프라로 자리한 많은 영역에서의 빅데이터 플랫폼을 간단히 살펴보고 소셜빅데이터 분석 플랫폼을 기관 차원에서 인프라로 활용해봄으로써 보다 많은 정책연구 분야에서의 기초 데이터 분석 적용 가능성을 알아보하고자 한다.

## 2. 주요 연구결과

정부는 2019년 3월 ‘빅데이터 플랫폼 및 센터 구축’사업을 시작하여 2021년 10월 현재 금융, 환경, 문화, 교통, 헬스케어, 유통, 통신, 지역경제, 중소기업, 산림, 해양수산, 소방안전, 스마트치안, 라이프로그, 디지

#### 4 2021년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

털산업혁신, 농식품 등 16개 영역별 빅데이터 플랫폼이 운영 중이다.

소셜빅데이터 분석 플랫폼 중 하나인 LUCY2.0을 기관차원에서 활용하여 장애인 정책관련 키워드 분석을 실시한 결과, 2021년 1년동안의 수집된 관련 문서량은 총 674,818건이었으며 이 중 ‘장애인의 날’이 있는 4월에 문서량이 급증하였고 ‘2022년 대통령선거 관련하여 정당 후보자 확정’ 후 11월을 기점으로 다시 증가하였다. 채널별 문서비율은 뉴스, 블로그, 카페 순으로 나타났으며 부정적 키워드로는 ‘어렵다’, ‘장애’ 등의 키워드가, 긍정적 키워드로는 ‘노력’, ‘강화’, ‘도움’ 등이 추출되었다.

노인 정책관련 키워드 분석을 실시한 결과, 2021년 1년동안의 수집된 관련 문서량은 총 844,335건이었으며 이 중 ‘노인학대 예방의 날’인 6월 15일 영향으로 6월달 관련 문서량이 증가하였으며 11월을 기점으로 대통령 선거관련 각 정당후보자 확정 관련으로 크게 증가하였다. 부정적 키워드로는 ‘어렵다’가 추출되었고, 긍정적 키워드로는 ‘노력’과 ‘행복’이 추출되었다. ‘노인’ 키워드만으로 해시태그 검색을 해본 결과, ‘#어르신’, ‘#치매’, ‘#노인복지’ 순으로 많이 언급되었다.

아동 정책관련 키워드 분석을 실시한 결과, 2021년 1년동안의 수집된 관련 문서량은 총 1,151,297건이었으며 이 중 부모 학대로 생후 16개월에 사망한 ‘정인이 사건’과 ‘3살 여자아이가 내복차림으로 길거리에서 발견된 사건’ 영향으로 1월에 가장 많은 문서량이 추출되었다. 부정적 키워드로는 ‘아동학대’, ‘학대’, ‘어렵다’, ‘못한다’가 추출되었고 긍정적 키워드로는 ‘노력’과 ‘강화’가 추출되었다.

건강 보장 관련 키워드 분석을 실시한 결과, 2017년 1월부터 12월까지 수집된 건강 보장 키워드 문서량은 122,496건, 2018년은 112,411건, 2019년 106,220건, 2020년은 86,760건, 마지막으로 2021년은 95,515건이다. ‘건강보험 보장성 강화 정책 발표’에 따른 2017년 8월,

대규모 반대 집회와 관련하여 2017년 12월, 대한의사협회에서의 대규모 시위와 관련하여 2018년 4월, 건강보험 보장성강화 관련 국정감사와 관련하여 2018년 10월, 건강보험 보장성 강화 성과관련하여 2019년 7월, 건강보험 보장률 상승관련하여 2019년 12월, 건강보험 보장성 강화대책 4주년 성과보고대회 관련하여 2021년 8월에 각각 문서량이 증가하였다. 2021년 해시태그 검색결과, ‘#건강보장’, ‘#건강보험’, ‘#문재인’, ‘#국민건강보험’ 등이 많이 언급되었다.

### 3. 결론 및 시사점

사실상 빅데이터 분석에서는 나타난 결과에 대한 정확한 해석과 판단이 중요하다. 서론에서도 언급하였듯이 유용하지 않은 정보를 걸러내고 전체 대상에 대해 연관성을 적용하지 않고 의미있는 관련성을 골라내기 위해 많은 노력이 필요하다. 이를 위해서는 정책 영역에 대한 보다 치밀한 검토를 통해 그 연관성을 찾아내야 할 것이다.

이번 연구에서는 상세한 정책 영역에 대해 깊이 있는 검토를 적용하지는 않은 상태로 다만 주변에 많은 인프라로 널려있는 빅데이터 분석플랫폼, 소셜빅데이터 분석플랫폼에 대한 활용가능성을 제시한 것으로 앞으로 이러한 많은 플랫폼을 좀 더 적극적으로 활용하기를 기대한다.

주요 용어 : 근거기반 정책 연구, 빅데이터 분석 플랫폼, 소셜빅데이터 분석 플랫폼





사람을  
생각하는  
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



# 제 1 장

## 서론



# 제 1 장 서론

디지털사회에서 생산되는 데이터 규모는 날로 증가하고 있고 이에 따라 규모와 가치에 주안점을 둔 빅데이터에 대한 관심은 더욱 커지고 있다.

‘빅데이터’라는 용어가 하드웨어와 소프트웨어 기술 발달에 따라 1997년 처음 등장한 이후 데이터의 역할과 비중은 크게 바뀌었다. 빅데이터를 자료의 규모와 구성, 그리고 분석근거 측면에서 설명해보았을 때, 우선 자료 규모 측면에서는 엄청난 양적 변화를 나타낸다. 자료 구성 측면에서는 자료가 일관적이거나 체계적이지 않은 특성을 지니고 있으며, 분석근거 측면에서는 기존의 인과관계에서 상관관계로 변화함을 나타내고 있다. 이러한 빅데이터는 기존의 전통적 자료 수집과 분석 방식을 대체하기 보다는 상호보완할 때 효용이 나타난다(한신갑, 2015).

빅데이터는 사회 곳곳에서 핵심자원으로서 자리를 잡아가고 있으며 산업측면에서도 데이터를 처리해주는 빅데이터 수집, 분석 시장이 눈에 띄게 성장하고 있다.

이에 힘입어 우리나라의 ‘20년도 데이터산업 시장규모는 20조원에 이르렀으며 최근 5년간 연평균 성장률 9.8%에 비해 최근 3년간 연평균 성장률이 13.3%를 보이는 등 계속 높은 성장률을 보이고 있다([그림 1-1] 참조).

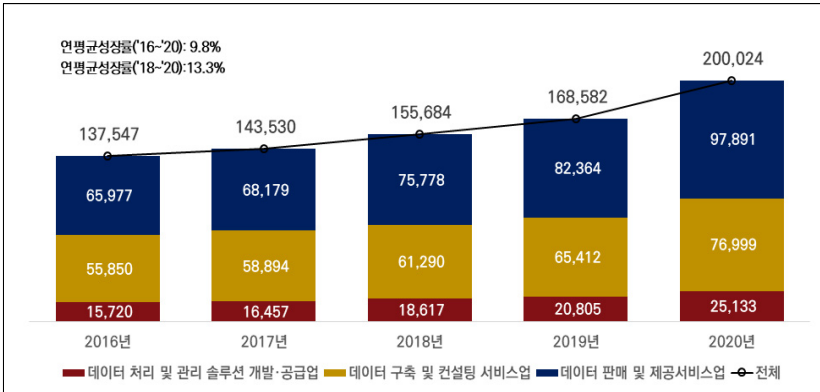
한편, 전 세계 빅데이터 시장은 2020년 1,388억 8,600만 달러에서 연평균 성장률 10.6%로 증가하여, 2025년에는 2,294억 2,300만 달러에 이를 것으로 전망되고 있으며 특히 ‘헬스케어 및 생명과학’ 산업분야는

10 2021년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

연평균 성장률이 12.5%에 이르고 있다([그림 1-2] 참조).

[그림 1-1] 2016-2020 데이터산업 직접매출 시장규모

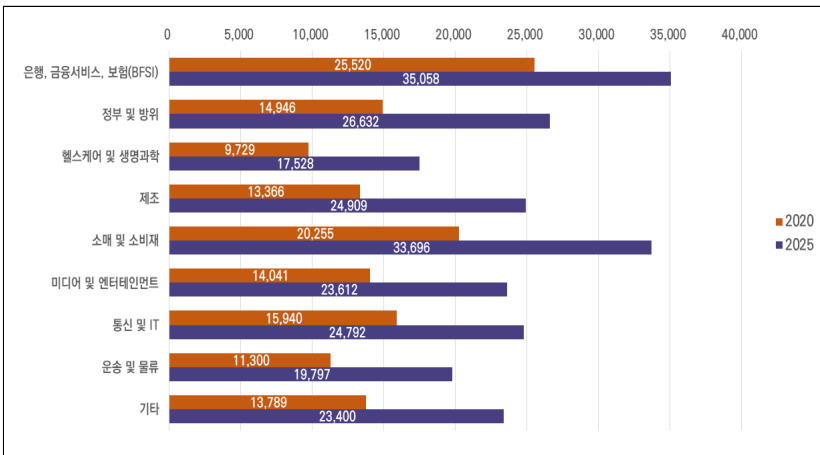
(단위: 억원)



자료: 과학기술정보통신부 보도자료. (2021.12.30.). 데이터 댐이 이끈 데이터산업 시장, 2020년 2019년 대비 18.9% 성장한 20조원 달성. p.3.

[그림 1-2] 글로벌 빅데이터 시장의 산업 분야별 시장 규모 및 전망

(단위: 백만 달러)



출처: MarketsandMarkets. Big Data Market. 2020.

자료: 연구개발특구진흥재단. (2021). 빅데이터 시장. 글로벌 시장동향보고서. p.9.

이러한 빅데이터를 이용할 경우 사회과학 분야에서는 이전에 수행하기 어려웠던 인과관계 분석을 위한 실험 등 새로운 연구 설계가 가능하여 정책 수립과 운용상 효율성을 높일 수 있는 장점이 부각되고 있다. 그 외에도 빅데이터는 전통적인 데이터를 완전히 대체할 수는 없으나 기존의 정형 데이터가 전달할 수 있는 정보량과 범위의 한계를 극복하여 정책분석에서의 장점을 다음과 같이 제시하고 있다(Stephens-Davidowitz, 2017; 유경원, 2020. 재인용). 첫째, 기존 정형화된 데이터와 달리 이미지, 텍스트 등 새로운 유형의 데이터를 제공함으로써 설명하고자 하는 현상에 대해 추가적인 정보를 제공한다. 둘째, 이용자의 솔직한 내용이 그대로 제공됨으로써 조사자료에서 발생할 수 있는 응답편의가 줄어든다. 셋째, 작은 집단도 확대해 볼 수 있는 ‘디지털 확대’ 기능이 작동되어 정책의 타겟그룹 설정 등에 있어 용이하다. 넷째, 큰 비용이나 복잡함 없이도 인과관계 파악을 위한 실험 시행이 가능하여 보다 정확한 정책의 효과 분석이 가능하다.

그러나 한신갑(2015)은 빅데이터 분석에서의 유의점에 대해 다음과 같은 네가지를 언급하기도 하였다. 첫째, 빅데이터에서 자료의 양과 질은 비례하지 않으며, 자료 중에서도 유용하지 않은 정보(noise)를 제거하여 유용한 정보(signal)를 거르는 것이 중요하다고 하였다. 둘째, 빅데이터는 연구대상에 대한 대표성을 보장하기보다는 특정 영역에 국한됨을 이해하고 있어야 한다고 하였다. 셋째, 자료의 규모가 커질수록 의사(疑似) 상관관계 또한 커지므로 이 중에서 의미있는 관련성을 골라내는 것이 중요하다고 하였다. 넷째, 빅데이터가 산출되고 다루는 영역은 기술적 가능성, 상업적 필요성 등에 의해 사실상 편향되어 있으므로 모든 영역에서 빅데이터를 통한 해석이 가능하지는 않다고 하였다.

인터넷이 활성화되면서 기존 오프라인이 아닌 온라인상에서 커뮤니티

와 네트워크를 형성하는 서비스를 소셜 네트워크 서비스(Social Network Service)라 하였으며, 이러한 서비스를 제공하는 사이트는 소셜 네트워크 사이트(Social Network Sites)로 불리게 되었다(최민재, 2009). 즉 SNS라 함은 소셜네트워크 서비스, 소셜네트워크사이트, 소셜 미디어, 소셜네트워킹 등의 다양한 용어와 혼재되어 사용되며 업계에서도 학계에서도 통일된 정의를 내리지 못하고 있다(이윤희, 2014).

이러한 온라인상의 네트워크 개념에서 출발한 SNS는 점차 콘텐츠 생산과 공유 개념이 강화된 소셜 미디어(Social media) 개념으로 확장되면서 미디어적 측면이 부각되었으며(〈표 1-1〉 참조) 초창기 마케팅 영역을 중심으로 주목을 받게 되었다(최민재, 2009).

〈표 1-1〉 소셜 미디어의 특성

구분	내용
참여(Participation)	소셜미디어는 관심있는 모든 사람들의 기여와 피드백을 촉진하며, 미디어와 오디언스의 개념을 불명확하게 함
공개(Openness)	대부분의 소셜 미디어는 피드백과 참여가 공개되어 있으며 투표, 피드백, 코멘트, 정보 공유를 촉진함으로써 콘텐츠 접근과 사용에 대한 장벽이 거의 없음
대화(Conversation)	전통적인 미디어가 'broadcast'이고 콘텐츠가 일방적으로 오디언스에게 유통되는 반면, 소셜 미디어는 쌍방향성을 띠
커뮤니티(Community)	소셜미디어는 빠르게 커뮤니티를 구성케 하고 커뮤니티로 하여금 공통의 관심사에 대해 이야기함
연결(Connectedness)	대부분의 소셜미디어는 다양한 미디어의 조합이나 링크를 통한 연결상에서 변성함

자료: FkII 조사연구팀. (2006.12.). 소셜 미디어(Social Media)란 무엇인가?. 『정보산업지』, p.53.; 최민재. (2009). 소셜 미디어의 확산과 미디어 콘텐츠에 대한 수용자 인식연구. 한국 언론정보학회 학술대회. pp.5~31. 재인용.

인터넷 이용이 일반화되고 특히 무선환경이 확산되면서 SNS 분야는 급속히 성장하였으며 이에 따라 소셜 미디어에서 생산되는 공적, 사적 정

보들이 교류됨으로써 데이터 또한 빅데이터화되어 소셜 빅데이터에 이르게 되었다. 소셜빅데이터란 소셜미디어와 빅데이터를 결합한 용어로 뉴스, 블로그, 트위터 등과 같은 소셜미디어에서 나오는 방대한 양의 데이터를 뜻하며 빅데이터 특징에 소셜 미디어에서 생산되는 정보 특성을 더하고 있다.

빅데이터 중에서도 이러한 소셜 미디어에서의 데이터, 즉 소셜 빅데이터는 디지털 기술 발달과 스마트 디바이스의 확대보급에 따라 시간과 장소에 무관하게 개인 혹은 조직에서 정형 데이터 뿐 아니라 비정형 데이터까지 보다 빠르게 축적되고 확산되어 가고 있다(윤유동 외, 2017).

소셜 빅데이터 분석은 초창기 민간에서 마케팅 전략을 수립하기 위한 목적으로 활발하게 적용되어가면서 점차 다양한 정책분야, 근거기반 정책연구에서도 활용가치를 더해가고 있다. 그러나 행정 빅데이터와 달리 소셜 빅데이터는 데이터 생산과정에서 부산물로 얻어지다 보니 데이터 품질, 데이터의 대표성, 데이터의 신뢰성, 특히 개인의 사적정보가 포함되어 있어 개인정보침해 등에 대한 문제 등 한계를 드러내기도 한다. 그럼에도 불구하고 기존의 정형화된 조사데이터와 행정데이터를 보완하는 수단으로 활용가치를 더해가고 있다.

이제 소셜 빅데이터 분석은 기존의 정형화된 데이터를 기반으로 한 공공영역의 조사데이터나 행정데이터, 그리고 민간영역의 다양한 데이터들과 함께 정책연구에서 중요한 요소로 자리잡아 가고 있다.

한편 2018년 6월 데이터 산업 활성화 전략(4차산업혁명위원회)에서 빅데이터 유통 플랫폼 논의가 제기되어 정책적으로 공공, 민간의 데이터를 거래할 수 있는 민관협동 데이터 거래시장 형성 전략이 제시되었다. 이에 의해 2019년 3월 '빅데이터 플랫폼 및 센터 구축'에 관한 1차 공모 이후 금융, 환경, 문화, 교통, 헬스케어, 유통 소비, 통신, 중소기업, 지역

경제, 산림 등 10대 분야에 대해 '빅데이터 플랫폼'을 개설하여 이용자들에게 유·무상으로 데이터를 제공하는 사업을 추진하고 있다(국회입법조사처, 2020). 이후 2020년 농식품, 해양수산, 소방안전 등 6개 플랫폼이 추가 선정되어 2021년 10월 기준 16개 분야 빅데이터 플랫폼과 총 180여 개 센터가 운영 중이다.

이러한 빅데이터는 모으는 것도 중요하지만 가치를 추출하기 위해 데이터를 분석하는, 즉 마이닝하는 작업은 복잡하면서도 중요한 작업이다. 전 국민의 일상생활 데이터를 가장 많이 생산하고 있는 보건복지분야에서도 일상생활에서 겪는 다양한 문제해결을 위해 유효한 데이터 분석이 많이 활용되는 분야이기도 한 만큼 빅데이터 분석, 소셜 빅데이터 분석에 대해 당 연구원에서는 몇 년간 비정형 빅데이터 분석방법론을 중심으로 살펴보았다.

그러나 이러한 분석방법론을 살펴보는 것도 중요하지만 이제는 인프라로 자리한 많은 영역에서의 빅데이터 플랫폼에서 제공하는 다양한 빅데이터를 수집하고 분석하여 정책에 제대로 활용하는 것도 바람직하다. 그러므로 이번 사업에서는 영역별 빅데이터 플랫폼을 간단히 살펴보고 또한 소셜빅데이터 분석 플랫폼을 기관 차원에서 인프라로 활용해봄으로써 앞으로 다양한 영역에서, 보다 많은 정책연구 분야에서의 기초 데이터 분석 적용 가능성을 알아보고자 한다.



사람을  
생각하는  
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS



## 제2장

### 영역별 빅데이터 플랫폼



## 제 2장    영역별 빅데이터 플랫폼

빅데이터에 대한 효능과 가치는 곧 빅데이터 수집과 구매에 대한 관심을 증가시키고 있다. 정부에서는 이와 같은 빅데이터에 대한 수집·유통을 원활하게 하기 위해 분야별 빅데이터 플랫폼 구축 사업을 진행하고 있는 바, 제2장에서는 과학기술정보통신부가 수행하고 있는 ‘빅데이터 플랫폼 및 센터 구축사업’의 금융, 환경, 문화, 교통, 헬스케어, 유통, 통신, 지역경제, 중소기업, 산림, 해양수산, 소방안전, 스마트치안, 라이프로그, 디지털산업혁신, 농식품 등 16개 영역의 빅데이터 플랫폼과 이들의 전체 현황을 파악할 수 있는 통합데이터지도에 대해 간략하게 정리하였다.

### 1. 통합데이터지도(<http://www.bigdata-map.kr/>)

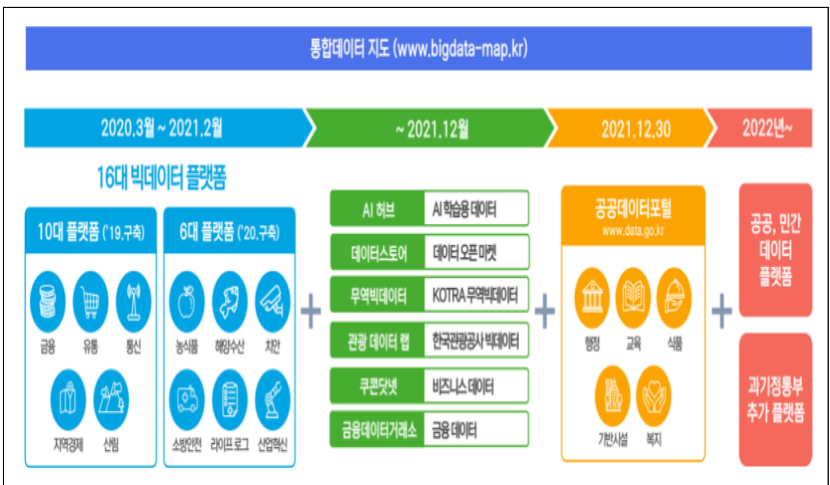
과학기술정보통신부가 디지털 뉴딜 ‘데이터댐’ 사업의 일환으로 추진한 플랫폼으로 ‘통합데이터지도’는 16개 분야별 빅데이터 플랫폼 및 센터를 구축하여 공공과 민간에서 제공하는 데이터의 수집·유통을 활성화하고 누구나 양질의 데이터에 쉽게 접근·활용할 수 있도록 지원하고 있다.

통합데이터지도는 공공과 민간에서 제공하는 데이터를 쉽게 검색·활용할 수 있도록 지원하며 ‘19년 10개 영역에 대한 플랫폼을 구축한 이후 ‘20년 6개 영역에 대한 플랫폼을 추가 구축하였으며 이는 AI Hub, 데이터스토어, KOTRA, 한국관광공사, 쿠콘 등 데이터 파트너들과 함께 하고 있다([그림 2-1] 참조). 2021년 10월 기준 총 180여개 센터가 운영 중에

있다(〈표 2-1〉 참조).

이러한 통합데이터지도는 ‘데이터 현황’, ‘데이터 분석’, ‘이슈데이터’, ‘게시판’, ‘커뮤니티’, ‘데이터지도’ 메뉴로 구성되어 있다(〈표 2-2〉 참조).

[그림 2-1] 통합데이터지도 구성도



자료: 통합데이터지도 사이트 중 ‘데이터지도 > 데이터지도 소개’ 메뉴, <http://www.bigdata-map.kr/intro>.

〈표 2-1〉 빅데이터 플랫폼 및 센터 현황(2021년 10월 기준)

분야	센터
금융	노타, 닐슨코리아, 다음소프트, 바이브컴퍼니, 여기어때컴퍼니, 이노핀, KT, 해빗팩토리, 기용정보통신, 한국감정평가사협회, 한국금융솔루션, 부산시청 (12)
환경	한국기상산업기술원, 국립생태원, 환경정책평가연구원, 한국지질자원연구원, GDS 컨설팅그룹, 그린에코스, 한국과학기술원, 에코엔파트너스, 국립중앙과학관, 강원대학교 삼척산학협력단, 노바코스 (11)
문화	국립중앙도서관, 국민체육진흥공단, 문화예술위원회, 청소년활동진흥원, 부산정보산업진흥원, 야놀자, 레드태이블, 레드타이, 데이터마케팅코리아, ㈜원투스엠, 제주관광공사, ㈜텐엔엠에스, 컨슈머인사이트, 충남정보문화산업진흥원, 한국문화원연합회 (15)

분야	센터
교통	울산정보산업진흥원, 포항테크노파크, 아이나비시스템즈, KT, 코리아크레딧뷰로, 우, 성남시, 진주시, 대전시, 서울시, 한국부동산원, 오픈메이트온 (11)
헬스케어	삼성서울병원, 연대 세브란스병원, 건양대학교병원, 길병원, 대구가톨릭대학교의료원, 전북대학교병원, 분당서울대학교병원, 화순전남대병원, 아주대학교병원, 서울대학교 산학협력단 (10)
통신	경기대학교, 넥스트이지, 네스, 두잉랩, 모토브, BC카드, 소상공인연합회, 어메이징푸드솔루션, 오픈메이트, 인천테크노파크, 제로투원파트너스, 코난테크놀로지, 한국스마트그리드사업단, 한국인터넷진흥원 (14)
유통	삼성카드, CJ올리브네트웍스, SK텔레콤, SK플래닛, 웰컴에프앤디, GS리테일, KDX한국데이터거래소, NICE디앤알, 바이브컴퍼니, 데이블, 로플렛, 빌트온, 식신, 온누리에이치엔씨, 지인플러스, 코리아크레딧뷰로, 한국우편사업진흥원, 경동도시가스, 나이스디앤비, 어반유니온 (20)
중소기업	빅밸류, 한국생산성본부, NICE평가정보, 한국무역정보통신, 와이즈넷, 한국산업기술진흥협회, 인크루트, 녹색기술센터, 선도소프트, 한국물가정보, 한국법령정보원, 한국경영인증원 (12)
지역경제	경기일자리재단, 경기신용보증재단, 한국기업데이터, 코나아이, 더아이앤씨, 국토연구원, 한양대학교 산학협력단, 한국생산기술연구원, 한국신용데이터, 펌프킨 (10)
산림	비글, 아로정보기술, 인포보스, 한국산림복지진흥원, 시선아이티, 우림엔알, 삼아항업, 에이콘컴퍼니, 마켓링크, 한국한의학연구원, 에코아이, 데이터사이언스랩 (12)
농식품	이지팜, 유로지스넷, 장보고식자재마트, 나이스지니데이터, 케이플러스, 이지아이치엘디, 한국무역통계진흥원, 전남정보문화산업진흥원, 원데이터기술 (9)
해양수산	환경과학기술, 빌리언21, 목포대학교, 광주과학기술원, 울산과학기술원, 램오투원, 에이치더블유, 해운항만물류정보협회, 한국해양대학교, 해양교통안전공단, 올시데이터 (11)
소방안전	세종시소방본부, 전북소방본부, 울산시소방본부, 제주소방안전본부, 한국소방산업기술원, 한국화재보험협회, 올라이트라이프, 업데이트, 한방유비스, 예측진단기술 (10)
스마트 치안	수원시청, 한국토지주택공사, 서울신용보증재단, 에스투더블유패, 더치트, 지란지교데이터, 엠진시큐러스, 이투온, 아이티로, 연합뉴스 (10)
라이프 로그	강원대학교산학협력단, 고려대의료원, 굿닥, 대한청각학회, 베이글랩스, 아이센스, 케이웨더, 한림대병원, 헬스맥스, 휴레이포지티브, 헬스브릿지, LG유플러스 (12)
디지털 산업혁신	한국기업데이터, 한국산업지능화협회, 한국생산성본부, 한국인사이트연구소, 애플엔가이드, 가이온, 한국M&A거래소, 대덕넷, 알리콘, 에스티에이치아이에스, 이앤씨지엘에스, 헤브론스타, 광개토연구소, 벅스인텔리전스 (14)

자료: 통합데이터지도 홈페이지 중 '데이터지도' 메뉴. <https://www.bigdata-map.kr/intro>

〈표 2-2〉 통합데이터지도 구성 메뉴

데이터현황	데이터분석	데이터 스토리	이슈데이터	게시판	커뮤니티	데이터지도
-전체현황 -데이터분류별 -지역별 -키워드별 -플랫폼분야별	- 데이터 시각화 - 데이터 프로파일링	- 데이터 스토리	-이슈?있슈! -코로나19 데이터	-공지사항 -혁신서비스사례 -이벤트 -FAQ -Q&A -데이터114	-오픈커뮤니티 -데이터활동커뮤니티	-데이터지도소개 -데이터등록

자료: 통합데이터지도. <http://www.bigdata-map.kr/>.

## 2. 금융 빅데이터 플랫폼(<https://www.bigdata-finance.kr/>)

금융 빅데이터 플랫폼은 11개의 빅데이터 기업이 생산하는 데이터를 수집하여 유통하는 빅데이터 거래소로써, 카드소비(BC카드), 보험 보장·가입 분석(해빗팩토리), 증권·투자 지표(SBCN, 한국금융솔루션) 등 금융 데이터부터 유통인구(KT), 부동산·아파트(기용정보통신, 한국감정평가사협회), 소셜데이터(바이브컴퍼니, 구 다음소프트), 맛집 데이터(망고플레이트), 소비자 구매 트렌드 분석 데이터(닐슨컴퍼니 코리아) 등 금융데이터와 결합해 시너지 창출이 가능한 주요 데이터를 한 곳에 담고 있다. 또한 융합데이터 기반 특화서비스를 개발·제공하고, 데이터 분석 환경을 제공하며, 데이터 분석교육을 진행하는 데이터 플랫폼이다(〈그림 2-2〉 참조).

데이터제공기관으로는 BC카드, KT, NOTA, 한국신용정보원, 여기어때컴퍼니, 닐슨코리아, 티머니, 부산광역시, 한국감정사평가협회, 한국금융솔루션 등이며 데이터 제공영역은 분석Report, 컨설팅, 융합/결합, 카드, 맛집, 소셜, 유통인구, 보험, 부동산, 주거, 대출 등으로 구분하고 있다. 금융 빅데이터 플랫폼은 ‘데이터상품’, ‘데이터활용’, ‘플랫폼서비스’, ‘플랫폼소개’, ‘이용안내’ 메뉴로 구성되어 있다(〈표 2-3〉 참조).

[그림 2-2] 금융 빅데이터 플랫폼 구성도



자료: 금융 빅데이터 플랫폼 중 ‘플랫폼소개’ 메뉴.

<https://www.bigdata-finance.kr/platform/introPlatform.do>

<표 2-3> 금융 빅데이터 플랫폼 구성 메뉴

데이터 상품	데이터 활용	플랫폼서비스	플랫폼소개	이용안내
-전체데이터 -무료데이터 -유료데이터 -맞춤데이터 -추천데이터 -OPEN API	-GIS분석 -분석소프트웨어 - 분석 / 시각 화 Report - 금융 빅 데이터 플랫폼 Lab -빅데이터공모전	-융합서비스 -Biz Credit서비 스 -트렌드 키워드 분석	-금융 빅 데이터 플랫폼 소개 -제공기관안내 -데이터활용 서비스안내	-공지사항 -FAQ -Q&A -DB META INFO -OPEN API안내 -금융빅데이터 플 랫폼 Lab 예약하기 -데이터요청

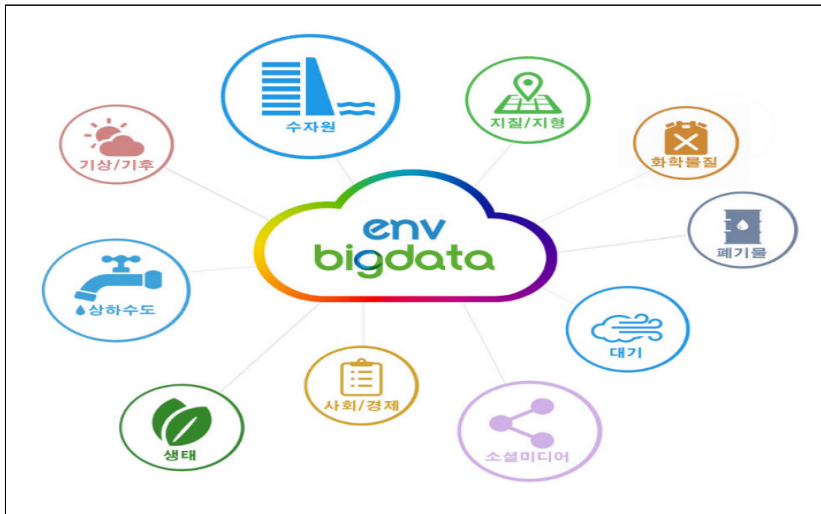
자료: 금융빅데이터플랫폼. <https://www.bigdata-finance.kr/main.do>

### 3. 환경 비즈니스 빅데이터 플랫폼 (<https://www.bigdata-environment.kr/>)

환경 비즈니스 빅데이터 플랫폼은 한국수자원공사가 수행하는 사업으로 환경분야(8개)와 사회경제분야(2개)의 총 601종 전문데이터를 축적·유통하여 데이터기반 환경 비즈니스를 육성하고 새로운 일자리 창출로 사회적 가치 실현을 목표로 하는 데이터마켓 플랫폼이다(그림 2-3) 참조).

데이터 제공기관으로는 한국수자원공사, 국립생태원, 한국환경연구원, 한국과학기술원, 노바코스, 강원대학교 등 11개이며 영역은 상하수도, 수자원, 기상/기후, 대기, 생태, 지질/지형, 사회경제, 소셜미디어, 폐기물 등 총 10개 영역으로 구분하고 있다.

[그림 2-3] 환경 비즈니스 빅데이터 플랫폼 개념도



자료: 박지성. (2021.4.20.). 한국수자원공사 '환경 빅데이터 플랫폼'. 전자신문.  
<https://m.etnews.com/20210414000050>.



환경 비즈니스빅데이터 플랫폼은 ‘환경데이터마켓’, ‘데이터서비스’, ‘데이터분석’, ‘지원프로그램’, ‘플랫폼소개’ 메뉴로 구성되어 있다(〈표 2-4〉 참조).

〈표 2-4〉 환경 비즈니스 빅데이터 플랫폼 구성 메뉴

환경데이터 마켓	데이터 서비스	데이터 분석	지원프로그램	플랫폼 소개
-데이터검색 -데이터맵 -타플랫폼 데이터 검색 -데이터유통거래안내 -데이터보유 현황 -데이터수요 조사 -데이터패키지 상품	-서비스소개 -지하수를 찾아 줘 -도시침수예측 서비스 -환경법령제공 서비스 -산책하기좋은날 -내 손안의 물 -수상레저 활동 지수 -언론기반 환경 이슈 분석 -환경소비 트렌드분석	-분석서비스 도구 -데이터시각화	-교육계획 -창업&연구 컨설팅 -학술대회특별 세션 -공모전 -신규센터모집 신청 -기업홍보용팝업 신청	-플랫폼소개 -플랫폼활용사례 -공지사항 -Q&A -FAQ

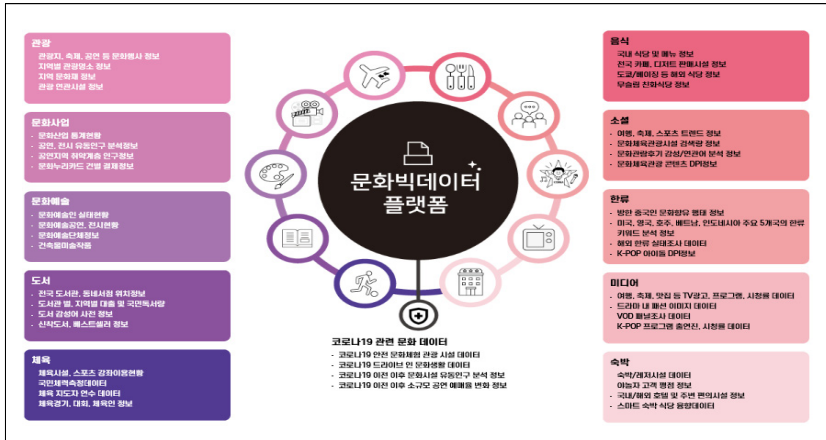
자료: 환경 비즈니스빅데이터 플랫폼. <https://www.bigdata-environment.kr/>.

#### 4. 문화 빅데이터 플랫폼(<https://www.culture.go.kr/bigdata/>)

공공기관 및 민간기업으로 구성된 데이터센터와 함께 도서, 체육, 예술, 숙박, 레저, 음식 등 문화체육관광분야 데이터를 개방하고 데이터 유통거래 생태계를 조성하는 플랫폼으로 수행기관은 한국문화정보원이다(〈그림 2-4〉 참조). 데이터는 문화예술, 문화유산, 문화산업, 관광, 체육, 도서관, 여건조성, 홍보지원 영역으로 구분하고 있으며 데이터 제공기관은 제주관광공사, 국립중앙도서관, 한국문화정보원, 국민체육진흥공단, 한국문화원연합회, 데이터마케팅코리아, 한국청소년활동진흥원, 야놀자 등이 있다.

24 2021년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

[그림 2-4] 문화 빅데이터 플랫폼 구성도



자료: 문화 빅데이터 플랫폼 중 '플랫폼 이용안내' 메뉴.  
[https://www.culture.go.kr/bigdata/user/data\\_market/agency/list.do](https://www.culture.go.kr/bigdata/user/data_market/agency/list.do).

문화 빅데이터 플랫폼은 '데이터 상품', '데이터 상담소', '데이터 분석', '문화서비스', '데이터활용', '플랫폼 이용안내' 메뉴로 구성되어 있다(〈표 2-5〉 참조).

〈표 2-5〉 문화 빅데이터 플랫폼 구성 메뉴

데이터상품	데이터상담소	데이터분석	문화서비스	데이터활용	플랫폼 이용안내
-전체상품 -데이터리포트	-상담소 소개	-시각화하기 -시각화공유 -머신러닝 -데이터 시각화	-문화시설 방문분석 -우리동네 책방 -문화역세권 -문화트렌드 -지역문화 소비분석 -K-Food -durlcncjs	-활용사례 -공모전	-플랫폼소개 -공지사항 -Q&A -FAQ

자료: 문화 빅데이터 플랫폼. <https://www.culture.go.kr/bigdata/>.

## 5. 교통빅데이터 플랫폼: 국가교통데이터 오픈마켓 (<https://www.bigdata-transportation.kr/>)

국가교통데이터 오픈마켓은 한국도로공사에서 운영하고 있으며 교통 관련데이터를 기반으로 새로운 교통서비스를 발굴하기 위한 플랫폼으로 ([그림 2-5] 참조) 데이터는 도로, 버스, 철도, 시설안전, 자율주행, 공간, 융합, 일반기타 영역으로 구분하고 있으며 데이터 제공기관은 한국도로공사, 한국교통연구원, 한국철도공사, 포항테크노파크, 진주시청, 성남시청, KT, SK텔레콤, 코리아크레딧뷰로 등이 있다.

[그림 2-5] 국가교통데이터 오픈마켓 개념도



자료: 국가교통데이터 오픈마켓 홈페이지 중 '이용안내' 메뉴.  
<https://www.bigdata-transportation.kr/howto/index>.

국가교통데이터 오픈마켓은 '데이터 상품', '데이터 활용', '공모전', '커뮤니티', '이용안내' 메뉴로 구성되어 있다(<표 2-6 참조).

〈표 2-6〉 국가교통데이터 오픈마켓 구성 메뉴

데이터상품	데이터활용	공모전	커뮤니티	이용안내
-데이터검색 -지도기반검색	-데이터활용사례 -데이터분석 툴 -모빌리티 리빙랩	-공모전개요 -공모전현황	-전문가포럼 -기업홍보 -찾아가는 설명회	-데이터거래 안내 -국토교통결합 전문기관 -고객센터

자료: 국가교통데이터 오픈마켓. <https://www.bigdata-transportation.kr/>.

## 6. 헬스케어 빅데이터 플랫폼: 국립암센터CONNECT (<https://www.bigdata-cancer.kr/ncc/main.do>)

헬스케어 빅데이터 플랫폼인 국립암센터CONNECT는 국립암센터에서 운영하고 있으며 암 빅데이터를 상호 공유하고 호환성을 강화하여 아이디어·지식·자원을 공유하는 개방형 플랫폼이다(〔그림 2-6〕 참조).

제공하는 데이터는 폐쇄환경에서 IRB 승인받은 연구자에게 제공하는 ‘폐쇄형 데이터’, 연구자를 대상으로 검색조건별 메타집계 정보를 조회해 볼 수 있는 ‘공유형 데이터’, 그리고 일반 국민을 대상으로 메타포털을 통해 활용할 수 있는 ‘개방형 데이터’로 구분할 수 있다.

제공하는 암종류로는 유방암, 갑상선암, 난소암, 폐암, 대장암, 신장암, 위암, 간암, 전립선암, 췌장암, 담도암 등이다.

국립암센터CONNECT는 ‘플랫폼소개’, ‘데이터 시각화’, ‘데이터 조회’, ‘데이터 신청’, ‘응용서비스’, ‘미디어홍보’, ‘이용안내’ 메뉴로 구성되어 있다(〈표 2-7〉 참조).

[그림 2-6] 국립암센터CONNECT 구성도



자료: 국립암센터CONNECT 홈페이지 중 '플랫폼 소개' 메뉴.  
<https://www.bigdata-cancer.kr/ncc/summary.do>

<표 2-7> 국립암센터CONNECT 구성 메뉴

플랫폼소개	데이터 시각화	데이터조회	데이터신청	응용서비스	미디어홍보	이용안내
-CONNE CT란? -플랫폼 구성도 -포털기능 -데이터 소개	-데이터 대시보드 -데이터 상세 그래프 -콘텐츠 정보	-테이블 관계도 -테이블 현황 -컬럼조회 -코드조회 -용어사전 -타플랫폼 데이터 조회	-헬프 데스크 서비스 -암라이브 러리 데이터 신청 -오픈API 신청 -DATA FREE BOX 서비스	-간암환자 병기별 치료정보 -헬스케어 알고리즘 분석허브 -유방암 환자재발 예측 -폐암환자 사망예측	-미디어 홍보	-공지사항 -FAQ -Q&A

자료: 국립암센터CONNECT. <https://www.bigdata-cancer.kr/ncc/main.do>

## 7. 유통 빅데이터 플랫폼: 한국데이터거래소(https://kdx.kr/)

한국데이터거래소 플랫폼은 민간데이터 거래소로 MBN(매일방송)이 주관하고 있으며, 데이터 공급자와 수요자 사이의 거래를 중개하고 AI학습 데이터를 생산 및 유통하고 있다([그림 2-7] 참조).

데이터는 경제/산업, 금융/증권, 통신/인구, 소비/상권, 이커머스, 유통/마케팅, 물류/교통, 보건의료, 부동산/지리, 자동차 영역으로 구분하고 있으며 데이터 제공기관은 MBN, KDX한국데이터거래소, CJ올리브 네트워크, KB캐피탈, MarketLink, SK텔레콤, 경동도시가스, 나이스디엔비, 나이스디엔알 등이 있다.

한국데이터거래소는 ‘데이터 상품’, ‘인공지능’, ‘플랫폼 서비스’, ‘데이터루’, ‘이용안내’, ‘KDX소개’로 구성되어 있다(<표 2-8> 참조).

[그림 2-7] 한국데이터거래소 데이터 보유현황

<p><b>금융/부동산</b></p> <ul style="list-style-type: none"> <li>☑ 웰컴에프앤디 대표 데이터 대부업 대출 분포 정보 데이터</li> <li>☑ 지인플러스 대표 데이터 아파트 매수가능 잠재수요 분석 데이터</li> <li>☑ 코리아크레딧뷰로 대표 데이터 소비 정보 데이터</li> </ul>	<p><b>소비</b></p> <ul style="list-style-type: none"> <li>☑ 삼성카드 대표 데이터 고객 유형별 소비 형태 데이터</li> <li>☑ 빌트온 대표 데이터 온라인마켓카테고리별 베스트 100 가격노출순위</li> <li>☑ 영코퍼레이션 대표 데이터 온라인 상품 구매 데이터</li> <li>☑ 신한카드 대표 데이터 카드소비(X이동통신) 데이터</li> </ul>	<p><b>언어/동영상</b></p> <ul style="list-style-type: none"> <li>☑ MBN 대표 데이터 경제뉴스 데이터</li> <li>☑ KDX 대표 데이터 MBN 뉴스 분석(NLP)</li> <li>☑ 바이브컴퍼니 대표 데이터 이슈 키워드</li> <li>☑ 데이블 대표 데이터 미디어 방문자 추정 데이터</li> </ul>
<p><b>유통/물류</b></p> <ul style="list-style-type: none"> <li>☑ CJ올리브네트웍스 대표 데이터 택배 송장 유통 인덱스 데이터</li> <li>☑ GS리테일 대표 데이터 특성별 매출 구성비 데이터</li> <li>☑ 한국우편사업진흥원 대표 데이터 우편 데이터</li> </ul>	<p><b>유통인구</b></p> <ul style="list-style-type: none"> <li>☑ SK텔레콤 대표 데이터 월별 서비스인구 데이터</li> <li>☑ 로플렛 대표 데이터 코로나19 여파 유통인구 데이터</li> <li>☑ TDI 대표 데이터 상권 유통인구 분석 자료 데이터</li> </ul>	<p><b>기타</b></p> <ul style="list-style-type: none"> <li>☑ SK플래닛 대표 데이터 상권 순위 데이터</li> <li>☑ 나이스디엔알 대표 데이터 코로나19 전후 앱 이용량 변화 데이터</li> <li>☑ 식신 대표 데이터 분류별 점심 식대 금액 통계 데이터</li> <li>☑ 온누리H&amp;C 대표 데이터 온누리스토어(온라인몰) 판매 내역 데이터</li> </ul>

자료: 한국데이터거래소 홈페이지 중 ‘커뮤니티’ 메뉴. <https://kdx.kr/introduce>.

〈표 2-8〉 한국데이터거래소 구성 메뉴

데이터상품	인공지능	플랫폼서비스	데이터루	이용안내	KDX소개
-추천데이터 -전체데이터 -무료데이터 -유료데이터 -맞춤데이터 -데이터 리포트	-KDX-AI -AI데이터 상품 -MBN동영상 -AI학습용 데이터 -KDX저작 도구 -KDX-AIVA	-데이터 솔루션 -데이터 인사이트	-데이터루 -데이터 스토리	-회원가입 안내 -이용약관 -공지사항 -FAQ -문의하기	-KDX소개 -KDX소식 -참여기업

자료: 한국데이터거래소 홈페이지. <https://kdx.kr/>.

## 8. 통신 빅데이터 플랫폼 (<https://www.bigdata-telecom.kr/>)

통신 빅데이터 플랫폼은 KT와 16개 센터에서 생산된 데이터가 수집되고, 클라우드와 인공지능 기반으로 분석된 데이터가 유통되는 거래 플랫폼으로 유동인구와 소비, 상권 등 기업의 디지털 마케팅 활동에 필수인 데이터와 함께 기업 수요에 따라 맞춤형 분석 리포트와 컨설팅 등을 제공하고 있다(그림 2-8) 참조).

데이터는 지역/위치, 생활, 공공, 사업·업무, 사회, 사람, 세대, 기업, 시간, 공간 등 영역으로 구분하고 있으며 데이터 제공기관은 한국인터넷진흥원, (주)네스트이지, 한국스마트그리드, 두잉랩, 롯데카드, (주)케이티, 한국궤협조사연구소, (주)비씨카드, 소상공인연합회 등이다.

통신 빅데이터 플랫폼은 ‘통신 마켓플레이스’, ‘분석/시각화’, ‘커뮤니티’, ‘빅데이터 서비스’, ‘이용안내’, ‘마이페이지’로 구성되어 있다(〈표 2-9〉 참조).

[그림 2-8] 통신 빅데이터 플랫폼 개념도



자료: 통신 빅데이터 플랫폼 중 '이용안내' 메뉴.  
<https://www.bigdata-telecom.kr/invoke/SOKBP2517/>.

<표 2-9> 통신 빅데이터 플랫폼 구성 메뉴

통신 마켓플레이스	분석/시각화	커뮤니티	빅데이터 서비스	이용안내	마이페이지
-데이터마켓 -분석해듀오	-기본분석 -전문분석	-오픈랩 이용 신청 -분석리포트 요청 -사업협력 요청 -공지사항 -Q&A -FAQ	-kt BigSight -유레카 -잘나가게 -식품 Trend Sensing -AI 큐레이션 식물관리 -마켓인텔리 전스 -Social감정 읽어주기 -웹기반 AI 분석서비스 -배달최적화	-서비스이용 안내 -개인정보 처리방침	-데이터신청/ 이용현황 -분석/시각화 관리 -회원정보 관리 -내글 현황 -이용권현황 -분석해듀오 관리 -AI분석 서비스관리

자료: 통신 빅데이터 플랫폼. <https://www.bigdata-telecom.kr/>.



## 9. 지역경제 빅데이터 플랫폼: 경기지역경제포털 (<https://bigdata-region.kr/#/>)

경기지역 경제포털은 경기지역경제 활성화를 위해 경기지역의 데이터를 한곳에서 확인해볼 수 있고 사용가능한 플랫폼으로 데이터 개방과 데이터 공유, 그리고 데이터 기여도에 따라 사용자에게 데이터 거래일부를 지역화폐로 배당해주고 있다.

데이터는 영상, 지역현황, 친환경충전인프라, 취업, 지역트랜드, 지역화폐, 기업신용보증, 소상공인, 기업 영역으로 구분하고 있으며 데이터 제공기관은 ㈜아임클라우드, 경기신용보증재단, 경기도경제과학진흥원, 경기도일자리재단, 한양대학교, 한국생산기술연구원, 국토연구원 등이다.

경기지역경제포털은 ‘데이터셋’, ‘데이터활용’, ‘서비스’, ‘플랫폼소개’, ‘이용안내’ 메뉴로 구성되어 있다(〈표 2-10〉 참조).

〈표 2-10〉 경기지역경제포털 구성 메뉴

데이터셋	데이터 활용	서비스	플랫폼소개	이용안내
-데이터셋 -융합데이터 -통합데이터 검색	-활용갤러리	-분석서비스신청 -기업입지추천 서비스 -소상공인투자 환경분석 -분석경험공유 -기업신용분석 -데이터시각화	-소개 -데이터배당	-공지사항 -문고답하기 -자주하는질문 -데이터문의 -이벤트

자료: 경기지역경제포털. <https://bigdata-region.kr/#/>.

## 10. 중소중견기업 빅데이터 플랫폼 (<https://www.bigdata-sme.kr/#/datastore/landing>)

중소·중견기업 빅데이터 플랫폼은 마켓 플레이스를 통해 데이터를 유통하고, AI 알고리즘으로 학습된 서비스 및 API를 제공하여 데이터 소비와 공급의 연계시스템을 구축하는 플랫폼이다([그림 2-9] 참조).

[그림 2-9] 중소중견기업 빅데이터 플랫폼 구성도



자료: 중소중견기업 빅데이터 플랫폼 중 '플랫폼 소개' 메뉴.  
<https://www.bigdata-sme.kr/#/datastore/intro>.

데이터는 부동산, 산업경제, 교육, 무역, R&D, 고용(채용), 기업정보, 생산성, 회계, 뉴스, 법령 등 영역으로 구분하고 있으며 데이터 제공기관은 (주)빅밸류, 한국무역정보통신, 한국생산성본부, 한국산업기술진흥협회, 더존비즈온, 선도소프트, 인크루트, NICE평가정보, 한국물가정보 등이다.

중소·중견기업 빅데이터 플랫폼은 ‘플랫폼소개’, ‘데이터’, ‘커뮤니티’, ‘경진대회’ 메뉴로 구성되어 있다(〈표 2-11〉 참조).

〈표 2-11〉 중소기업 빅데이터 플랫폼 구성 메뉴

플랫폼소개	데이터	커뮤니티	경진대회
-플랫폼소개	-데이터마켓 -데이터맵 -데이터분석도구 -혁신서비스	-커뮤니티	-지난경진대회

자료: 중소기업 빅데이터 플랫폼 <https://www.bigdata-sme.kr/#/datastore/intro>.

## 11. 산림빅데이터 플랫폼: 산림빅데이터 거래소 (<https://www.bigdata-forest.kr/>)

산림 빅데이터거래소는 서비스개발, 데이터분석 등 사업에 필요한 산림, 임업, 임산업 관련 각종 데이터를 온라인으로 판매하거나 구매할 수 있는 플랫폼으로 한국임업진흥원에서 주관하고 있다.

데이터는 산림휴양/복지, 산림자원관리, 산림재해안전 등 영역으로 구분하여 임상도, 산림탄소흡수량과 같은 산림자원에 대한 데이터뿐만 아니라 트래킹 코스, 산림치유정보, 산림관광, 교통, 생물종 유전체, 약용생물자원, 산림재해, 공간정보, 항공영상, 기상·미세먼지, 임산물유통정보, 산림 소비트렌드 정보 등도 제공하고 있다.

데이터 제공기관은 한국산림복지진흥원, 한국한의학연구원, 한국임업진흥원, nemopartnersBA, (주)마켓링크, (주)아로정보기술, ECOEYE 등이다.

산림 빅데이터거래소는 ‘데이터상품’, ‘소개’, ‘빅데이터서비스’, ‘데이터맵’, ‘커뮤니티’, ‘분석환경’, ‘타플랫폼상품’ 메뉴로 구성되어 있다(〈표 2-12〉 참조).

〈표 2-12〉 산림 빅데이터거래소 구성 메뉴

데이터상품	소개	빅데이터 서비스	데이터맵	커뮤니티	분석환경	타플랫폼 상품
-데이터 상품 상세 검색	-데이터 거래소 -참여기업 -생태계 프로그램 -플랫폼 브로셔 -홍보 동영상	-시각화 포털 -혁신 서비스 -활용사례 -검색통계	-데이터맵 (전체) -데이터맵 (키워드)	-공지사항 -생태계 프로그램 -문의 게시판 -데이터 구매상담 신청 -FAQ	-분석환경 소개 (Python, RStudio, Zeppelin)	-태플랫폼 상품 상세 검색

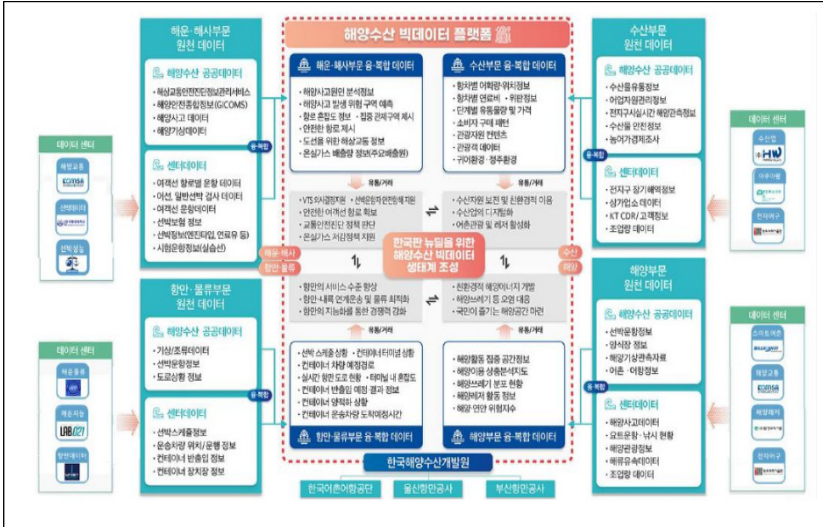
자료: 산림 빅데이터거래소. <https://www.bigdata-forest.kr/>.

## 12. 해양수산 빅데이터 플랫폼: 해양수산 빅데이터 거래소 (<https://www.bigdata-sea.kr/>)

해양수산 빅데이터 거래소는 해양수산 분야의 데이터 수집, 가공, 연계, 유통 판매를 위한 데이터 유통 거래 마켓 플레이스로(〔그림 2-10〕 참조) 한국해양수산개발원이 주관하고 있으며 플랫폼 자동화, 데이터 가치화, 데이터 유통 편의성, 신규 가치창출을 위한 유연한 플랫폼을 통해 플랫폼과 센터 간 실시간 데이터 연계 강화, 선순환되는 해양수산 데이터 유통체계를 마련하고자 한다.

데이터는 해양산업, 해양환경, 국제협력, 항만운영, 해사안전, 항만건설, 공동행정 등 영역으로 구분하며 데이터 제공기관은 해운항만물류정보협회, 울산과학기술원, 광주과학기술원, 환경과학기술, 랩오투원, 목포대학교 산학협력단, 한국해양교통안전공단 등이다.

[그림 2-10] 해양수산 빅데이터 거러소 구성도



자료: 해양수산 빅데이터 거러소 중 '소개' 메뉴.  
[https://www.bigdata-sea.kr/main/contentsView.do?contentsId=CON\\_000007](https://www.bigdata-sea.kr/main/contentsView.do?contentsId=CON_000007).

해양수산 빅데이터거러소는 '소개', '데이터검색', '데이터분석', '빅데이터 교육', '커뮤니티' 메뉴로 구성되어 있다(〈표 2-13〉 참조).

〈표 2-13〉 해양수산 빅데이터거러소 구성 메뉴

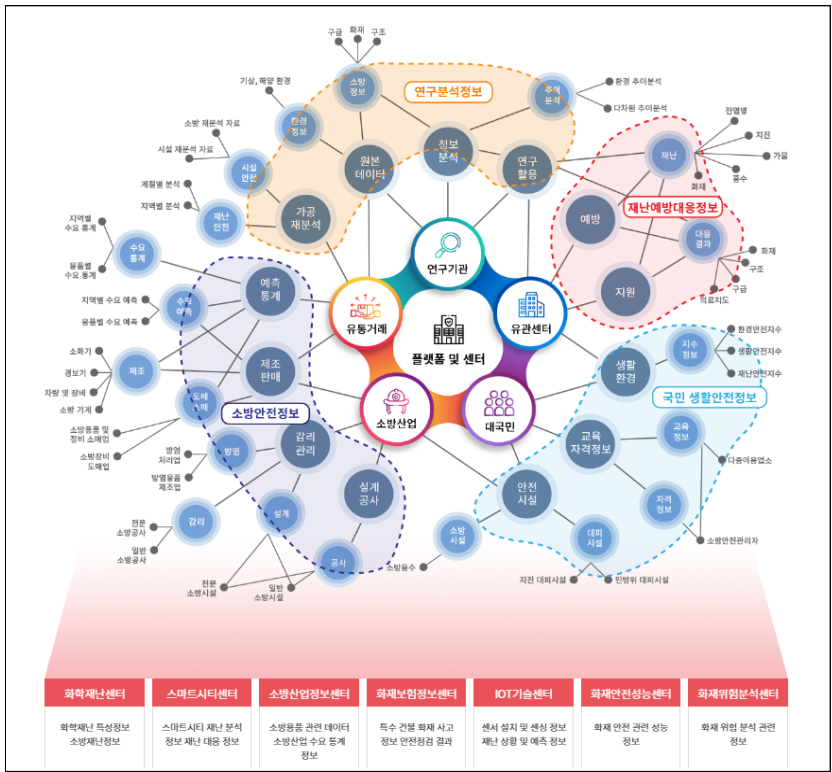
소개	데이터검색	데이터분석	빅데이터 교육	커뮤니티
<ul style="list-style-type: none"> <li>-해양수산플랫폼 소개</li> <li>-이용안내</li> <li>-플랫폼참여기관 소개</li> <li>-센터 참여기관 소개</li> <li>-뉴스정보</li> </ul>	<ul style="list-style-type: none"> <li>-전체데이터</li> <li>-이슈데이터</li> <li>-신규데이터</li> </ul>	<ul style="list-style-type: none"> <li>-데이터상황판</li> <li>-시각화도구</li> <li>-데이터활용사례</li> <li>-데이터 혁신 서비스</li> </ul>	<ul style="list-style-type: none"> <li>-교육소개</li> <li>-교육자료</li> <li>-시각화 교육 신청</li> </ul>	<ul style="list-style-type: none"> <li>-공지사항</li> <li>-자주하는질문</li> <li>-질의응답</li> <li>-자료실</li> <li>-데이터요청</li> <li>-데이터 안심 구역 이용신청</li> <li>-사용자 의견 청취</li> </ul>

자료: 해양수산 빅데이터거러소. <https://www.bigdata-sea.kr/>.

### 13. 소방안전 빅데이터 플랫폼(<https://www.bigdata-119.kr/>)

소방안전 빅데이터 플랫폼은 각종 소방안전 및 소방산업 분야 데이터를 수집·저장·유통하는 시스템으로 화학재난센터, 스마트시티센터, 소방산업센터, 재난보험센터, IoT기술센터, 화재위험분석으로 구성되어 있다 ([그림 2-11] 참조).

[그림 2-11] 소방안전 빅데이터 플랫폼 구성도



자료: 소방안전 빅데이터 플랫폼 중 '플랫폼 소개' 메뉴.  
<https://www.bigdata-119.kr/info/introduction>.

데이터는 융합데이터, 소방안전, 소방산업, 재난보험, 소방IoT, 화재안전성능, 화재위험분석, 지하배관안전 영역으로 구분되어 있으며 데이터 제공기관은 한국소방안전원, 한국소방산업기술원, (사)한국화재보험협회, 한방유비스, 울산광역시소방본부, 제주소방본부, 세종특별자치시소방본부 등이다.

소방안전 빅데이터 플랫폼은 ‘데이터마켓’, ‘데이터서비스’, ‘지원&신청’, ‘플랫폼소개’ 메뉴로 구성되어 있다(〈표 2-14〉 참조).

〈표 2-14〉 소방안전 빅데이터 플랫폼 구성 메뉴

데이터마켓	데이터서비스	지원&신청	플랫폼소개
-상품검색 -데이터마켓소개	-데이터 서비스 소개 -분석서비스 도구 -SAFETY119 -화재 연관성 분석 -데이터 시각화 -지하매설배관 안전 관리 -공간정보 데이터 시각화	-교육 -창업지원	-소개 -공지사항 -Q&A -FAQ -이벤트 -맞춤데이터 신청

자료: 소방안전 빅데이터 플랫폼. <https://www.bigdata-119.kr/>.

#### 14. 스마트치안 빅데이터 플랫폼 (<https://www.bigdata-policing.kr/>)

스마트치안 빅데이터 플랫폼은 생활안전·수사·교통 등 기능별 빅데이터를 통합하고 위험예측·분석이 가능한 민간·공공 데이터를 결합하여 맞춤형 빅데이터 서비스를 제공하는 플랫폼으로 경찰대학 치안정책연구소에서 주관하고 있다(〈그림 2-12〉 참조).

[그림 2-12] 스마트치안 빅데이터 플랫폼 서비스 구성도



자료: 이관주. (2020.12.27.). 현실로 다가온 '스마트 치안'...순찰 경로·방법·시설 설치 "인공지능 추천". 아시아경제. <https://www.asiae.co.kr/article/2020122619243856497>.

제공하는 데이터는 치안/범죄, 사이버치안/범죄, 교통, 재난, 지역환경, 기타 등 영역으로 구분하고 있으며 제공기관은 한국토지주택공사, 지란지교, 이투스, 연합뉴스, 아이티로, 수원시청, 세종특별자치시, 경찰청 등이다.

스마트치안 빅데이터 플랫폼은 '빅데이터마켓', '분속/시각화', '혁신 서비스', '커뮤니티', '스마트치안', 'G자유' 메뉴로 구성되어 있다(표 2-15) 참조).



〈표 2-15〉 스마트치안 빅데이터 플랫폼 구성 메뉴

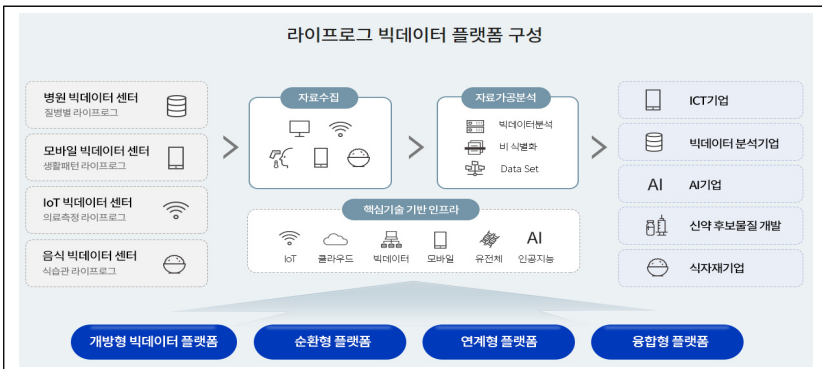
빅데이터마켓	분석/시각화	혁신서비스	커뮤니티	스마트치안	G자유
-데이터검색 -맞춤데이터 신청 -오픈API	-분석/연구 보고서 -분석/학습 신청 -데이터 시각화 -BIX5 -SAP BO	-우리동네 안심지도 -112신고 예측 -마이폴 -종합안전 서비스 -스탑퍼싱	-활용사례 -자료실 -기간별 인기 키워드검색 -치안연구 기획실	-플랫폼소개 -거버넌스 소식 -경찰24시	-자유그룹

자료: 스마트치안 빅데이터 플랫폼. <https://www.bigdata-policing.kr/>.

### 15. 라이프로그 빅데이터 플랫폼 (<https://www.bigdata-lifelog.kr/>)

라이프로그(lifelog)는 개인의 일상에 대한 기록을 의미하는 것으로 라이프로그 빅데이터 플랫폼은 병원 안의 의료데이터와 병원 밖의 일상 데이터를 포괄적으로 적재, 관리, 유통하는 플랫폼으로 원주연세의료원이 주관하고 있다(〔그림 2-13〕 참조).

〔그림 2-13〕 라이프로그 빅데이터 플랫폼 구성도



자료: 라이프로그 빅데이터 플랫폼 중 '서비스소개' 메뉴.  
<https://www.bigdata-lifelog.kr/portal/intro/dbIntro>.

제공하는 데이터는 성별, 연령대별(60대미만, 60대이상), 질병별(당뇨, 일반), 측정항목별(공복혈당, 총 콜레스테롤, LDL, HDL, 중성지방, 헤모글로빈, 당화혈색소, 혈당, 나이, 키, 체중, 허리둘레, 수축기혈압, 이완기혈압, 신체질량지수)로 구분하고 있으며 제공기관은 원주연세의료원, 강원대학교, 고려대의료원, 굿닥, 대한청각학회, 아이센스, 한림대학교, 헬스브릿지, LG유플러스 등이다.

라이프로그 빅데이터 플랫폼은 ‘데이터찾기’, ‘데이터활용’, ‘커뮤니티’, ‘서비스소개’, ‘교육프로그램’ 메뉴로 구성되어 있다(〈표 2-16〉 참조).

〈표 2-16〉 라이프로그 빅데이터 플랫폼 구성 메뉴

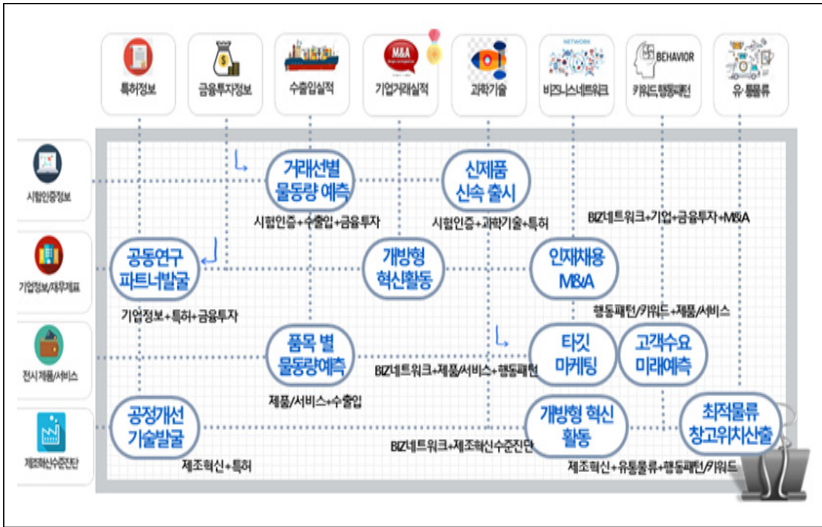
데이터찾기	데이터활용	커뮤니티	서비스소개	교육프로그램
-데이터목록 -데이터맵 -데이터개방요청	-데이터시각화 -데이터분석도구 -데이터분석 서비스 -Open API -활용사례	-공지사항 -보도자료 -FAQ -Q&A	-라이프로그DB 소개 -서비스가이드 -데이터구축현황	-교육프로그램

자료: 라이프로그 빅데이터 플랫폼 <https://www.bigdata-lifelog.kr/>

## 16. 디지털산업혁신 빅데이터 플랫폼 (<https://www.bigdata-dx.kr/>)

디지털산업혁신 빅데이터 플랫폼은 지역산업 생태계를 분석하여 기업 간거래(B2B) 관계, 인수합병(M&A) 금융·투자 정보, 기술개발 정보 등을 종합해 시장 및 경쟁 기업을 분석 가능하며 각 기업은 이 같은 정보를 신규 제품이나 서비스를 개발하는데 적용할 수 있다(〈그림 2-14〉 참조).

[그림 2-14] 디지털산업혁신 빅데이터 활용사례



자료: 산업통상자원부 보도자료(2021.2.25.). 기업의 비즈니스 혁신, 산업 빅데이터 플랫폼으로 빠르고 쉽게!. p.2.

제공하는 데이터는 M&A정보, 과학기술, 금융, 기업정보, 물류, 수출입, 시험인증, 유통, 인증, 재무, 투자유치, 특허, 펀드 영역으로 구분하고 있으며 제공기관은 한국산업기술시험원, STHIS, 가이온, 예프엔가이드, 알리콘, 한국M&A거래소, 한국평가데이터, 엔에이치엔데이터, 대한전자공학회 등이다.

디지털산업혁신 빅데이터 플랫폼은 ‘데이터몰’, ‘스튜디오’, ‘오픈스테이지’, ‘플레이그라운드’, ‘캠퍼스’, ‘더보기’ 메뉴로 구성되어 있다(〈표 2-17〉 참조).

〈표 2-17〉 디지털산업혁신 빅데이터 플랫폼 구성 메뉴

데이터	스튜디오	오픈스테이지	플레이 그라운드	캠퍼스	더보기
-데이터마켓 -데이터 라운지	-데이터 스토리 -데이터 갤러리 -데이터 솔루션	-데이터 챌린지	-재능마켓 -피플인피플 -데이터톡	-데이터 강의실 -온라인 연구실 -데이터 오픈랩	-플랫폼소개 -공지사항 -데이터지도 -랭킹보드 -문의하기

자료: 디지털산업혁신 빅데이터 플랫폼. <https://www.bigdata-dx.kr/>.

### 17. 농식품 빅데이터 플랫폼: 농식품 빅데이터 거래소 (<https://kadx.co.kr/>)

농식품 빅데이터 거래소는 농식품 데이터를 온라인으로 판매하거나 구매할 수 있는 농식품 빅데이터 플랫폼으로 농식품 분야의 데이터덤 역할을 수행하고 있으며 한국농수산물유통공사가 주관하고 있다.

거래소의 데이터 영역으로는 농업, 축산, 유통, 소비, 국제, 농식품 공공데이터 영역으로 구분하고 있으며 데이터 거래소 회원은 상품을 판매하는 ‘판매회원’과 상품을 구매하는 ‘기업회원’으로 구분하고 있다. 판매사(센터)로는 이지팜, 이지 에이치엘디, 나이스지니데이터, 전남문하정보산업진흥원, 케이플러스, 장보고식자재마트, 한국농수산물유통공사, 유로지스넷, 원데이터기술, 한국무역통계진흥원, 플랜인피닛, 한국마사회 등이 있다.

농식품 빅데이터 거래소 플랫폼 메뉴는 ‘데이터상품’, ‘데이터서비스’, ‘데이터분석’, ‘이용안내’, ‘플랫폼소개’로 구성되어 있다(〈표 2-18〉 참조).

〈표 2-18〉 농식품 빅데이터 거래소 구성 메뉴

데이터상품	데이터서비스	데이터분석	이용안내	플랫폼 소개
-데이터상품 -데이터 신청 및 문의	-전국 도매시장 가격 비교 -농산물 물류정보 -농산물별 출하지 추천 -양돈농가 자가진단 -출하반장	-시각화서비스 -구독서비스	-공지사항 -문의 게시판 -FAQ	-데이터 거래소 소개 -센터 소개 -명예센터 소개

자료: 농식품 빅데이터 거래소. <https://kadx.co.kr/>.



사람을  
생각하는  
사람들



KOREA INSTITUTE FOR HEALTH AND SOCIAL AFFAIRS

# 제3장

## 소셜빅데이터 분석





## 제 3 장 소셜빅데이터 분석

제3장에서는 소셜빅데이터 분석 플랫폼 중 하나인 LUCY 2.0을 기관 차원에서 인프라로 활용한 바 장애인, 노인, 아동, 건강보장이라는 주제에 대해 관련 키워드 분석 수행 결과와 함께 보건복지데이터 포털에서 활용한 사례를 제시하였다.

### 1. 자료원

소셜빅데이터는 분석하고자 하는 목적에 따라 데이터를 수집해야 하는데, 윤용찬(2018)에 따르면 소셜데이터 분석시 정확도를 높이기 위해서는 첫째 많은 양의 데이터를 단순하게 수집하기보다 분석의 목적과 주제에 따라 분석 대상(채널) 데이터를 선정하고, 둘째 데이터의 품질을 높이기 위해 수집한 전체 데이터에서 가비지(garbage) 데이터를 제거하여 분석하는 것이 중요하다.

이에 이번 연구에서는 소셜빅데이터 분석을 위해 알에스엔이 자체 개발한 LUCY 2.0을 활용하였다(〈표 3-1〉 참조). LUCY 2.0은 유튜브, 트위터, 인스타그램, 카페, 블로그 등 12개 채널의 소셜 데이터를 인공지능(AI)으로 분석·데이터화해 제품 소비, 사용 패턴, 트렌드를 발굴할 수 있도록 실시간 분석·통계 자료를 대시보드로 제공하는 소셜빅데이터 분석 플랫폼이다.

48 2021년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

〈표 3-1〉 LUCY 2.0 수집 채널 개요(2021년 12월 9일 기준)

연번	채널	대상 사이트	수집 유형	수집범위	사이트 및 키워드 수
1	뉴스 (언론)	일간지, 월간지, 인터넷신문, 지방지, 전문지 등	전체	선정된 사이트의 모든 게시글	약 3,200개
2	커뮤니티	국내 주요 커뮤니티 (디시인사이드, 보배드림, 판지일보 등)	전체	지정 계정의 모든 게시글	약 1,300개
3	블로그	네이버 블로그, 다음 블로그, 티스토리	전체	지정 계정의 모든 게시글	-
			메타	키워드 검색결과 게시글	메타 키워드 방식 약 8,000개 키워드
4	카페	네이버 카페, 다음 카페	전체	지정 계정의 모든 게시글	-
			메타	키워드 검색결과 게시글	메타 키워드 방식 약 8,000개 키워드
5	트위터	트위터	전체	지정 계정의 모든 게시글	-
			메타	키워드 검색결과 게시글	메타 키워드 방식 약 6,000개 키워드
6	인스타그램	해시태그 포함 게시글	메타	메타 해시태그 방식	메타 해시태그 방식 약 9,000개 키워드
7	유튜브	유튜브	전체	지정 채널의 모든 게시글	-
			메타	키워드 검색결과 게시글	메타 키워드 방식 약 7,500개 키워드
8	페이스북	페이스북	지정된 계정	선정 팬페이지의 모든 게시글	약 15,000개
9	카카오스토리	카카오스토리	메타	메타 해시태그 방식	메타 키워드 방식 약 6,500개 키워드

연번	채널	대상 사이트	수집 유형	수집범위	사이트 및 키워드 수
10	지식인	네이버 지식인	전체	모든 지식카테고리	-
11	기업/단체	주요 기업, 은행 등의 공지/소식 등	지정된 사이트	지정된 사이트의 모든 게시물	약 2,200개
12	정부/공공	한국소비자원, 식품안전정보포털의 게시판 등	지정된 사이트	지정된 사이트의 모든 게시물	약 1,200개
-	Reply	네이버 뉴스, 다음 뉴스, 네이버 뉴스 기사의 댓글	댓글	기사에 작성되는 댓글을 최대 3일간 수집	2개(네이버/다음에 발표되는 모든 기사)

주: 채널별 데이터 수집 범위는 수시로 업데이트됨  
 자료: 알에스엔 내부자료.

LUCY 2.0의 주요 기능으로는 정기 분석이 필요한 대상(브랜드, 제품, 캠페인 등)을 사용자가 키워드로 등록하여 지속적으로 모니터링하는 아이템분석(비교할 아이템을 비교구간 내 모니터링)이 있으며, 사회 전반의 트렌드나 이슈 키워드 등 원하는 검색 키워드에 대한 정보량, 감성, 채널, 연관 키워드를 모니터링하는 고급검색, 해당 키워드에 맞는 토픽을 매칭<sup>1)</sup>시켜 모니터링하는 토픽검색, 해시태그를 사용하는 채널<sup>2)</sup>을 기준으로 키워드 현황을 모니터링하는 해시태그검색, 지역, 연령대, 성별 등 프로필 특성이 나타나는 키워드를 분석하는 프로필검색, 키워드에 대해서 시간(Time), 장소(Place), 상황(Occasion), 대상(Person)별 분포를 모니터링하는 TPOP검색이 있다(LUCY 2.0 브랜드 홈페이지, [그림 3-1], [그림 3-2] 참조).

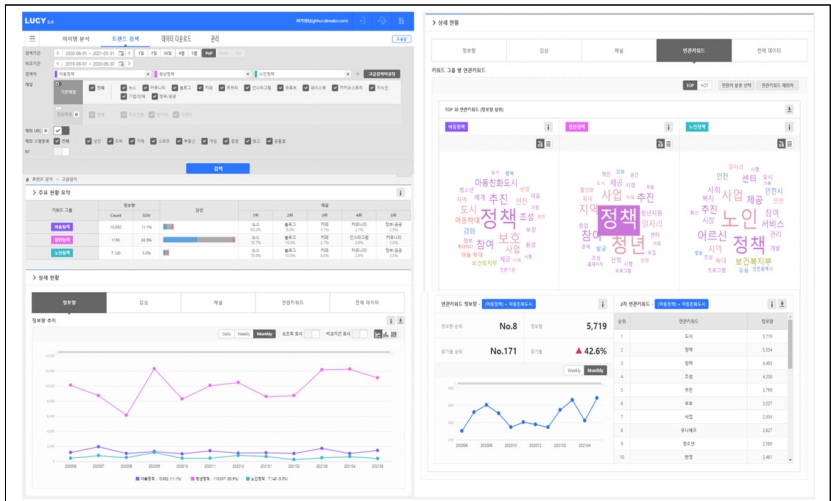
- 1) 알에스엔에서 분류한 정치, 경제, 사회, 스포츠, 연예, 문화, 생활, IT/과학, 산업, 환경 등 10개의 대항목과 53개의 소항목으로 분류하여 전체적인 언급 비율과 정보량을 확인할 수 있음. 예를 들어 정치 항목에는 정책, 정부, 정당, 지자체, 외교, 국방안보의 소항목으로 구성되며, 정치-정책에는 정부에서 추진 중이거나 예정인 정책 관련 정보(예, 일 자리창출, 의료민영화 등)가 분류됨(알에스엔 내부자료)
- 2) 인스타그램, 트위터, 페이스북, 유튜브, 카카오토리, 블로그, 카페가 해당됨.

[그림 3-1] LUCY 2.0 주요 기능



자료: 알에스엔(2021). LUCY 2.0 Product Introduction. p.13.

[그림 3-2] LUCY 2.0 트렌드 검색 - 고급검색 화면 예시



자료: LUCY 2.0. <https://lucy2.realsn.com/> 에서 인용.

## 2. 장애인 정책관련 키워드 분석

### 가. 키워드 설정

보건복지분야 장애인 정책에 대한 동향 분석을 하기 위해 ‘장애인 정책’ 키워드를 대표 키워드로 선정하였으며, 이외 장애인 정책과 관련있는 키워드는 보건복지부 홈페이지의 정책메뉴를 활용하여 전문가 회의를 통해 ‘장애인 서비스, 장애인 보건, 장애인 복지, 장애인 보육, 장애인 양육, 장애인 세제, 장애인 의료, 장애인 감면, 장애인 등록, 장애인 지역사회복지사업, 장애인 재활, 장애인 권익지원, 장애인 차별금지, 장애인 거주, 장애인 편의, 장애인 권리협약, 장애인 자립, 장애인 연금, 장애 수당, 장애인 우선구매, 장애인 일자리, 장애인 공공후견, 장애인 활동, 장애인 가족’ 키워드를 포함하였다.

키워드는 ‘인접’<sup>3)</sup>기능을 사용하여, ‘장애인 정책, 장애인 서비스, 장애인 보건, 장애인 복지’의 키워드는 5글자 내에 포함되는 문서를 모두 수집하였고, 그 외 키워드는 1글자 내에 포함되는 문서를 수집하였다. 예를 들어 ‘장애인 가족’ 키워드는 ‘장애인’과 ‘가족’이라는 두 키워드 사이에 1글자를 허용한 것으로 ‘...장애인의 가족...’ 또는 ‘...돌봄 가족 중 장애인...’의 문서들이 수집될 수 있다. 단, LUCY 2.0에서 지정한 성인, 도박, 거래, 스포츠, 부동산, 게임, 증권, 광고 등의 사이트의 문서는 제외하였

3) 검색어 설정 기능은 일반, 구문, 인접, 제외로 구분됨.

- 일반: 개별 단어는 정확히 포함(구문 조건), 한 줄 내에 띄어쓰기로 구분된 단어들은 모두 포함(AND 조건)된 경우 검색
- 구문: 한 줄 내에 띄어쓰기로 구분된 단어들 정확히 포함(구문 조건)된 경우 검색
- 인접: 한 줄 내에 띄어쓰기로 구분된 단어들 사용자가 원하는 글자 수 내에 모두 포함(인접 조건)된 경우 검색
- 제외: 개별 단어는 정확히 포함(구문 조건), 한 줄 내에 띄어쓰기로 구분된 단어들은 모두 포함(AND 조건)된 경우 검색에서 제외

으며, 성범죄 또는 학대관련 범죄의 경우 장애인 복지시설 취업제한<sup>4)</sup>의 문구로 인해 장애인 정책과 관련없는 문서가 수집되는 것을 방지하기 위해 ‘취업제한’의 키워드도 제외하였다. 트위터 채널에서 리트윗(Retweet)은 포함하지 않았다(〈표 3-2〉 참조).

〈표 3-2〉 장애인 정책과 관련하여 LUCY 2.0에서 설정된 최종 검색어

((("장애인 정책"~5) OR ("장애인 서비스"~5) OR ("장애인 보건"~5) OR ("장애인 복지"~5) OR ("장애인 보호"~1) OR ("장애인 양육"~1) OR ("장애인 세제"~1) OR ("장애인 의료"~1) OR ("장애인 감면"~1) OR ("장애인 등록"~1) OR ("장애인 지역사회복지사업"~1) OR ("장애인 재활"~1) OR ("장애인 권익지원"~1) OR ("장애인 차별금지"~1) OR ("장애인 거주"~1) OR ("장애인 편의"~1) OR ("장애인 권리협약"~1) OR ("장애인 자립"~1) OR ("장애인 연금"~1) OR ("장애 수당"~1) OR ("장애인 우선 구매"~1) OR ("장애인 일자리"~1) OR ("장애인 공공후견"~1) OR ("장애인 활동"~1) OR ("장애인 가족"~1))) NOT ("취업제한") OR ("취업" AND "제한"))

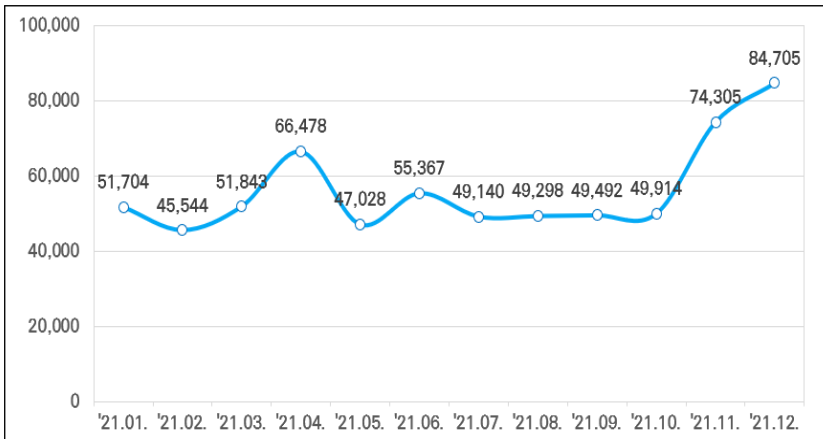
4) 장애인복지법 제59조의3(장애인 관련 기관에의 취업제한 등) ① 법원은 장애인학대관련 범죄나 성범죄(「성폭력범죄의 처벌 등에 관한 특례법」 제2조에 따른 성폭력범죄 또는 「아동·청소년의 성보호에 관한 법률」 제2조제2호에 따른 아동·청소년대상 성범죄를 말한다. 이하 같다)로 형 또는 치료감호를 선고하는 경우에는 판결(약식명령을 포함한다. 이하 같다)로 그 형 또는 치료감호의 전부 또는 일부의 집행을 종료하거나 집행이 유예·면제된 날(벌금형을 선고받은 경우에는 그 형이 확정된 날을 말한다)부터 일정기간(이하 “취업제한기간”이라 한다) 동안 다음 각 호에 따른 시설 또는 기관(이하 “장애인관련기관”이라 한다)을 운영하거나 장애인관련기관에 취업 또는 사실상 노무를 제공할 수 없도록 하는 명령(이하 “취업제한명령”이라 한다)을 장애인학대관련범죄나 성범죄(이하 “장애인학대관련범죄등”이라 한다) 사건의 판결과 동시에 선고(약식명령의 경우에는 고지를 말한다)하여야 한다. 다만, 재범의 위험성이 현저히 낮은 경우, 그 밖에 취업을 제한하여서는 아니 되는 특별한 사정이 있다고 판단하는 경우에는 그러하지 아니한다(국가법령정보센터-장애인복지법 <https://www.law.go.kr/법령/장애인복지법>).

## 나. 분석 결과

2021년 1월부터 12월까지 수집된 장애인 정책 관련 문서량은 총 674,818건이다. 장애인복지법 제14조 1항<sup>5)</sup>에 따라 장애인의 날이 있는 4월의 경우는 장애인 관련 문서량이 급증하였고, 2022년 대통령선거를 앞두고, 각 정당의 후보자 확정 후<sup>6)</sup> 관련 공약 및 선거활동으로 인해 11월을 기점으로 다시 증가하였다([그림 3-3], <표 3-3> 참조).

[그림 3-3] 2021년 월별 '장애인 정책' 문서량

(단위: 건)



자료: 소셜빅데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

5) 장애인복지법 제14조(장애인의 날)①장애인에 대한 국민의 이해를 깊게 하고 장애인의 재 활의욕을 높이기 위하여 매년 4월 20일을 장애인의 날로 하며, 장애인의 날부터 1주간을 장애인 주간으로 한다. ②국가와 지방자치단체는 장애인의 날의 취지에 맞는 행사 등 사 업을 하도록 노력하여야 한다.(국가법령정보센터-장애인복지법 <https://www.law.go.kr/법령/장애인복지법>).

6) 국민의 힘 대선 후보 윤석열 11월 5일 확정.

54 2021년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

〈표 3-3〉 2021년 4월 20일 '장애인의 날' 뉴스 예

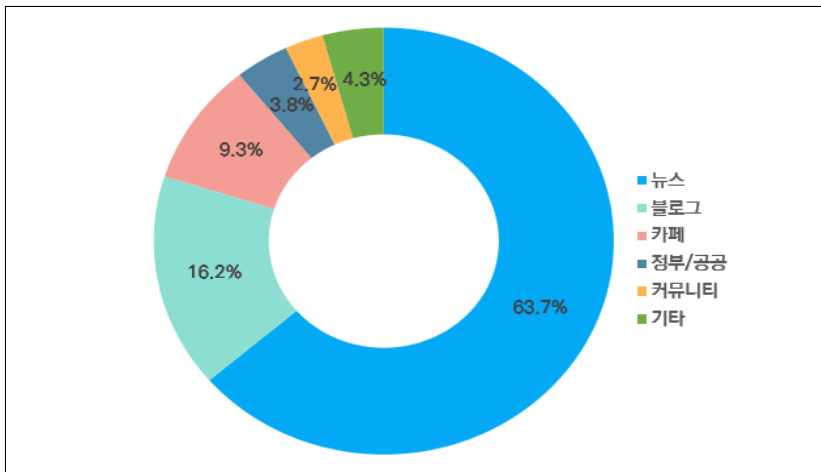
수집일시	제목
2021-04-20 19:29	공직사회서 여전한 '장애의 벽'
2021-04-20 18:59	[장애인의 날] "오늘은 내가 바리스타"
2021-04-20 16:49	순천시 '제41회 장애인의 날' 기념행사 개최
2021-04-20 16:49	"장애가 행복 걸림돌 안되도록..." 광주시 장애인의날 기념식
2021-04-20 15:58	장애인 5명 중 1명은 기초수급자, "코로나19로 더 힘들다"
2021-04-20 15:58	장애인, 코로나에 "외출 전혀 안 해" 2배↑...의료이용 빈도 '폭'
2021-04-20 12:05	오늘은 '장애인의 날'...혼자 사는 장애인만 71.3만명
2021-04-20 11:56	국민의힘 초선 "현실과 괴리된 정부 장애인 정책, 혼란만 가중시켜"

자료: 소셜 데이터 분석 시스템 LUCY 2.0에서 추출.

674,818건의 문서량에 대한 채널별 문서비율을 살펴보면 뉴스가 63.7%로 가장 많았고, 블로그 16.2%, 카페 9.3% 순으로 나타났다(〈그림 3-4〉 참조).

〈그림 3-4〉 2021년 채널별 '장애인 정책' 문서 비율

(단위: %)



주: 기타에는 트위터, 인스타그램, 유튜브, 페이스북, 카카오토리, 지식인, 기업/단체가 포함됨.  
 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.





장애인 정책 월별 정보량 상위 20개 연관 키워드 중 부정과 관련된 키워드는 ‘어렵다’, ‘장애’ 키워드가 추출되었고, 긍정과 관련된 키워드는 ‘노력’, ‘강화’, ‘도움’ 등이 추출되었다(〈표 3-4〉 참조).

토픽 검색 결과 전체 674,818 건 중 664,391건(98.4%)이 LUCY 2.0에서 설정한 토픽<sup>8)</sup>과 매칭되었고, 사회>보건복지(252,633건), 사회>교육(61,231건), 생활>의료(56,332건) 순으로 토픽 관련 언급이 가장 많았다(〈그림 3-6〉 참조).

---

8) 사회>보건복지 토픽은 보건복지부, 기초생활과 같은 보건복지 법·종류·기관명 관련 정보일 경우 분류되고, 사회>교육 토픽은 교육 담당 기관명, 교육의 분류, 국가 교육방침, 교육비, 교육의 문제점관련 정보일 경우 분류되고, 생활>의료 토픽은 국민건강보험공단, 금연과 같은 질병예방법, 생활 양상, 건강 관련 기관, 질병명 관련 정보일 경우 분류됨.

〈표 3-4〉 2021년 월별 '장애인 정책' 정보량 상위 20개 연관 키워드

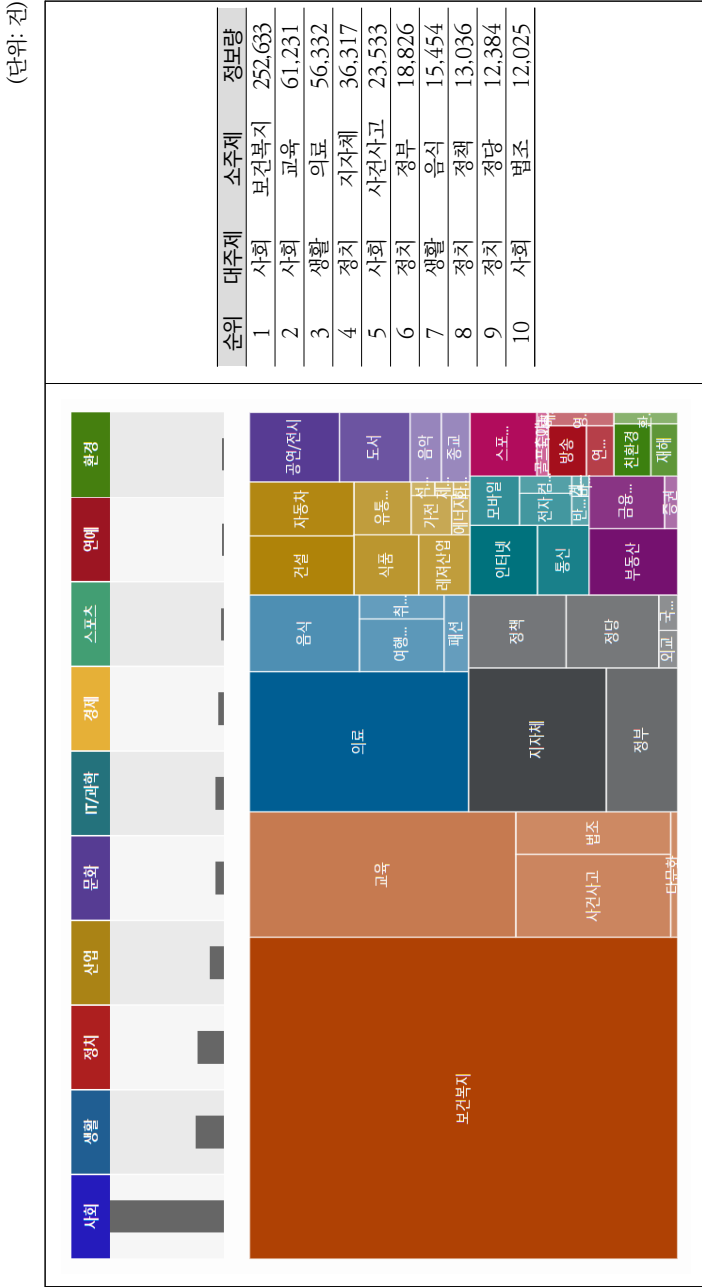
순위	전체	21.1.	21.2.	21.3.	21.4.	21.5.	21.6.	21.7.	21.8.	21.9.	21.10.	21.11.	21.12.
1	사업	사업	사업	사업	사업	사업	사업	센터	센터	지역	사업	사업	지역
2	지역	서비스	지역	지역	지역	지역	지역	사업	사업	사회	지역	센터	사업
3	센터	센터	서비스	서비스	센터	서비스	센터	지역	지역	활동	센터	지역	사회
4	사회	지역	서비스	센터	사회	어렵다	서비스	사회	사회	활동	사회	지역	센터
5	복지	복지	사회	사회	복지	사회	사회	어렵다	장에	사업	복지	활동	복지
6	서비스	사회	사회	복지	어렵다	센터	복지	어렵다	어렵다	어렵다	활동	복지	활동
7	어렵다	환대	방문	추진	서비스	복지관	어렵다	복지관	운영	복지	생활	참여	어렵다
8	복지관	가족	추진	복지관	복지관	복지	참여	운영	복지	생활	운영	생활	운영
9	참여	추진	복지관	참여	참여	참여	참여	어렵다	생활	장에	복지관	운영	기관
10	활동	기준	참여	가족	추진	추진	노력	참여	서비스	복지관	장에	복지관	서비스
11	추진	관리	환대	환대	가족	가족	추진	활동	기관	운영	서비스	장에	생활
12	운영	방문	관리	신청	단체	신청	가족	추진	활동	서비스	어렵다	어렵다	장에
13	장에	노인	가족	방문	정책	방문	프로그램	기관	교육	참여	기관	기관	참여
14	생활	참여	기준	관리	환대	확대	노력	실시	생활	가족	참여	서비스	복지관
15	가족	정책	실시	프로그램	실시	실시	정보	교육	방문	방문	교육	노력	노력
16	방문	신청	신청	노인	장에	장에	실시	방문	환정	참여	노력	추진	추진
17	기관	실시	어렵다	정보	정보	노력	확대	신청	관리	복지관	교육	교육	신청
18	관리	안전	실시	실시	방문	관계자	환경	가족	이용	사회복지	관리	종합	정책
19	환대	정보	관계자	정책	지역사회	관리	사용	장에	관리	가족	프로그램	확대	환경
20	신청	강화	사회복지	어렵다	기존	정책	정보	프로그램	도움	추진	방문	환경	확대

주: 1) '장애인'은 모든 달에서 가장 많이 추출된 연관키워드로 위 표에는 제외하였고, '에인' 키워드는 '장애인'과 밀접하게 연관되어 포함시키지 않았음.

2) LUCY 2.0 속성분류체계에 따라 빨간색 음영은 부정, 파란색 음영은 긍정을 나타내는 키워드임.

자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

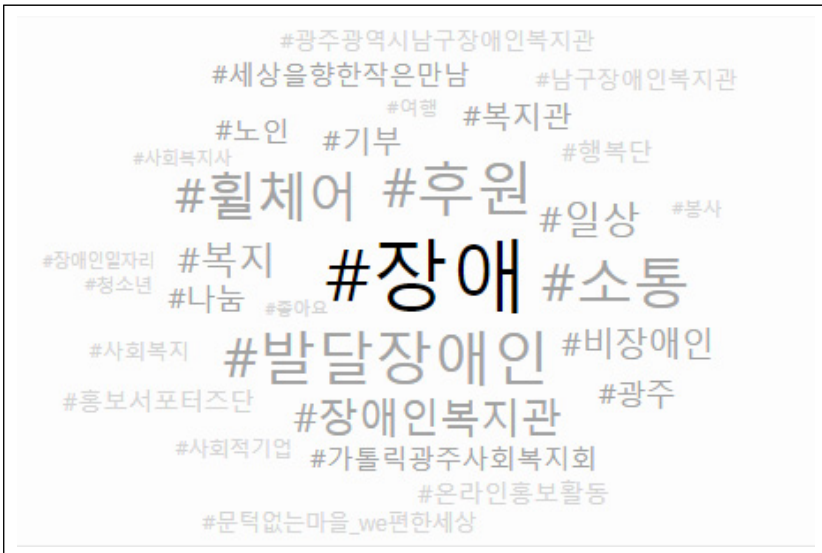
[그림 3-6] 2021년 '장애인 정책' 토픽 매칭 현황



주: LUCY 2.0 토픽검색 분류 방식에 따라 정치, 경제, 사회, 스포츠, 연예, 문화, 생활, IT/과학, 산업, 환경 등 10개의 대항목과 53개의 소항목으로 분류. 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

‘장애인’ 키워드만으로 인스타그램, 트위터, 페이스북, 유튜브, 카카오 스토리, 블로그, 카페 채널을 통한 해시태그 검색을 해본 결과, 언급된 데이터는 전체 19,870건이며, ‘#장애(1,228건)’, ‘#발달장애인(1,016건)’, ‘#후원(1,013건)’ 해시태그가 가장 많이 함께 언급되었다([그림 3-7] 참조).

[그림 3-7] 2021년 ‘장애인’ 정보량 상위 30 해시태그



자료: 소셜 데이터 분석 시스템 LUCY 2.0에서 추출.

### 3. 노인 정책관련 키워드 분석

#### 가. 키워드 설정

보건복지분야 노인 정책에 대한 동향 분석을 하기 위해 ‘노인 정책’ 키워드를 대표 키워드로 선정하였고, 이외 노인 정책과 관련있는 키워드는 보건복지부 홈페이지의 정책메뉴를 활용하여 전문가 회의를 통해 ‘노인 서비스, 노인 보건, 노인 복지, 노인 치매, 노인 실명, 노인 주거, 노인 돌봄, 노인 일자리, 노인 사회활동, 노인 자원봉사, 노인 여가, 장사제도, 노인 장기요양, 노인 의료, 노인 복지시설’ 키워드를 포함하였다.

키워드는 ‘장애인 정책’ 키워드 설정과 동일하게 ‘인접’기능을 사용하여, 키워드별로 5글자 또는 1글자 내에 포함되는 문서를 모두 수집하였고, LUCY 2.0에서 지정한 성인, 도박, 거래, 스포츠, 부동산, 게임, 증권, 광고 등의 사이트의 문서는 제외하였으며, 트위터 채널에서 리트윗(Retweet)은 포함하지 않았다(〈표 3-5〉 참조).

〈표 3-5〉 노인 정책과 관련하여 LUCY 2.0에서 설정된 최종 검색어

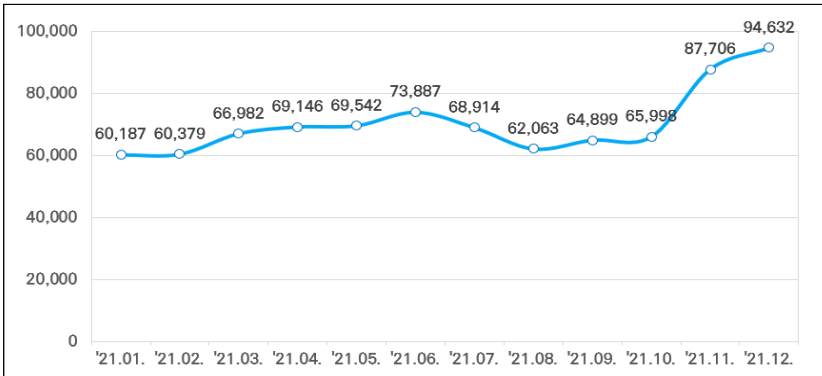
("노인 정책"~5) OR ("노인 서비스"~5) OR ("노인 보건"~5) OR ("노인 복지"~5) OR ("노인 치매"~1) OR ("노인 실명"~1) OR ("노인 주거"~1) OR ("노인 돌봄"~1) OR ("노인 일자리"~1) OR ("노인 사회활동"~1) OR ("노인 자원봉사"~1) OR ("노인 여가"~1) OR ("장사제도"~1) OR ("노인 장기요양"~1) OR ("노인 의료"~1) OR ("노인 복지시설"~1)

## 나. 분석 결과

2021년 1월부터 12월까지 수집된 노인 정책관련 문서량은 총 844,335건이다. 월별 약 6만건 이상의 문서량을 보이다가 2021년 6월 다소 증가한 경향을 보이는데, 이는 6월 초 코로나19로 인한 사회적 거리두기 단계에 따라 2차접종 완료자의 경우 경로당 및 문화센터 등의 이용 가능 영향과 노인 인권을 보호하고 노인학대를 예방하기 위해 제정된 법정 기념일인 6월 15일, '노인학대 예방의 날'<sup>9)</sup> 영향으로 관련 문서량이 증가하였다. 이후 장애인 키워드와 마찬가지로 2022년 대통령선거를 앞두고, 각 정당의 후보자 확정 후 관련 공약 및 선거활동으로 인해 11월을 기점으로 크게 증가하였다([그림 3-8], <표 3-6> 참조).

[그림 3-8] 2021년 월별 '노인 정책' 문서량

(단위: 건)



자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

9) 노인복지법 제6조 4항(노인의 날) 범국민적으로 노인학대에 대한 인식을 높이고 관심을 유도하기 위하여 매년 6월 15일을 노인학대예방의 날로 지정하고, 국가와 지방자치단체는 노인학대예방의 날의 취지에 맞는 행사와 홍보를 실시하도록 노력하여야 한다(국가법령정보센터-노인복지법 <https://www.law.go.kr/법령/노인복지법>).

〈표 3-6〉 2021년 6월 사회적거리두기 완화와 노인학대 관련 뉴스 예

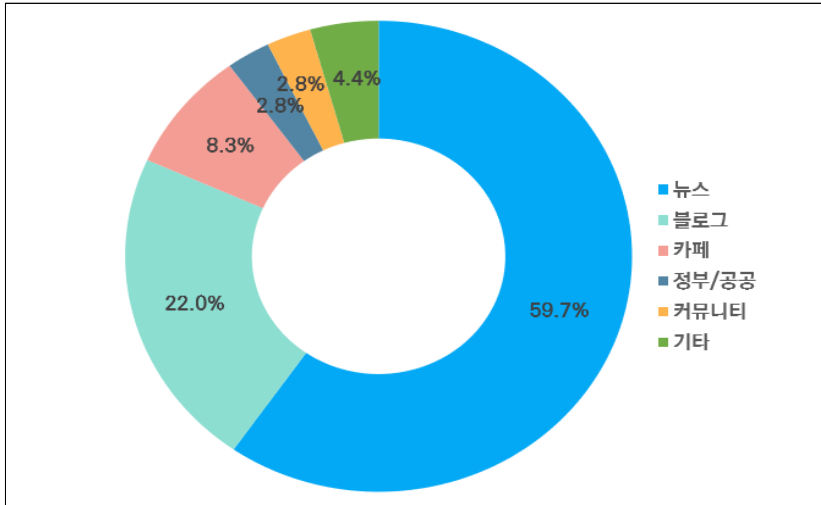
수집일시	제목
2021-06-01 3:10	"못참고 1시간이나 일찍왔수" 어르신 1년 반만에 웃었다 [르포]
2021-06-01 2:29	다시 문 연 경로당·문화센터..'백신 혜택' 본격 시작
2021-06-07 3:39	울산 노인시설 850여 곳 한 달 여 만에 운영 재개
2021-06-15 3:21	노인학대, 집콕으로 급증... 가해자는 아들이 1위
2021-06-15 1:27	명 자국 있어도 속앓이만..노인학대 증가세
2021-06-15 8:32	[인터뷰] "조사받느니 과태료 낸다"..요양시설 노인학대, 왜 해결 어렵나

자료: 소셜 데이터 분석 시스템 LUCY 2.0에서 추출.

844,335건 문서량에 대한 채널별 문서비율을 살펴보면 뉴스가 59.7%로 가장 많았고, 블로그 22.0%, 카페 8.3% 순으로 나타났다([그림 3-9] 참조).

[그림 3-9] 2021년 채널별 '노인 정책' 문서 비율

(단위: %)



주: 기타에는 트위터, 인스타그램, 유튜브, 페이스북, 카카오토리, 지식인, 기업/단체가 포함됨.  
 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.



LUCY 2.0의 탑재된 연관어사전으로 추출한 문서 내 언급 키워드를 연관키워드라 정의할 때, 정보량(연관키워드가 한 번 이상 언급된 문서 수량) 중 가장 많이 차지한 연관키워드는 ‘노인’으로 541,855건으로 나타났다. ‘노인’을 제외한 상위 20개 키워드를 살펴 보면, ‘어르신, 사업, 센터, 지역, 사회, 서비스, 복지, 참여, 활동, 추진, 어렵다, 관리, 운영, 생활, 기관, 방문, 돌봄, 일자리, 안전, 프로그램’순으로 나타났다([그림 3-10], <표 3-7> 참조).

[그림 3-10] 2021년 ‘노인 정책’ 정보량 상위 100개 연관 키워드



자료: 소셜 데이터 분석 시스템 LUCY 2.0 분석결과를 보건복지데이터포털(data.kihasa.re.kr) 소셜데이터 '사용자 시각 Tool' 사용하여 저자 작성.

노인 정책 월별 정보량 상위 20개 연관 키워드 중 부정과 관련된 키워드는 ‘어렵다’가 추출되었고, 긍정과 관련된 키워드는 ‘노력’과 ‘행복’이 추출되었다(<표 3-7> 참조).

토픽 검색 결과 전체 844,335 건 중 836,620건(99%)이 LUCY 2.0에

서 설정한 토픽<sup>10)</sup>과 매칭되었고, 사회>보건복지(405,453건), 생활>의료(106,986건), 사회>교육(62,924건) 토픽 관련 언급이 가장 많았다([그림 3-11] 참조).

---

10) 사회>보건복지 토픽은 보건복지부, 기초생활과 같은 보건복지 법·종류·기관명 관련 정보 일 경우 분류되고, 생활>의료 토픽은 국민건강보험공단, 금연과 같은 질병예방법, 생활 양상, 건강 관련 기관, 질병명 관련 정보 일 경우 분류되고, 사회>교육 토픽은 교육 담당 기고나명, 교육의 분류, 국가 교육방침, 교육비, 교육의 문제점관련 정보일 경우 분류됨.

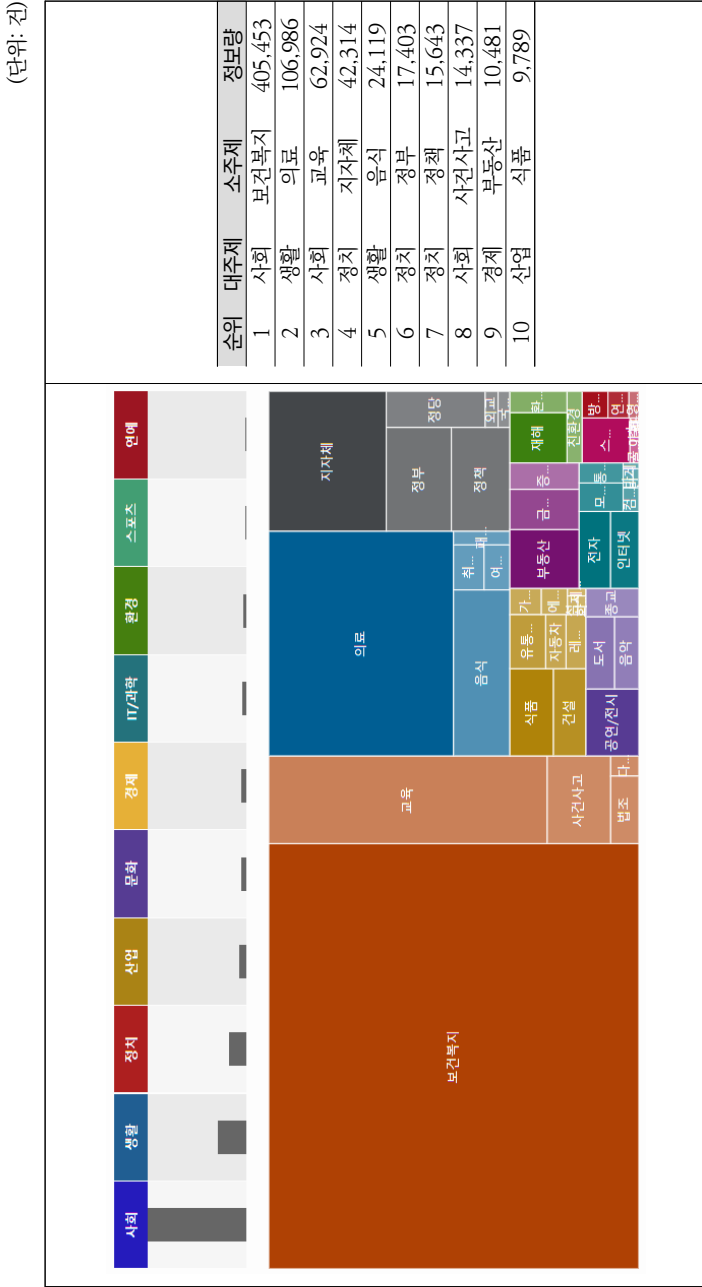
〈표 3-7〉 2021년 월별 '노인 정책' 정보량 상위 20개 연관 키워드

순위	전체	21.1.	21.2.	21.3.	21.4.	21.5.	21.6.	21.7.	21.8.	21.9.	21.10.	21.11.	21.12.
1	어르신	사업	센터	어르신	어르신	어르신	어르신	어르신	어르신	어르신	어르신	어르신	사업
2	사업	센터	지역	센터	사업	사업	사업	센터	사업	지역	사회	지역	어르신
3	센터	지역	어르신	사업	사업	지역	센터	사업	지역	사업	지역	사업	지역
4	지역	서비스	사업	지역	지역	센터	지역	지역	센터	센터	사업	사회	사회
5	사회	어르신	사회	서비스	서비스	서비스	서비스	사회	사회	사회	사업	센터	센터
6	서비스	사회	서비스	사회	사회	사회	사회	사회	활동	활동	활동	활동	활동
7	복지	복지	방문	복지	어렵다	복지	복지	복지	서비스	복지	운영	운영	서비스
8	참여	추진	복지	추진	복지	어렵다	참여	어렵다	생활	생활	서비스	복지	기관
9	활동	관리	추진	관리	돌봄	참여	관리	운영	운영	서비스	복지	복지	복지
10	추진	안전	관리	참여	참여	추진	프로그램	활동	기관	운영	생활	생활	운영
11	어렵다	확대	참여	돌봄	추진	돌봄	어렵다	관리	복지	기관	참여	참여	생활
12	관리	방문	안전	방문	관리	복지관	추진	돌봄	어렵다	어렵다	참여	기관	어렵다
13	운영	참여	실시	프로그램	방문	프로그램	실시	추진	관리	참여	어렵다	추진	참여
14	생활	일자리	장애인	안전	실시	방문	안전	참여	참여	추진	교육	어렵다	추진
15	기관	장애인	장애인	확대	장애인	가족	돌봄	생활	추진	추석	우리	교육	건강
16	방문	대응	대응	실시	안전	안전	노력	안전	방문	방문	프로그램	관리	일자리
17	돌봄	정책	돌봄	일자리	프로그램	관리	일자리	방문	일자리	노력	관리	일자리	관리
18	일자리	실시	관계자	장애인	복지관	노력	방문	교육	교육	종합	추진	노력	노력
19	안전	기준	제가	보호	일자리	실시	가족	프로그램	돌봄	관리	노력	돌봄	돌봄
20	프로그램	시행	접종	가족	보호	행복	보호	보호	도움	교육	행복	프로그램	교육

주: LUCY 2.0 속성분류체계에 따라 빨간색 음영은 부정, 파란색 음영은 긍정을 나타내는 키워드임.

자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

[그림 3-11] 2021년 '노인 정책' 토픽 매핑 현황



주: LUCY 2.0 토픽검색 분류 방식에 따라 정치, 경제, 생활, IT/과학, 산업, 환경 등 10개의 대항목과 53개의 소항목으로 분류. 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.



## 4. 아동 정책관련 키워드 분석

### 가. 키워드 설정

보건복지분야 아동 정책에 대한 동향 분석을 하기 위해 ‘아동 정책’ 키워드를 대표 키워드로 선정하였고, 이외 아동 정책과 관련있는 키워드는 보건복지부 홈페이지의 정책메뉴를 활용하여 전문가 회의를 통해 ‘아동 서비스, 아동 보건, 아동 복지, 아동 수당, 아동 입양, 아동 인권, 요보호 아동, 가정위탁 지원, 드림스타트, 아동 자립수당, 아동 교사, 공동생활 가정, 디딤씨앗통장, 아동 학대, 피해 아동, 아동 안전, 실종 아동’ 키워드를 포함하였다.

키워드는 ‘장애인 정책’ 키워드 설정과 동일하게 ‘인접’기능을 사용하여, 키워드별로 5글자 또는 1글자 내에 포함되는 문서를 모두 수집하였고, LUCY 2.0에서 지정한 성인, 도박, 거래, 스포츠, 부동산, 게임, 증권, 광고 등의 사이트의 문서는 제외하였으며, 트위터 채널에서 리트윗(Retweet)은 포함하지 않았다(〈표 3-8〉 참조).

〈표 3-8〉 아동 정책과 관련하여 LUCY 2.0에서 설정된 최종 검색어

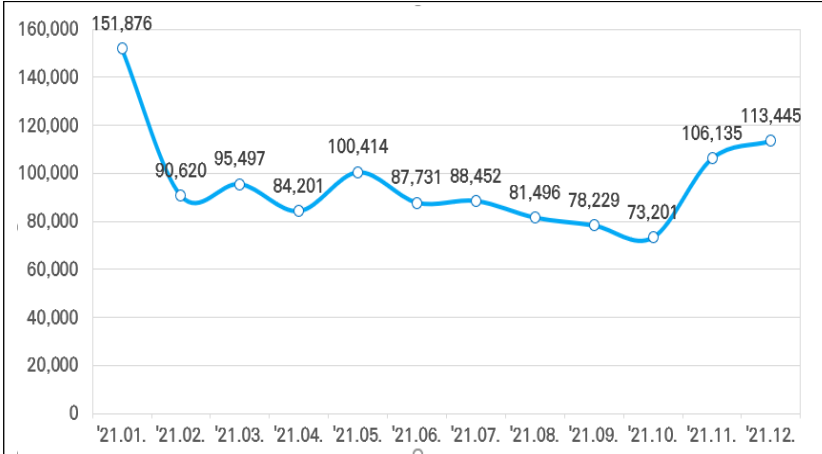
((“아동 정책”~5) OR (“아동 서비스”~5) OR (“아동 보건”~5) OR (“아동 복지”~5) OR (“아동 수당”~1) OR (“아동 입양”~1) OR (“아동 인권”~1) OR (“요보호 아동”~1) OR (“가정위탁 지원”~1) OR (“드림스타트”~1) OR (“아동 자립수당”~1) OR (“아동 교사”~1) OR (“공동생활가정”~1) OR (“디딤씨앗통장”~1) OR (“아동 학대”~1) OR (“피해 아동”~1) OR (“아동 안전”~1) OR (“실종 아동”~1))

## 나. 분석 결과

2021년 1월부터 12월까지 수집된 아동 정책관련 문서량은 총 1,151,297건이다. 전체 문서 중 1월에 151,876건으로 가장 많은 문서량이 추출되었는데, 이는 2020년 양부모 학대로 생후 16개월에 사망한 '정인이 사건'과 2021년 1월 초 3살 여자아이가 내복차림으로 길거리에서 발견된 사건의 영향으로 보인다. 이후 장애인 키워드와 마찬가지로 2022년 대통령선거를 앞두고, 각 정당의 후보자 확정 후 관련 공약 및 선거활동으로 인해 11월을 기점으로 증가하였다([그림 3-13], <표 3-9> 참조).

[그림 3-13] 2021년 월별 '아동 정책' 문서량

(단위: 건)



자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

70 2021년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

〈표 3-9〉 2021년 1월 아동학대 관련 뉴스 예

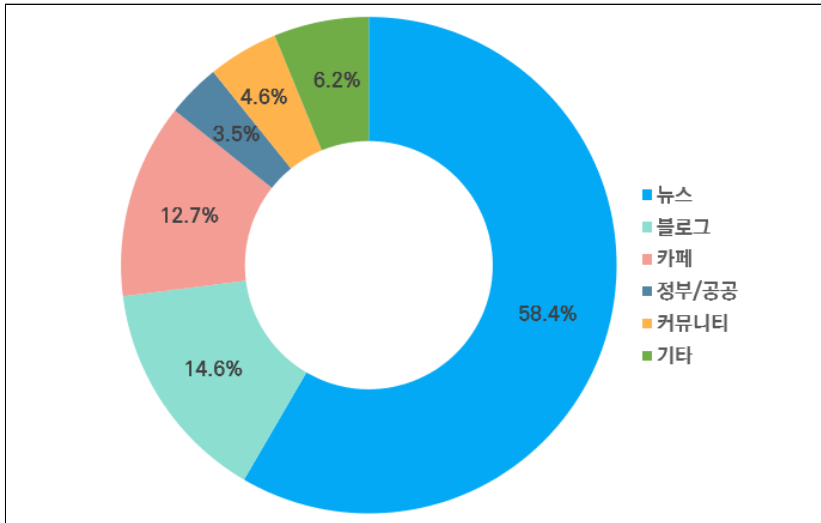
수집일시	제목
2021-01-05 3:58	'정인이 사건' 못매 맞는 경찰...靑국민청원 20만 돌파(종합2보)
2021-01-05 3:53	정인이 숨지기 열흘 전..입양기관 홀트 "잘지내고 있다" 기록(종합)
2021-01-09 3:44	'북극한파'에 내복차림 3세 여아 "도와주세요"
2021-01-09 2:14	"도와주세요" 영하 19도 한파에 내복입고 발견된 3살(종합)
2021-01-16 2:46	[단독] 학대 아동 의료 지원 '유명무실'..전담기관 없었다
2021-01-16 0:50	중학생 아들 때린 아버지 입건..가족과 분리 조치

자료: 소셜 데이터 분석 시스템 LUCY 2.0에서 추출.

1,151,297건의 문서량에 대한 채널별 문서비율을 살펴보면 뉴스가 58.4%로 가장 많았고, 블로그 14.6%, 카페 12.7% 순으로 나타났다(그림 3-14) 참조).

[그림 3-14] 2021년 채널별 '아동 정책' 문서 비율

(단위: %)

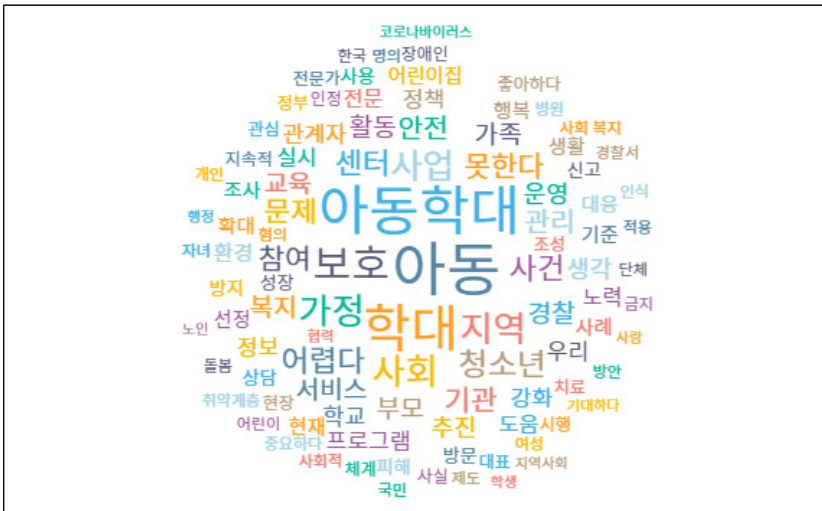


주: 기타에는 트위터, 인스타그램, 유튜브, 페이스북, 카카오토리, 지식인, 기업/단체가 포함됨.  
 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.



LUCY 2.0의 탑재된 연관어사전으로 추출한 문서 내 언급 키워드를 연관키워드라 정의할 때, 정보량(연관키워드가 한 번 이상 언급된 문서 수량) 중 가장 많이 차지한 연관키워드는 ‘아동’이 465,039건, ‘아동학대’가 427,343건 순이었다. ‘아동’을 제외한 상위 20개 키워드를 살펴 보면, ‘아동학대, 학대, 보호, 지역, 사회, 가정, 사업, 청소년, 사건, 어렵다, 센터, 못한다, 참여, 기관, 문제, 안전, 관리, 부모, 활동, 서비스’ 순으로 많이 나타났다([그림 3-15], <표 3-10> 참조).

[그림 3-15] 2021년 ‘아동 정책’ 정보량 상위 100개 연관 키워드



자료: 소셜 데이터 분석 시스템 LUCY 2.0 분석결과를 보건복지데이터포털(data.kihasa.re.kr) 소셜데이터 ‘사용자 시각 Tool’ 사용하여 저자 작성.

아동 정책 월별 정보량 상위 20개 연관 키워드 중 부정과 관련된 키워드는 ‘아동학대’, ‘학대’, ‘어렵다’, ‘못한다’가 추출되었고, 긍정과 관련된 키워드는 ‘노력’과 ‘강화’가 추출되었다(〈표 3-10〉 참조).

토픽 검색 결과 전체 1,151,297 건 중 1,140,805건 (99%)이 LUCY 2.0에서 설정한 토픽<sup>11)</sup>과 매칭되었고, 사회>사건사고 (305,053건), 사회>보건복지 (277,370건), 사회>교육 (178,799건) 토픽 관련 언급이 가장 많았다(〈그림 3-16〉 참조).

---

11) 사회>사건사고는 화재 등 사건사고 관련 정보일 경우 분류되고, 사회>보건복지 토픽은 보건복지부, 기초생활과 같은 보건복지 법·종류·기관명 관련 정보 일 경우 분류되고, 사회>교육 토픽은 교육 담당 기고나명, 교육의 분류, 국가 교육방침, 교육비, 교육의 문제점 관련 정보일 경우 분류됨.

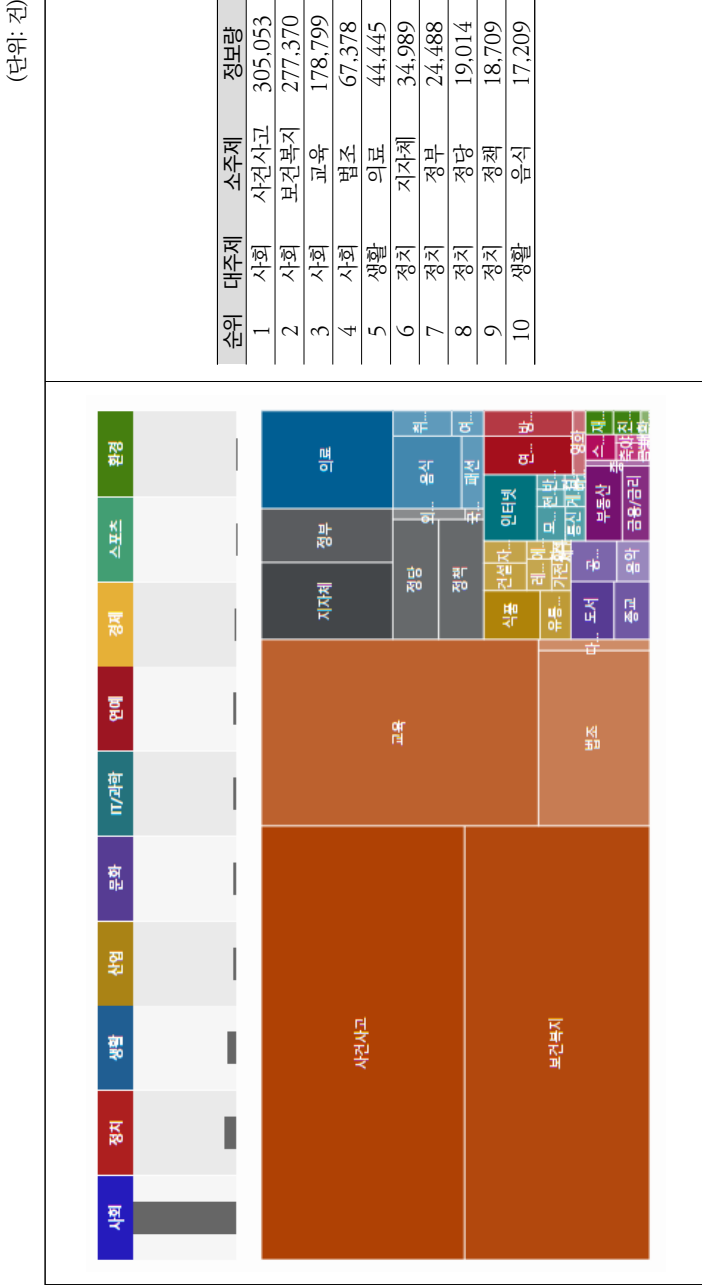
〈표 3-10〉 2021년 월별 '이동 정책' 정보량 상위 20개 연관 키워드

순위	전체	21.1.	21.2.	21.3.	21.4.	21.5.	21.6.	21.7.	21.8.	21.9.	21.10.	21.11.	21.12.
1	이동학대	이동학대	이동학대	이동학대	이동학대	이동학대	이동학대	학대	학대	학대	학대	학대	지역
2	학대	사건	학대	학대	학대	학대	학대	이동학대	이동학대	이동학대	이동학대	이동학대	사업
3	보호	학대	보호	보호	보호	보호	보호	청소년	활동	활동	보호	보호	사회
4	지역	보호	사건	경찰	가정	사회	지역	보호	보호	지역	기관	기관	활동
5	사회	못한다	경찰	사건	사업	지역	사회	지역	교육	보호	활동	지역	학대
6	가정	학대하다	못한다	가정	지역	가정	사업	사회	기관	사회	교육	활동	기관
7	사업	경찰	사회	지역	사회	어렵다	가정	사업	운영	기관	지역	사회	가정
8	청소년	부모	지역	사업	참여	못한다	어렵다	가정	청소년	가정	운영	교육	어렵다
9	사건	정인이	가정	사회	어렵다	참여	청소년	문제	사회	운영	사회	참여	운영
10	어렵다	양부모	문제	못한다	사건	사업	센터	센터	지역	청소년	가정	운영	보호
11	센터	가정	사업	청소년	센터	청소년	참여	기관	센터	어렵다	청소년	사업	센터
12	못한다	문제	부모	센터	못한다	가족	복지	활동	가정	교육	사업	센터	이동학대
13	참여	임양	청소년	문제	청소년	사건	서비스	관리	사업	사업	센터	가정	교육
14	기관	생각	생각	안진	추진	부모	안진	안진	어렵다	센터	참여	청소년	청소년
15	문제	사회	관계자	참여	서비스	센터	프로그램	서비스	문제	우리	문제	어렵다	복지
16	안전	방지	가족	가족	안전	생각	관리	어렵다	관리	참여	어렵다	안전	추진
17	관리	강화	학대하다	추진	복지	안전	못한다	교육	가족	문제	관리	노력	참여
18	부모	대응	복지	어린이집	관리	서비스	부모	참여	조사	관리	안전	추진	긴급
19	활동	관리	서비스	서비스	가족	노력	가족	학교	정보	추진	프로그램	복지	생활
20	서비스	안전	안전	관리	프로그램	경찰	노력	부모	서비스	생활	복지	우리	서비스

주: LUCY 2.0 속성분류체계에 따라 빨간색 음영은 부정, 파란색 음영은 긍정을 나타내는 키워드임.

자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

[그림 3-16] 2021년 '아동 정책 토픽 매핑 현황



주: LUCY 2.0 토픽검색 분류 방식에 따라 정치, 경제, 사회, 스포츠, 연예, 문화, 생활, IT/과학, 산업, 환경 등 10개의 대항목과 53개의 소항목으로 분류 자료. 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

## 5. 건강 보장 관련 키워드 분석

### 가. 키워드 설정

건강보험 보장성 강화는 가계의 병원비 부담을 낮추기 위한 국민 의료비 부담 완화 정책으로 문재인 대통령이 2017년 8월 9일 서울성모병원을 방문해 2022년까지 전 국민의 의료비 부담을 평균 18% 낮추는 '건강보험 보장성 강화 정책'을 발표하면서 시작되었다(대한민국정책브리핑 중 '한눈에 보는 정책', 2021.8.20.).

이번 절에서는 건강 보장에 대한 동향분석을 하기 위해 '건강 보장' 키워드를 대표 키워드로 선정하였고, 이외 관련있는 '문재인 케어'를 키워드에 포함하였다.

키워드 검색방식은 '장애인 정책' 키워드 방식과 동일하게 '인접'기능을 사용하여, 각 키워드가 5글자 내에 포함되는 문서를 모두 수집하였고, LUCY 2.0에서 지정한 성인, 도박, 거래, 스포츠, 부동산, 게임, 증권, 광고 등의 사이트의 문서는 제외하였으며 민간보험의 키워드인 '생명 보험', '손해 보험'도 제외하였다. 또한 트위터 채널에서 리트윗(Retweet)은 포함하지 않았다(〈표 3-11〉 참조).

〈표 3-11〉 건강 보장과 관련하여 LUCY 2.0에서 설정된 최종 검색어

```
((("건강 보장"~5) OR ("건강보장"~5) OR ("문재인 케어"~5) OR ("문재인케어"~5)))
NOT (("생명" AND "보험") OR ("손해" AND "보험"))
```

## 나. 분석 결과

2017년 1월부터 12월까지 수집된 건강 보장 키워드 문서량은 122,496건, 2018년은 112,411건, 2019년 106,220건, 2020년은 86,760건, 마지막으로 2021년은 95,515건이다. [그림 3-17]을 살펴보면 2017년 8월 9일 '건강보험 보장성 강화 정책'을 발표한 시점을 포함하여 총 7번 문서량이 급증한 것을 확인할 수 있다.

문서량이 급증한 기간의 주요 이슈를 살펴보면 첫 번째는 전체기간에 가장 많은 문서량을 차지한 2017년 8월(30,920건)이다. 이는 앞서 언급한대로 건강보험 보장성 강화 정책이 발표되면서 이에 관한 기사의 예시는 아래와 같다(〈표 3-12〉 참조).

〈표 3-12〉 건강보험 보장성 강화 정책 발표 기사

초음파, 자기공명영상장치(MRI), 로봇수술, 2인실 등 그간 건강보험의 적용을 받지 못했던 3천800여개의 비급여 진료항목들이 단계별로 보험급여를 받게 된다. 문재인 정부는 여기에 2022년까지 31조원을 투입할 계획이다. 이를 통해 국민 비급여 의료비 부담을 2015년 13조5천억원에서 2022년 4조8천억원으로 64% 낮춘다는 목표를 세웠다. 국민 의료비 상승의 주범으로 지목되는 비급여 진료를 획기적으로 줄이고자 이른바 '문재인 케어'에 본격 시동을 거는 것이다....

자료: 서한기(2017.8.19.) '문재인 케어' 시동...미용·성형 빼고 모든 치료에 건보 적용. 연합뉴스. <https://www.yna.co.kr/view/AKR20170725140100017?input=1195m>

두 번째는 2017년 12월(21,138건)로, 문재인 케어 시행으로 대한의사협회는 대형 병원에 환자 쏠림현상이 심해지고 건보 재정이 고갈될 것으로 주장하며 대규모 반대 집회를 개최하여 문서량이 급증하였다. 반대집회와 관련한 기사의 예시는 다음과 같다(〈표 3-13〉 참조).

## 〈표 3-13〉 건강보험 보장성 강화 반대 관련 기사1

...건강보험 보장성을 높이는 이른바 '문재인 케어'에 반대하는 의사들이 지난 주말 거리로 나섰습니다. '문재인 케어' 시행으로 대형 병원에 환자가 몰려 동네 병원은 망하고, 건보 재정도 파탄날 거라는 주장입니다. 이에 대해 문재인 대통령은 의료계의 염려는 이해한다면서도 의학적으로 필요한 모든 진료를 건강보험으로 해결하는 것이 '문재인 케어'의 취지라고 설명했습니다....

자료: 김도원(2017.12.11.) '문재인 대통령 "의료계가 '문재인 케어' 지혜 모아달라'. YTN.  
<https://news.naver.com/main/read.naver?mode=LSD&mid=sec&sid1=001&oid=052&aid=0001092882>

세 번째는 2018년 4월(14,144건)으로 상복부 초음파비용의 건강보험 적용을 두고 대한의사협회에서는 문재인케어 본격 시행으로 보고 집단행동(대규모 시위)을 예고하면서 정부와의 갈등을 야기했다. 이와 관련된 기사의 예시는 아래와 같다(〈표 3-14〉 참조).

## 〈표 3-14〉 건강보험 보장성 강화 반대 관련 기사2

...전국 13만 명 의사들의 권익을 대변하는 법정단체인 대한의사협회는 정부가 상복부 초음파 건강보험 적용 강행을 예고하자 지난 달 30일 '문재인 케어와의 전쟁'을 선포했다. 의사협회는 집단휴진(의료파업) 등 단체행동에 나선다는 방침이다. ....

자료: 이연호(2018.4.1.) '醫-政, 문케어 정면 충돌...의협 4월중 의료파업 예고. 이데일리.  
<https://news.naver.com/main/read.naver?mode=LSD&mid=sec&sid1=001&oid=018&aid=0004068757>

네 번째는 2018년 10월(9,245건)로 국회 보건복지위원회의 보건복지부 질병관리본부 등의 국정감사에서 국민연금과 문재인케어와 관련하여 여당과 야당에 대립에 관한 이슈로 문서량이 급증했다. 이와 관련한 기사

의 예시는 다음과 같다(〈표 3-15〉 참조).

〈표 3-15〉 건강보험 보장성 강화 관련 국정감사 이슈 기사

...자유한국당을 중심으로 한 일부 야당은 국민연금 소득대체율을 올릴 경우 재정고갈 시점이 앞당겨지고 보험료 인상이 불가피하며, '문재인 케어'로 결국 국민 부담이 가중될 것이라며 공세에 나섰다. 이에 여당인 더불어민주당은 국민연금 지급보장 시 국민에게 보험료 인상을 설득할 수 있고, '새 정부 출범 후 가장 호감도 높은 정책이 문재인 케어'라며 정부정책을 옹호했다. ....

자료: 이슬기(2018.10.10.) '복지위, 국민연금·文케어 놓고 대립..野 '강공'·與 '엄호'(종합). 연합뉴스. <https://news.v.daum.net/v/20181010200634153>

다섯번째는 2019년 7월(14,088건)로 건강보험 보장성 강화대책 시행 2주년을 맞이하여 주요 성과를 발표했다. 이와 관련한 기사의 예시는 다음과 같다(〈표 3-16〉 참조).

〈표 3-16〉 건강보험 보장성 강화 성과 관련 기사

지난 2017년부터 추진된 건강보험 보장성을 강화하는 '문재인 케어' 시행 2년 동안 우리 국민이 약 2조 2천 억 원의 의료비 경감 혜택을 봤다는 정부 설명이 나왔습니다. 보건복지부는 오늘 '건강보험 보장성 강화대책' 시행 2주년을 맞아 주요 성과를 발표하며 이같이 밝혔습니다. 정부는 앞서 지난 2017년 8월, 의학적으로 필요한 비급여 진료에 대해 건강보험을 적용해 급여화하고 노인과 아동 등 취약계층의 의료비를 대폭 낮추는 '문재인 케어' 대책을 발표한 후 선택진료비를 폐지하고 MRI와 초음파 급여화 등을 단계적으로 실시해왔습니다.

이에 따라 복지부는 그동안 환자가 전액 부담하던 비급여진료 검사 등에 대해 건강보험을 적용해 1조 4천억 원이 경감됐고, 노인과 아동 등 의료취약계층의 본인 부담금 인하로 8천 억 원이 경감되는 등 지난 2년 동안 총 2조 2천억 원의 의료



*비 경감혜택을 본 것으로 집계됐다고 밝혔습니다....*

자료: 박진주(2019.7.2.) '문케어' 2년..국민 의료비 2조 2천억 경감 혜택 봐. MBC.  
<https://news.v.daum.net/v/20190702194605860>

여섯번째는 2019년 12월(10,066건)로 국민건강보험공단에서 2018년도 건강보험환자 진료비실태조사 결과를 발표했고, 건강보험 보장률이 63.8%로 전년보다 1.1%p 상승한 것으로 나타났다. 이와 관련된 기사는 다음과 같다(〈표 3-17〉 참조).

〈표 3-17〉 건강보험 보장률 상승 관련 기사

*... 국민건강보험공단은 16일 이 같은 내용을 담은 '2018년도 건강보험환자 진료비 실태조사' 결과를 발표했다. 결과에 따르면 백혈병과 혈액암 등 중증·고액 진료비 상위 30위 내 질환 보장률은 81.2%로 역대 최대를 기록했다. 또한, 아동·노인 의료비에도 건강보험 재정을 집중 투입해 5세 이하 보장률은 전년보다 2.5%p 상승한 69.3%를 기록했고, 65세 이상 보장률은 1.4%p 상승한 70.2%로 나타났다. ...*

자료: 장지민(2019.12.16.) 건강보험 보장률 63.8%..전년대비 1.1%p ↑ '문케어 2년차 성적표'. 한국경제. <https://news.v.daum.net/v/20191216220117049>

마지막으로 일곱 번째는 2021년 8월(13,083건)로 8월 12일의 문서량이 3,231건으로 '건강보험 보장성 강화대책 4주년 성과보고대회'의 영향인 것으로 분석된다. 이와 관련된 기사는 다음과 같다(〈표 3-18〉 참조).

〈표 3-18〉 건강보험 보장성 강화관련 성과대회 관련 기사

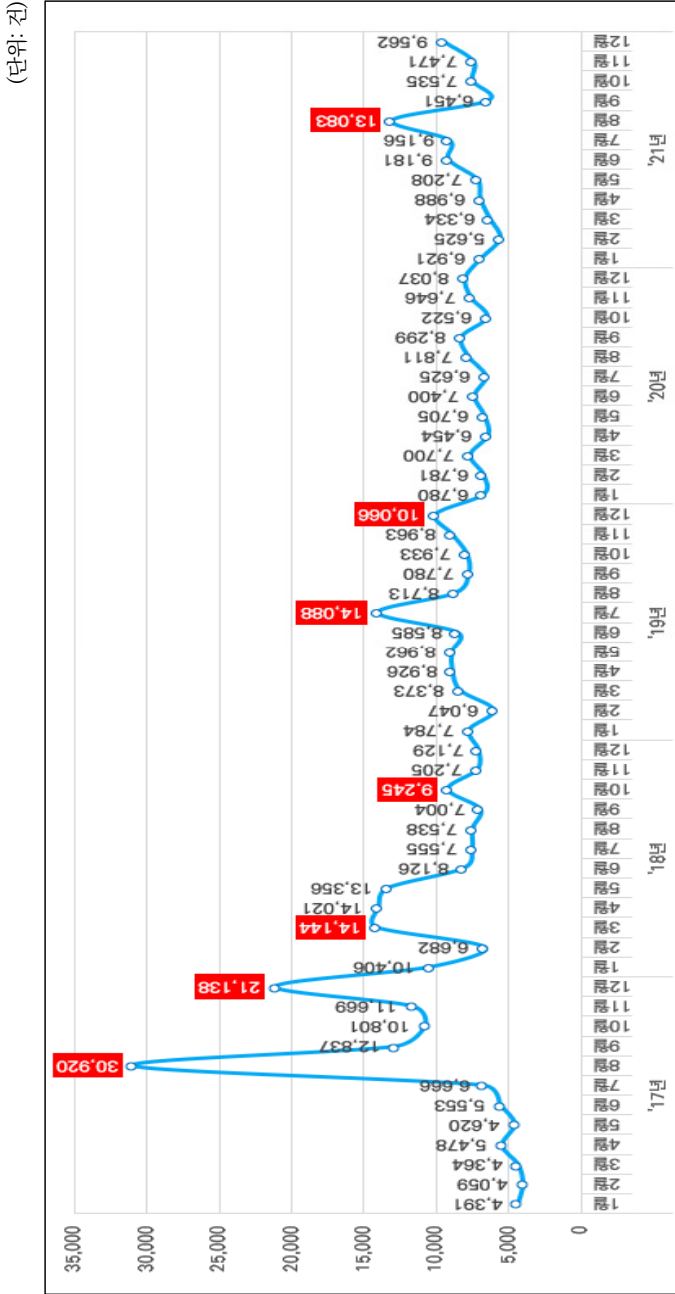
...문 대통령은 모두발언에서 '병원비 걱정 없는 든든한 나라'를 만들어 의료비로 인한 가계 파탄을 최소화하고자 노력했다며, 그 결과 고액·중증질환자, 건강 취약계층, 저소득층 중심으로 의료비를 경감할 수 있었다고 했다. 이어 코로나19 위기 상황에서도 건강보험이 보건의료의 버팀목으로 자리잡고 있었기에 예방, 진단·검사, 치료 및 각종 의료현장 지원 측면에서 신속하고 적절한 방역 대응이 가능했다고 말했다. ...

모두발언에 이어 김용익 국민건강보험공단 이사장은 건강보험 보장성 강화대책 성과를 보고했다. 김 이사장은 2018~2020년 3700만명의 국민이 약9조2000억원의 가계 의료비 부담을 덜었다고 보고했다.....

자료: 뉴시스(2021.8.12.) 문 대통령, '문케어 4주년' 성과보고...“병원비 걱정 없는 나라”. 동아일보.

<https://www.donga.com/news/Politics/article/all/20210812/108519812/1>

[그림 3-17] 2017년~2021년 월별 '건강 보장' 문서량

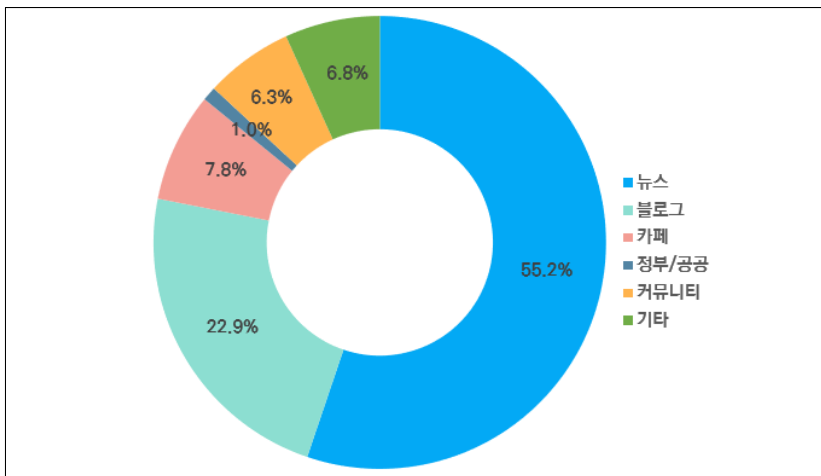


주: 문서량이 증가한 시기를 빨간색으로 표기함.  
 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

2021년 95,515건의 문서량의 채널별 문서비율을 살펴보면 뉴스가 55.2%로 가장 많았고, 블로그 22.9%, 카페 7.8% 순으로 나타났으며, 이 전년도도 유사한 패턴으로 추출되었다([그림 3-18] 참조).

[그림 3-18] 2021년 전체기간의 채널별 '건강 보장' 문서 비율

(단위: %)



주: 기타에는 트위터, 인스타그램, 유튜브, 페이스북, 카카오토리, 지식인, 기업/단체가 포함됨.  
자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

LUCY 2.0의 탑재된 연관어사전으로 추출한 문서 내 언급 키워드를 연관키워드라 정의할 때, 정보량(연관키워드가 한 번 이상 언급된 문서 수량)이 가장 많이 차지한 연관키워드는 '보장'이 2017년 54,824건, 2018년 51,740건, 2019년 64,712건, 2020년 65,591건, 2021년 74,994건으로 가장 많이 나왔다([그림 3-19] 참조).



‘보장’을 제외한 상위 20개 키워드를 살펴 보면, 정책 추진 첫해인 2017년과 2018년은 ‘건강보험, 문재인, 정책, 정부, 의료’ 등이 상위권에 추출되었고, 2019년 이후부터는 ‘정책, 강화, 관리, 사업, 확대’ 등이 정보량이 상위권에 나타났다(〈표 3-19〉 참조).

〈표 3-19〉 2017~2021년 ‘건강 보장’ 정보량 상위 20개 연관 키워드

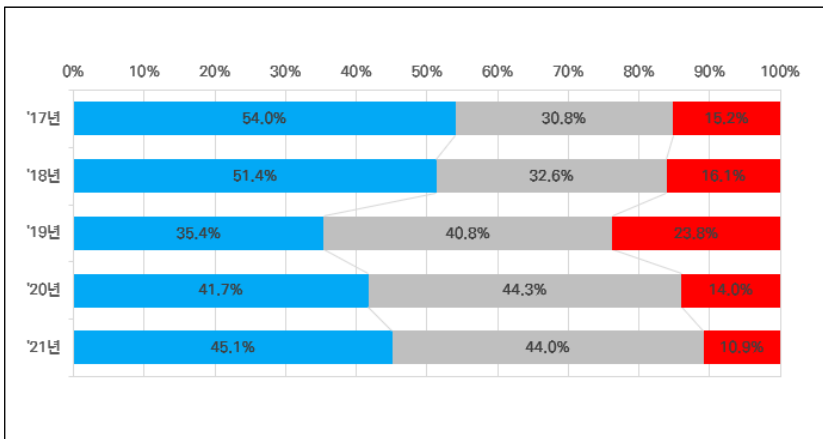
순위	‘17년	‘18년	‘19년	‘20년	‘21년
1	건강보험	문재인	건강	정책	정책
2	문재인	건강보험	강화	강화	관리
3	정책	정책	건강보험	정부	강화
4	정부	의료	정책	관리	사업
5	의료	정부	정부	의료	확대
6	추진	추진	확대	확대	추진
7	대책	병원	문재인	건강보험	지역
8	국민	적용	케어	지역	건강보험
9	병원	보건복지부	의료	추진	정부
10	개선	대책	부담	사회	사회
11	적용	개선	추진	못한다	못한다
12	치료	의사	관리	사업	의료
13	의사	국민	우리	문제	어렵다
14	보건복지부	치료	사회	건강	부담
15	인상	진료	병원	기준	기준
16	진료	의원	못한다	병원	치료
17	의원	비용	국민	안전	우리
18	비용	참여	기준	치료	시행
19	증가	논의	문제	어렵다	문제
20	국회	증가	높다	국민	국민

주: 1) ‘보장’은 모든 연도에서 가장 많이 추출된 연관키워드로 위 표에는 제외하였음.  
 2) LUCY 2.0 속성분류체계에 따라 빨간색 음영은 부정, 파란색 음영은 긍정을 나타내는 키워드임.  
 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

감성분석은 LUCY 2.0의 탑재된 감성사전으로 긍정, 중립, 부정으로 분류되며, 긍정은 문서 내 언급된 ‘긍정 패턴 수량’이 ‘부정 패턴 수량’보다 많은 경우, 중립은 문서 내 언급된 긍정 패턴 수량과 부정 패턴 수량이 동일한 경우, 부정은 문서내 언급된 부정 패턴 수량이 긍정 패턴 수량보다 많은 경우 분류된다. 2017년에서 2021년까지 감성분석을 수행한 결과, 2017년과 2018년은 긍정 정보량의 점유율이 50% 이상이었고, 2019년에 35.4%로 낮아졌다가 점점 높아지는 경향성을 보인다([그림 3-20] 참조).

[그림 3-20] 2017년~2021년 ‘건강 보장’ 감성분석

(단위: %)



자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

2021년의 정보량(연관키워드가 한 번 이상 언급된 문서 수량)이 가장 많이 차지한 상위 20개 연관 키워드를 살펴 보면, ‘보장’을 제외하고, ‘정책, 관리, 강화, 사업, 확대, 추진, 지역, 건강보험, 정부, 사회, 못한다, 의료, 어렵다, 부담, 기준, 치료, 우리, 시행, 문제, 국민’순으로 많이 나타났

다. 월별 키워드를 살펴보면 상반기에는 ‘못한다’, ‘어렵다’, ‘부담’ 등의 부정적 키워드가 상위권에 속한 반면에 하반기에는 ‘강화’, ‘건강한’, ‘노력’ 등의 긍정적 키워드가 상위권에 속한 것으로 나타났다(〈표 3-20〉 참조).

토픽 검색 결과 데이터 전체 95,515 건 중 93,469건 (97.8%) LUCY 2.0에서 설정한 토픽<sup>12)</sup>과 매칭되었고, 생활〉의료 (28,759건), 사회〉보건복지 (17,567건), 정치〉정부 (7,709건) 순으로 언급되었다(〈그림 3-21〉 참조).

---

12) 생활〉의료 토픽은 국민건강보험공단, 금연과 같은 질병예방법, 생활 양상, 건강 관련 기관, 질병명 관련 정보 일 경우 분류되며, 사회〉보건복지 토픽은 보건복지부, 기초생활과 같은 보건복지 법·종류·기관명 관련 정보 일 경우 분류되고, 정치〉정부는 중앙부처, 각 산하기관 및 위원회 등과 관련된 정보가 분류됨

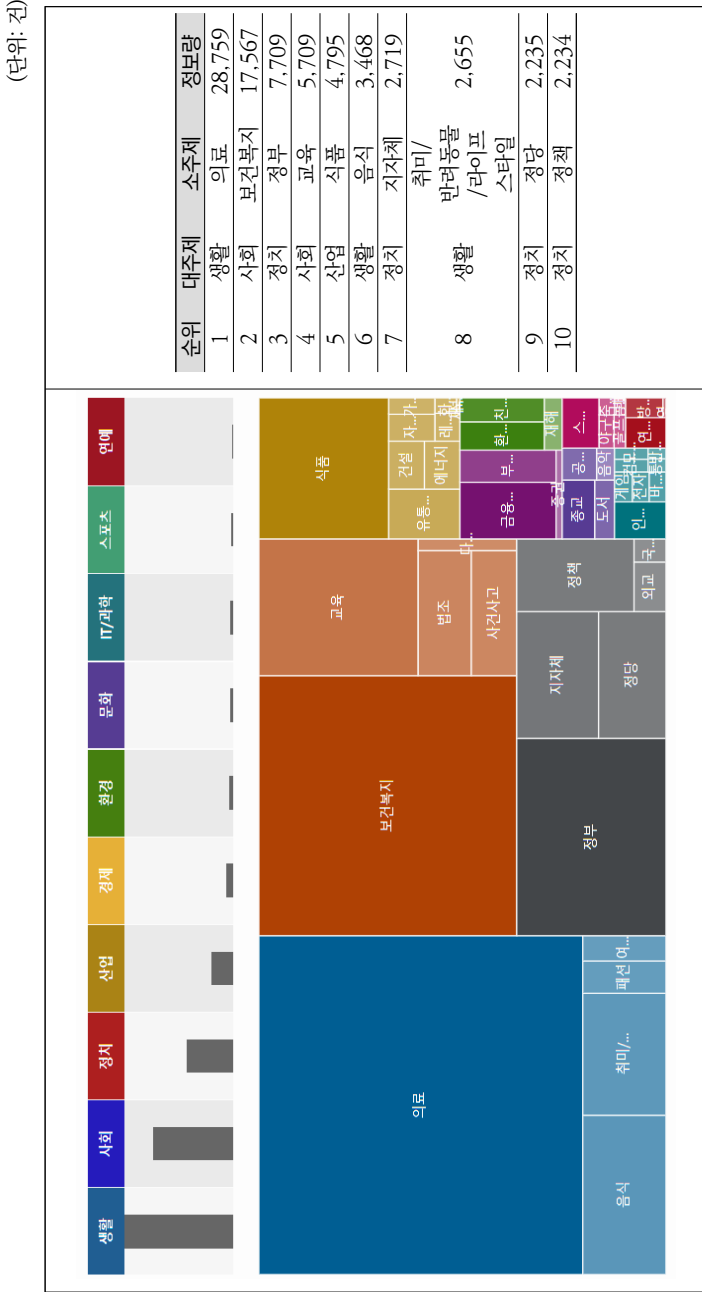


(표 3-20) 2021년 월별 '건강 보강' 정보량 상위 20개 연관 키워드

순위	전체	21.1.	21.2.	21.3.	21.4.	21.5.	21.6.	21.7.	21.8.	21.9.	21.10.	21.11.	21.12.
1	정책	기준	관리	관리	관리	정책	사업	사업	강화	정책	사업	건강	건강
2	관리	추진	의로	정책	어렵다	강화	확대	확대	문제인	강화	정책	지역	관리
3	강화	정책	못한다	추진	못한다	부담	강화	관리	간병협	관리	추진	관리	사업
4	사업	관리	정책	사업	사업	못한다	정부	부담	케어	확대	관리	사업	추진
5	확대	사업	기준	못한다	정책	추진	추진	의로	정부	정부	강화	건강한	노력
6	추진	확대	강병협	사회	추진	간병협	의로	지역	정책	간병협	지역	노력	건강한
7	지역	강화	강화	정부	문제	어렵다	정책	강화	부담	국민	어렵다	정책	지역
8	간병협	못한다	사용	지역	지역	확대	부담	어렵다	확대	부담	사회	사회	정책
9	정부	지역	생각	치로	치로	추진	관리	간병협	성과	우리	의로	우리	서비스
10	사회	사회	생각	안전	환경	노력	간병협	치로	대통령	시행	기준	운영	사회
11	못한다	안전	병원	어렵다	사회	치로	지역	사회	국민	사회	기관	기관	센터
12	의로	의로	사회	기준	확대	사회	기준	정책	대책	건강한	확대	문제	운영
13	어렵다	정부	문제	강화	의로	정부	못한다	기준	추진	문제	못한다	추진	강화
14	부담	문제	확대	서비스	정부	의로	어렵다	시행	못한다	의로	우리	서비스	기관
15	기준	간병협	사업	확대	강화	지역	실시	추진	4주년	추진	건강한	강화	확대
16	치로	증가	우리	의로	기준	생각	치로	건강한	우리	기관	운영	생각	개선
17	우리	사용	국민	문제	중요하다	사업	시행	노력	치로	기준	시행	확대	안전
18	시행	시행	치료	사용	시행	국민	사회	중요하다	관리	못한다	간병협	생활	의료
19	문제	제도	추진	유지	우리	중요하다	문제	못한다	어렵다	사업	서비스	의료	정부
20	국민	중요하다	제도	우리	서비스	시행	환경	환경	청와대	지역	사용	환경	어렵다

주: 1) '보장'은 모든 달에서 가장 많이 추출된 연관키워드로 위 표에는 제외하였음.  
 2) LUCY 2.0 속성분류체계에 따라 빨간색 음영은 부정, 파란색 음영은 긍정을 나타내는 키워드임.  
 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성.

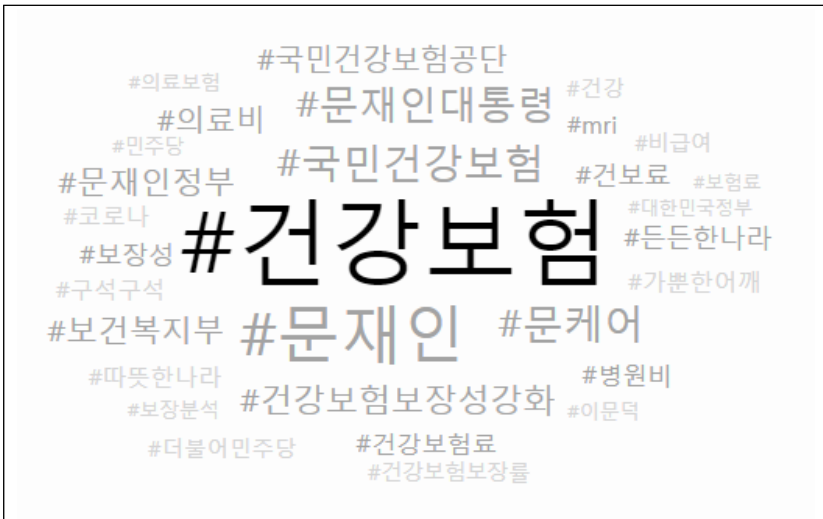
[그림 3-21] 2021년 '건강 보장 토픽 매칭 현황



주: LUCY 2.0 토픽검색 분류 방식에 따라 정치, 경제, 경제, 사회, 스포츠, 연예, 문화, 생활, IT/과학, 산업, 환경 등 10개의 대항목과 53개의 소항목으로 분류. 자료: 소셜 데이터 분석 시스템 LUCY 2.0을 기반으로 저자 작성

2021년 동안 인스타그램, 트위터, 페이스북, 유튜브, 카카오톡, 블로그, 카페 채널을 통한 해시태그 검색 결과 '#건강보장'에 해시태그가 언급된 데이터는 전체 377건이며, '#건강보험'(75건), '#문재인'(49건), '#국민건강보험'(24건) 해시태그가 가장 많이 함께 언급되었다([그림 3-22] 참조).

[그림 3-22] 2021년 '건강 보장' 정보량 상위 30 해시태그



자료: 소셜 데이터 분석 시스템 LUCY 2.0에서 추출

## 6. 보건복지데이터포털에서의 LUCY 2.0 활용 사례

한국보건사회연구원에서는 국민의 보건복지 통계 수요가 늘어나고, 효율적인 통계 제공 및 관리의 필요성이 대두되면서 2009년 중장기 정보화 전략계획을 세운 이후 2011년부터 본격적인 보건복지데이터포털(data.kihasa.re.kr) 서비스를 시작하여 2021년까지 11년간 운영되고 있다([그림 3-23] 참조).

[그림 3-23] 보건복지데이터포털 메인화면



자료: 한국보건사회연구원 보건복지데이터포털. data.kihasa.re.kr

본 포털에서는 2018년부터 소셜데이터의 수집을 통한 소셜통계 현황(시각화) 서비스를 제공하였으며, 2021년은 LUCY 2.0을 활용하여 ‘보건’, ‘복지’, ‘사회보장’ 키워드의 소셜빅데이터 분석을 통해 현황을 제시하였다. 메인화면에는 3개의 연관키워드의 상위 50개의 키워드를 워드클라우드기법으로 제시하고 있으며, 소셜데이터 소메뉴에서는 일별, 상위 키워드갯수별(50~200개)로 조절하여 조회할 수 있다. 또한 워드클라우드

드 뿐만 아니라 막대그래프, 선그래프, 영역그래프 등 다양한 형태로 소셜데이터 분석결과를 시각화하고 있다(그림 3-24 참조).

[그림 3-24] 보건복지데이터포털 소셜데이터 메뉴 화면



자료: 한국보건사회연구원 보건복지데이터포털 > 통계시각화 > 소셜데이터.  
<https://data.kihasa.re.kr/kihasa/kor/social/socialList.html>

## 7. 마치며

사실상 빅데이터, 소셜빅데이터 분석에서는 나타난 결과에 대한 정확한 해석과 판단이 중요하다. 서론에서도 언급하였듯이 자료의 양이 많다고 하여 무조건적인 질을 보장하지는 않으므로 유용하지 않은 정보를 걸러내고 대표성의 문제로 인해 전체 대상에 대해 연관성을 적용하지 않으며, 의미있는 관련성을 골라내기 위해 많은 노력이 필요하다. 빅데이터가 만능일 수는 없는 것이다. 그러므로 정책을 수립하고 평가하는데, 빅데이터를 활용함에 있어서 분석도구를 통한 결과를 보고 정책 영역에 대한 보다 치밀한 검토를 통해 그 연관성을 찾아내야 할 것이다.

이번 연구에서는 정책 영역에 대해 분석결과를 보다 상세하게 검토하지 않은 상태로 주변에 많은 인프라로 널려있는 빅데이터 분석플랫폼, 소셜빅데이터 분석플랫폼에 대해 소개하는 것에 의미를 두었다. ‘빅데이터’라는 단어가 의미하듯이 우리 주변에는 너무나 많은 데이터가 곳곳에서 생산되고 있지만, 과연 적재적소에서 잘 활용하고 있는지는 장담할 수 없다. 다행히 정부사업으로 영역별 공공, 민간데이터가 유통될 수 있는 기반이 마련되어 있는 만큼 이에 대한 관심과 보다 적극적인 활용이 필요하다 하겠다.



- 과학기술정보통신부 보도자료. (2021.12.30.). 데이터 댐이 이끈 데이터산업 시장, 2020년 2019년 대비 18.9% 성장한 20조원 달성.
- 국회입법조사처. (2020). 빅데이터 플랫폼의 운영 실태와 개선과제. 입법·정책보고서 제55호.
- 김도원. (2017.12.11.). '문재인 대통령 "의료계가 '문재인 케어' 지혜 모아달라'. YTN.  
<https://news.naver.com/main/read.naver?mode=LSD&mid=sec&sid1=001&oid=052&aid=0001092882>
- 뉴시스. (2021.8.12.). 문 대통령, '문케어 4주년' 성과보고...“병원비 걱정 없는 나라”. 동아일보.  
<https://www.donga.com/news/Politics/article/all/20210812/108519812/1>
- 대한민국정책브리핑 중 '한눈에 보는 정책'. (2021.8.20.). 문재인케어.  
<https://www.korea.kr/special/policyCurationView.do?newsId=148863918>.
- 박지성. (2021.4.20.). 한국수자원공사 '환경 빅데이터 플랫폼'. 전자신문.  
<https://m.etnews.com/20210414000050>.
- 박진주. (2019.7.2.). '문케어' 2년..국민 의료비 2조 2천억 경감 혜택 봐. MBC.  
<https://news.v.daum.net/v/20190702194605860>
- 산업통상자원부 보도자료. (2021.2.25.). 기업의 비즈니스 혁신, 산업 빅데이터 플랫폼으로 빠르고 쉽게!. p.2.
- 서한기. (2017.8.19.). '문재인 케어' 시동...미용·성형 빼고 모든 치료에 건보 적용. 연합뉴스.  
<https://www.yna.co.kr/view/AKR20170725140100017?input=1195m>
- 알에스엔. (2021). LUCY 2.0 Product Introduction.

- 연구개발특구진흥재단. (2021). 빅데이터 시장. 글로벌 시장동향보고서.
- 유경원. (2020). 빅데이터를 이용한 정책분석 방법론에 관한 연구. 텔코저널 제8권 0호. pp.39-75.
- 윤용찬. (2018.11.). 극도로 복잡한 한국어 소셜버즈분석 필요없는 데이터 잘버리는게 핵심. Issue2 261호.  
[https://dbr.donga.com/article/view/1202/article\\_no/8894](https://dbr.donga.com/article/view/1202/article_no/8894)
- 윤유동, 조재춘, 허윤아, 임희석. (2017). 소셜 미디어 데이터분석을 활용한 빅데이터에 대한 인식변화 비교 분석. 소프트웨어 및 데이터공학. 제6권 제7호 pp.371~378. 정보처리학회논문지.
- 이관주. (2020.12.27.). 현실로 다가온 '스마트 치안'...순찰 경로·방법시설 설치 "인공지능 추천". 아시아경제.  
<https://www.asiae.co.kr/article/2020122619243856497>.
- 이슬기. (2018.10.10.). '복지위, 국민연금·文케어 놓고 대립..野 '강공'·與 '엄호'(종합). 연합뉴스. <https://news.v.daum.net/v/20181010200634153>
- 이연호. (2018.4.1.). '醫-政, 문케어 정면 충돌...의협 4월중 의료파업 예고. 이데일리.  
<https://news.naver.com/main/read.naver?mode=LSD&mid=sec&sid1=001&oid=018&aid=0004068757>
- 이윤희. (2014). 국내 SNS의 이용현황과 주요 이슈 분석. INTERNET & SECURITY FOCUS. pp. 56-78. 한국인터넷진흥원.
- 장지민. (2019.12.16.). 건강보험 보장률 63.8%..전년대비 1.1%p ↑ '文케어 2년차 성적표'. 한국경제.  
<https://news.v.daum.net/v/20191216220117049>
- 최민재. (2009). 소셜 미디어의 확산과 미디어 콘텐츠에 대한 수용자 인식연구. 한국언론정보학회 학술대회, pp. 5-31.
- 한신갑. (2015). 빅데이터와 사회과학하기: 자료기반의 변화와 분석전략의 재구성. 한국사회학 제49집 제2호. pp.161-192. 한국사회학회.



FkII 조사연구팀. (2006.12.). 소셜 미디어(Social Media)란 무엇인가?. 『정보 산업지』. pp.52-55.

MarketsandMarkets. (2020).Big Data Market.

Seth Stephens-Davidowitz. (2017). “Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are”, Foreword by Steven Pinker,

교통빅데이터 플랫폼: 국가교통데이터 오픈마켓.

<https://www.bigdata-transportation.kr/>

국가법령정보센터-노인복지법, <https://www.law.go.kr/법령/노인복지법>

국가법령정보센터-장애인복지법 <https://www.law.go.kr/법령/장애인복지법>

금융 빅데이터 플랫폼. <https://www.bigdata-finance.kr/>

농식품 빅데이터 플랫폼: 농식품 빅데이터 거래소. <https://kadx.co.kr/>

디지털산업혁신 빅데이터 플랫폼. <https://www.bigdata-dx.kr/>

라이프로그 빅데이터 플랫폼. <https://www.bigdata-lifelog.kr/>

문화 빅데이터 플랫폼. <https://www.culture.go.kr/bigdata/>

산림빅데이터 플랫폼: 산림빅데이터 거래소. <https://www.bigdata-forest.kr/>

소방안전 빅데이터 플랫폼. <https://www.bigdata-119.kr/>

스마트치안 빅데이터 플랫폼. <https://www.bigdata-policing.kr/>

유통 빅데이터 플랫폼: 한국데이터거래소. <https://kdx.kr/>

중소중견기업 빅데이터 플랫폼.

<https://www.bigdata-sme.kr/#/datastore/landing>

지역경제 빅데이터 플랫폼: 경기지역경제포털.<https://bigdata-region.kr/#/>

통신 빅데이터 플랫폼. <https://www.bigdata-telecom.kr/>

통합데이터지도. <http://www.bigdata-map.kr/>

한국보건사회연구원 보건복지데이터포털. [data.kihasa.re.kr](http://data.kihasa.re.kr)

해양수산 빅데이터 플랫폼: 해양수산 빅데이터 거래소.

<https://www.bigdata-sea.kr/>

헬스케어 빅데이터 플랫폼: 국립암센터CONNECT.

<https://www.bigdata-cancer.kr/ncc/main.do>

환경 비즈니스 빅데이터 플랫폼. <https://www.bigdata-environment.kr/>

LUCY 2.0 브랜드 홈페이지. <https://lucy2.realsn.com/>