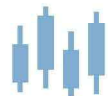


WP 20-09

# 빅데이터를 이용한 국토 민생현안 모니터링 연구

| 장요한 부연구위원 (ycanns@krihs.re.kr) |



※ 이 Working Paper의 내용은 국토연구원의 공식 견해가 아니며, 저자 개인의 의견입니다. 연구 내용에 대하여 궁금한 점은 저자의 이메일로 문의하여 주시고, 인용 시에는 저자 및 출처를 반드시 밝혀주시기 바랍니다.

※ 본 연구는 서울시와 서울신용보증재단의 자료 협조를 통하여 수행되었습니다.

## 차례

01 연구의 배경 및 목적	05
02 국토 민생현안 모니터링과 빅데이터	07
03 국토 민생현안 모니터링 사례분석	15
04 정책제언 및 향후 연구방향	31



# 01 연구의 배경 및 목적

## ■ 국토정책 환경의 변화

- ‘공급’ 중심으로 이루어졌던 고도성장 이후 지역 간 불균형 등의 문제점 대두, 기존 물리적 성장방식의 한계 직면 및 지속적인 균형발전의 필요성 증대
- 자원의 효율적인 재분배와 다변화하는 사회현상에 대한 구체적인 대안을 마련하고 미래 현상에 대한 중·장기적 예측과 대응전략의 선제적 수립 필요
  - 저출산·고령화를 비롯한 인구감소, 생활사막(Life desert) 확대 등 국토환경 변화에 대응한 정책수립 필요(이영주·황명화 외 2016)

## ■ 시민의식의 제고와 국민체감형 정책의 필요성 증대

- 시민의식<sup>1)</sup>은 2014년 이후 매년 평균 0.64%씩, 정치적 역량감<sup>2)</sup>은 2016년 이후 매년 평균 7.36%씩 지속적으로 상승(삶의 질 지표, 국가지표체계)
- 국민 의식수준의 향상과 더불어 삶의 질 개선을 위해 국민이 직접 체감할 수 있는 정책 수립의 필요성 증대
- 체감형 정책 수립을 위한 지속적인 국민의 소리 수렴 필요성 증대

## ■ 대·내외 환경의 변화

- 4차 산업혁명의 정보통신기술(Information & Communication Technology: ICT) 정책 패러다임 변화와 국민수요에 대응할 수 있는 양방향 소통의 필요성 증대, 이에 따른 사회·경제 분야의 효율적이고 체계적인 대응물 필요
- 정책 패러다임 변화에 따른 공공부문 클라우드(Cloud) 서비스, 인공지능(Artificial Intelligence: AI), 데이터경제(Data Economy) 구축 등에 대한 범적인 중요성이 강조

1) 시민으로서 가져야 할 책무에 대해 중요하다고 인식하는 인구비율 (국가지표체계 [www.index.go.kr](http://www.index.go.kr), 2020년 3월 25일 검색).  
 2) 스스로 정치적 역량감을 가지고 있다고 인식하는 인구비율 (국가지표체계 [www.index.go.kr](http://www.index.go.kr), 2020년 3월 25일 검색).

## ■ 국정운영과 국민 소통방식의 패러다임 전환

- 국민의 소리를 정책으로 환류시켜 국민 행복에 기여할 수 있는 지능형 정부 구현의 필요성 증대
- 디지털 정부의 구현이 진전됨에 따라 정책 수립에 있어 국민과의 소통방식이 최근 국민청원 및 SNS 등을 통해 다변화되었고 국민의 정책수요 또한 다양화

## ■ 참신한국토 분야 연구 아이디어 발굴

- 기존 전문가 중심의 국토정책 연구 아이디어 발굴 방법은 지속적인 모니터링이 어렵고 한시적이어서 최근 변화하는 정보화 기반 정책 패러다임에 한계점 직면
- 지속가능하고 참신한 수단인 ‘국민의 소리’ 수렴을 위해 빅데이터와 정보통신기술(ICT)을 이용한 국민 소리 모니터링을 시도하는 데에 이 연구의 목적이 있음
- ‘국토’와 관련된 중요 키워드들을 중심으로, 대한민국 전국의 주요 중앙지 및 지역별 뉴스 기사를 수집하고, 오픈소스(Open-source)와 빅데이터(Big Data) 분석기법을 이용하여 지역별·시기별 연구과제 발굴을 제안하고자 함
- 이를 통해 지역별·분야별 구체적인 연구방향 도출과 지속적인 모니터링을 진행하는 데 기여할 수 있을 것으로 예상

# 02 국토 민생현안 모니터링과 빅데이터

## 1) 민생현안과 모니터링의 정의

■ 국립국어원과 메리엄 웹스터(Merriam-Webster) 사전의 정의에 따르면, ‘민생’, ‘현안’, ‘모니터링’의 각 사전적 의미는 아래 <표 1>과 같고, 정의된 내용들을 바탕으로 ‘민생현안 모니터링’의 개념을 다음과 같이 조작적으로 정의함

- 일반 국민의 안정된 생활 및 생계유지를 위해, 논쟁의 여지가 있는 (이전부터 의논하여 오면서도 아직 진행 중이거나 또는 새로운) 문제들을 지속적으로 예의주시하여 해결방안을 강구하는 일

**표 1** 국립국어원과 메리엄 웹스터(Merriam-Webster) 사전의 ‘민생, 현안, 모니터링’에 대한 정의

<p><b>민생(民生, Public Welfare)</b></p> <ul style="list-style-type: none"> <li>- 일반 국민의 생활 및 생계 (국립국어원)</li> <li>- the state of doing well especially in respect to good fortune, happiness, well-being, or prosperity: 대중의 안정된 생활영위 (Merriam-Webster)</li> </ul> <p><b>현안(懸案, issue 또는 current issue)</b></p> <ul style="list-style-type: none"> <li>- 이전부터 의논하여 오면서도 아직 해결되지 않은 채 남아있는 문제나 의안 (국립국어원)</li> <li>- 이슈(issue): 논쟁의 중심이 되는 문제 (국립국어원)</li> <li>- a vital or unsettled matter; concern; problem; a matter that is in dispute between two or more parties; the point at which unsettled matter is ready for a decision. (중략): 중대하거나 조율 또는 의사결정이 필요하지만 대립되는 이해관계 등으로 진행되지 않은 문제 또는 걱정들 (Merriam-Webster)</li> </ul> <p><b>모니터링(monitoring)</b></p> <ul style="list-style-type: none"> <li>- to watch, keep track of, or check usually for a special purpose: 특정목적을 가지고 예의주시하고 주기적인 확인을 하는 행위 (Merriam-Webster)</li> </ul>
---

출처: 국립국어원 (www.korean.go.kr); Merriam-Webster (https://www.merriam-webster.com, 2020년 4월 1일 검색).

## 2) 빅데이터

- 빅데이터(Big Data)의 의미에 관해 국립국어원은 양적인 특성에 집중하고 있는 반면, 메리엄 웹스터(Merriam-Webster)는 양적인 특성과 더불어 데이터의 구조적인 특성을 함께 강조
- 하지만 데이터의 성격, 활용 분야 및 목적 등에 따라서 파생될 수 있는 데이터의 관점과 특성이 다양해지기 때문에, 빅데이터의 사전적 정의는 한계가 있음(김동한 외 2014, 14)
- 최근에 고려되고 있는 빅데이터의 개념은 3차원 데이터관리(3-D Data Management)로써, 3V\*의 개념을 기반으로(Laney 2001), 복잡성(Complexity)과 가치성(Value) 등을 추가로 고려(김동한 외 2014, 14; 김미정 외 2013 재인용)
  - \* 데이터 크기(Volume), 속도(Velocity), 다양성(Variety)
- 다시 말해 빅데이터는 사전에서 정의하는 데이터의 양적인 특성뿐만 아니라, 해당 데이터의 다양성과 복잡성이 설명될 수 있는 구성 및 구조적인 특성, 시간적 흐름에 따라 변화하는 동태적 특성, 그리고 해당 데이터로부터 유도해낼 수 있는 활용가치들이 함께 존재하는 질적인 특성 등의 속성을 가지는 데이터로 확장하여 정의 가능 (김대중·윤서연 2013; 김동한 외 2014; 황명화 2014)

**표 2** 국립국어원과 메리엄 웹스터(Merriam-Webster) 사전의 '빅데이터'에 대한 정의

<p><b>빅데이터(Big Data)</b></p> <ul style="list-style-type: none"> <li>- 기존의 데이터베이스로는 수집·저장·분석 따위를 수행하기가 어려울 만큼 방대한 양의 데이터 (국립국어원)</li> <li>- An accumulation of data that is too large and complex for processing by traditional database management tools: 기존의 데이터베이스 관리 도구 등으로 분석하고 처리하기에는 지나치게 방대하고 복잡한 축적 데이터 (Merriam-Webster)</li> </ul>
--

출처: 국립국어원. www.korean.go.kr; Merriam-Webster. https://www.merriam-webster.com (2020년 4월 1일 검색).

## 3) 사례 및 선행연구

- 사회관계망(Social Network Service: SNS) 데이터, 뉴스기사 기록 등의 자료를 분석하여 다양한 정책적·인과적 의미를 도출해내는 방법들에 관한 여러 선행연구가 존재
- 분석하고자 하는 연구의 목적과 방향에 따라 다양한 빅데이터와 분석방법을 활용



■ 국민권익위원회는 국민신문고 등을 통해 수집된 민원정보를 주별·월별·연도별로 분석하여, 그 결과를 ‘빅데이터로 보는 국민의 소리’라는 간행물을 통해 정기적으로 공개

- 주간·월간·연간 민원동향, 주요 민원사례, 민원예보(모니터링), 국민불편 개선요청 및 개선사례 등을 취합하여 발간
- 민원동향에서는 해당 기간 동안의 전반적인 민원추이와 신청인 현황, 민원이 접수된 기관 유형별 현황 등의 통계지표를 제공하고, 주요민원 사례에서는 이슈가 되고 있는 사례들을 집중 조명
- 국민신문고에서 주력으로 운영하고 있는 민원예보는 국민신문고 민원을 통해 수집된 모니터링 내용들 중 ‘국민의 생명과 안전을 위협하거나 대규모 갈등 및 피해 등을 유발하는 민원이 발생하는 경우’를 지속적으로 확인하고 해당 관계부처에 시정명령을 요청하는 방식으로 운영(국민권익위원회, <http://www.acrc.go.kr/acrc/index.do>)

그림 1 빅데이터로 보는 국민의 소리



출처: 국민권익위원회. <http://www.acrc.go.kr/acrc/index.do> (2020년 8월 27일 검색).

■ 서울시는 ‘응답소’와 ‘서울시 여론조사’ 서비스를 통해 국민의 소리 환류 및 서울시민 민생 현안파악을 위한 서비스를 진행 중

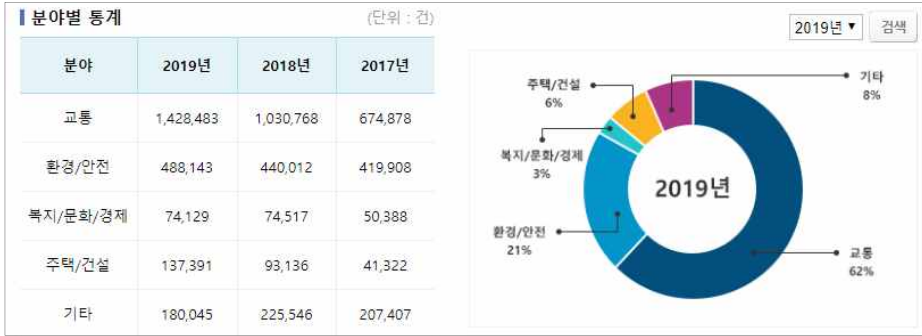
- 2011년 서울시의 민원 환류 정책방안 중 하나로 공식 개시된 ‘서울시 SNS 서포터즈’는 이후 시민들과의 소통방법과 현안파악의 효율적 운영 고도화를 통해 현재의 ‘응답소’와 ‘서울시 여론조사’ 사업으로 발전
- ‘응답소’와 ‘서울시 여론조사’ 모두 시민과 공공기관의 정책 소통도구로 활용된다는 점에서는 역할이 동일하지만, ‘응답소’는 시민의 민원수렴에 ‘서울시 여론조사’는 서울시의 정책의지와 그에 따른 시민들의 수요의식을 보다 자세하게 파악하는 데 초점을 둠
- (응답소) 전화, 문자, 챗봇, SNS, 유선전화로 상시 접수되는 다양한 민원들에 대해 온라인 기반의 ‘응답소’ 홈페이지를 통해 대응하고, 누적된 민원들을 수렴하여 서울시 정책에 반영하는 서비스를 진행 중
  - 신청수단별, 민원분야별, 분야별, 처리기관별, 자치구별 민원현황에 대해 2017년부터 상시 통계정보를 제공
- (서울시 여론조사) 온라인으로 진행되는 여론조사 시스템으로, 서울, 경기 및 인천에 거주하는 만 14세 이상의 시민들의 자발적 참여로 운영
  - 여론조사 패널 가입 후 각종 여론조사에 대한 참여도에 따라 T-Money (수도권의 후불 교통카드) 또는 도서문화상품권 등으로 전환할 수 있는 마일리지 제공
  - 전화, 면접, 온라인, 좌담회 등으로 수집된 주제별 여론조사 결과는 일정 기간 후 서울시 여론조사 홈페이지를 통해 공개

그림 2 서울시의 여론조사 분석을 통한 정책 발굴 및 운영 사례

조사결과		여론조사 결과																																																						
<p>여론조사가 종료되고, 결과 데이터 검증 후 일부 문항 결과가 공개됩니다.                      여론조사 종료 후 공개까지 기간은 약 3-6개월 정도 소요됩니다.                      조사결과 세부내용 문의 및 활용관련 사항은 담당부서와 협의하시고, 활용 시 출처를 명기하여 주시기 바랍니다.                      여론조사 결과 확인을 위해서 조사계약을 클릭하세요.</p>		<p>조사기간: 2009-01-01 ~ 2020-04-03    부서/제목    검색</p> <p>총 376 건    출력수    검색</p> <table border="1"> <thead> <tr> <th>번호</th> <th>구분</th> <th>부서</th> <th>조사제목</th> <th>조사 방법</th> <th>기간</th> <th>참여 인원</th> <th>공개 여부</th> </tr> </thead> <tbody> <tr> <td>376</td> <td>문화/관광</td> <td>문화본부 한양도성도감</td> <td>한양도성에 대한 시민인식 여론조사</td> <td>온라인</td> <td>2020-02-14 ~ 2020-02-19</td> <td>3685</td> <td>준비중</td> </tr> <tr> <td>375</td> <td>교통</td> <td>도시교통본부 버스정책과</td> <td>서울시 시내버스 혼잡 인식 및 이용 실태 설문조사</td> <td>온라인</td> <td>2020-02-11 ~ 2020-02-16</td> <td>3656</td> <td>준비중</td> </tr> <tr> <td>374</td> <td>홍보</td> <td>시민소통기획관 뉴미디어담당관</td> <td>서울시 홈페이지 이용에 대한 설문조사</td> <td>온라인</td> <td>2020-02-06 ~ 2020-02-13</td> <td>3593</td> <td>준비중</td> </tr> <tr> <td>373</td> <td>소방/안전</td> <td>소방재난본부 안전지원과</td> <td>2019년 서울시민 안전의식에 대한 여론조사</td> <td>온라인</td> <td>2019-12-31 ~ 2020-01-05</td> <td>2607</td> <td>공개</td> </tr> <tr> <td>372</td> <td>홍보</td> <td>시민소통기획관 뉴미디어담당관</td> <td>2019년 월간 '서울사랑' 독자 만족도 조사</td> <td>온라인</td> <td>2019-12-19 ~ 2019-12-27</td> <td>2232</td> <td>공개</td> </tr> </tbody> </table>							번호	구분	부서	조사제목	조사 방법	기간	참여 인원	공개 여부	376	문화/관광	문화본부 한양도성도감	한양도성에 대한 시민인식 여론조사	온라인	2020-02-14 ~ 2020-02-19	3685	준비중	375	교통	도시교통본부 버스정책과	서울시 시내버스 혼잡 인식 및 이용 실태 설문조사	온라인	2020-02-11 ~ 2020-02-16	3656	준비중	374	홍보	시민소통기획관 뉴미디어담당관	서울시 홈페이지 이용에 대한 설문조사	온라인	2020-02-06 ~ 2020-02-13	3593	준비중	373	소방/안전	소방재난본부 안전지원과	2019년 서울시민 안전의식에 대한 여론조사	온라인	2019-12-31 ~ 2020-01-05	2607	공개	372	홍보	시민소통기획관 뉴미디어담당관	2019년 월간 '서울사랑' 독자 만족도 조사	온라인	2019-12-19 ~ 2019-12-27	2232	공개
번호	구분	부서	조사제목	조사 방법	기간	참여 인원	공개 여부																																																	
376	문화/관광	문화본부 한양도성도감	한양도성에 대한 시민인식 여론조사	온라인	2020-02-14 ~ 2020-02-19	3685	준비중																																																	
375	교통	도시교통본부 버스정책과	서울시 시내버스 혼잡 인식 및 이용 실태 설문조사	온라인	2020-02-11 ~ 2020-02-16	3656	준비중																																																	
374	홍보	시민소통기획관 뉴미디어담당관	서울시 홈페이지 이용에 대한 설문조사	온라인	2020-02-06 ~ 2020-02-13	3593	준비중																																																	
373	소방/안전	소방재난본부 안전지원과	2019년 서울시민 안전의식에 대한 여론조사	온라인	2019-12-31 ~ 2020-01-05	2607	공개																																																	
372	홍보	시민소통기획관 뉴미디어담당관	2019년 월간 '서울사랑' 독자 만족도 조사	온라인	2019-12-19 ~ 2019-12-27	2232	공개																																																	

출처: 서울특별시 여론조사. <http://research.seoul.go.kr/surveyWebGuide.do> (2020년 4월 1일 검색).

**그림 3** 서울시의 민원 데이터 분석을 통한 정책 발굴 및 운영 사례



출처: 서울특별시 응답소. <https://eungdapo.seoul.go.kr/> (2020년 4월 1일 검색).

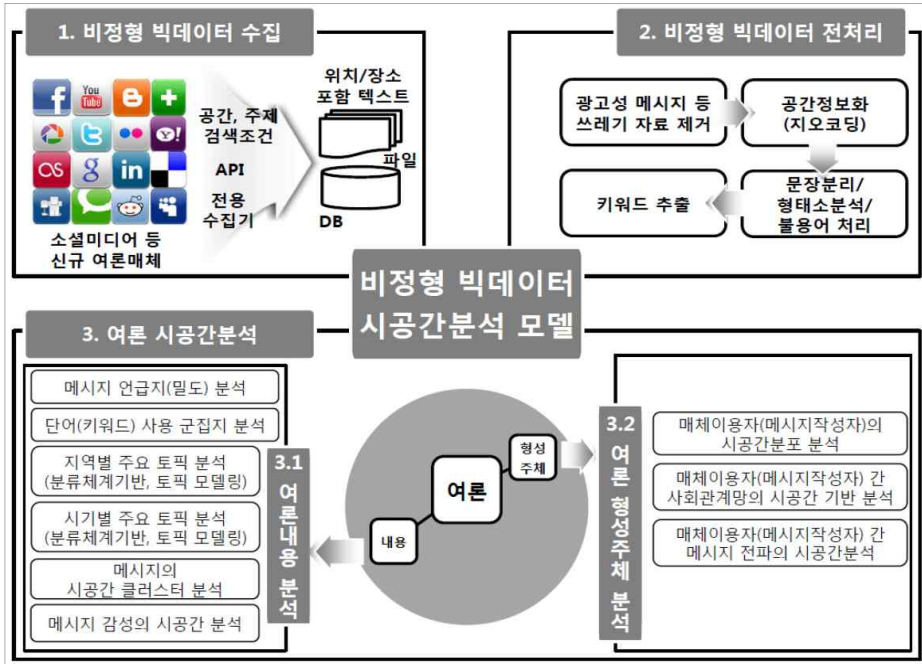
■ **황명화(2014)는 맞춤형 여론 모니터링을 위한 비정형 빅데이터 시공간 분석방법론에 관한 연구를 수행**

- 사람들이 자유롭게 공유하고 정보를 주고받는 글자 중심의 비정형 빅데이터를 크롤링 (Crawling)<sup>3)</sup>, 텍스트마이닝(Text mining)<sup>4)</sup> 등 다양한 비정형 데이터 분석기법들과 시공간 빅데이터 분석 등 정형 데이터 분석기법들을 함께 융합하여 시기별·지역별 여론 수요변화(Trend)를 도출하는 분석모델을 제안
- 소셜미디어, 언론매체 등 다양한 SNS 데이터를 시공간적으로 분석하고 지속적으로 모니터링하는 방법을 제안함으로써, 국민 수요 및 지역 맞춤형 국가·지역 정책수요 발굴에 이바지할 수 있을 것으로 기대

3) 소프트웨어 따위가 웹을 돌아다니며 유용한 정보를 찾아 특정 데이터베이스로 수집해 오는 작업. 또는 그러한 기술 (국립국어원, [https://opendict.korean.go.kr/dictionary/view?sense\\_no=837782](https://opendict.korean.go.kr/dictionary/view?sense_no=837782), 2020년 4월 3일 검색).

4) 정형 텍스트 데이터에서 새롭고 유용한 정보를 찾아내는 과정 또는 기술로 규범 표기는 미확정 (국립국어원, [https://opendict.korean.go.kr/dictionary/view?sense\\_no=1003292&viewType=confirm](https://opendict.korean.go.kr/dictionary/view?sense_no=1003292&viewType=confirm), 2020년 4월 3일 검색).

그림 4 비정형 빅데이터 시공간분석 모델의 구성



출처: 황명화 2014, 40.

■ 해외 여러 나라에서도 시민 의견 모니터링을 통한 국정운영을 위해 노력 중

- 2009년 미국의 제44대 대통령 오바마가 개시한 개방형정부운영(Open Government Initiative) 방침은 정보의 자유(The Freedom of Information Act: 5USC§552), 정부의 투명성(Transparency)과 시민참여(Participation), 협력(Collaboration)의 세 가지 목표를 추구5)
  - 연방정부는 각 기관별로 보유한 데이터들을 공공데이터 차원에서 통합적으로 관리하고 트위터(Twitter), 깃허브(Github), API(Application Programming Interface) 등 다양한 방법으로 시민에게 개방
  - 온라인 국민청원 창구6)를 이용한 시민의 의견 적극 수렴과 소통을 통한 국정운영의 극대화를 위해 시민과 정부기관과의 협의가 이뤄지는 모든 과정에 시민이 참여하고 접근할 수 있도록 온라인 게시판7) 플랫폼(Google groups)을 운영하고 개방
- 2017년부터 시작된 영국 디지털전략(UK Digital Strategy8) 중 범정부 전산행정 고도화

5) Obama white house. <https://obamawhitehouse.archives.gov/open> (2020년 4월 3일 검색).

6) Petitions, Obama white house. <https://petitions.obamawhitehouse.archives.gov/> (2020년 4월 3일 검색).

7) Groups Google. <https://groups.google.com/forum/#!forum/us-open-government> (2020년 4월 3일 검색).

8) Gouernment UK. <https://www.gov.uk/government/publications/uk-digital-strategy> (2020년 4월 3일 검색).

및 실증기반 정부운영 등의 사업 일환으로 시민들이 직접 서비스 및 국정운영 평가에 참여할 수 있고, 국가는 이 평가내용을 바탕으로 국정운영에 대한 시민들의 의식을 모니터링하는 서비스를 시범 운영 중<sup>9)</sup>

- 모니터링을 위해 수집 및 평가되는 정보는 비용 대비 정부의 업무 효율성, 기 청원된 문제의 해결여부와 시민의 만족도, 전화 및 우편 방식보다 온라인으로 지원되는 서비스의 편의성 등을 모니터링의 척도로 같음

■ 한편, 신뢰성 있는 뉴스미디어에 대한 구별과 누적된 비정형 뉴스미디어의 분석에 대한 시민 수요 부흥을 위해 한국언론진흥재단의 기획 아래 2016년부터 빅카인즈(Big KINDS) 서비스를 개시, 폭넓은 수요에 부흥할 수 있는 손쉬운 뉴스 분석 서비스를 제공

- 빅카인즈<sup>10)</sup>에서 수집하고 분석 중인 언론매체는 총 54곳으로 중앙지, 경제지, 지역종합지, 방송사 및 전문지 등 다양한 유형의 매체로 구성
- 1990년부터 현재까지 약 6천만 건의 비정형 데이터를 포함하는 텍스트, 사진, 또는 문서 등 다양한 형태의 뉴스기사들을 포함
- 실시간으로 발간되는 뉴스기사들을 주기적으로 수집하고 내부 시스템을 통해 뉴스카테고리를 분류한 뒤, 분류된 각 카테고리별 뉴스기사들을 형태소<sup>11)</sup> 및 바이그램<sup>12)</sup> 분석, 개체명 분석, 네트워크 분석 등을 통해 사용자들에게 무료로 공급

**표 3** 빅카인즈에서 수집 중인 지역별·특성별 54개 언론매체

구분	언론매체
중앙지	경향신문, 국민일보, 내일신문, 동아일보, 문화일보, 서울신문, 세계일보, 조선일보, 중앙일보, 한겨레, 한국일보
경제지	매일경제, 머니투데이, 서울경제, 아시아경제, 아주경제, 파이낸셜뉴스, 한국경제, 해럴드경제
지역종합지	강원도민일보, 강원일보, 경기일보, 경남도민일보, 경남신문, 경상일보, 경인일보, 광주매일신문, 광주일보, 국제신문, 대구일보, 대전일보, 매일신문, 무등일보, 부산일보, 영남일보, 울산매일, 전남일보, 전북도민일보, 전북일보, 제민일보, 중도일보, 중부매일, 중부일보, 충북일보, 충청일보, 충청투데이, 한라일보
방송사	KBS, MBC, OBS, SBS, YTN
전문지	디지털타임즈, 전자신문

출처: 빅카인즈. www.bigkinds.or.kr (2020년 4월 3일 검색).

9) Government UK. <https://www.gov.uk/performance> (2020년 4월 3일 검색).

10) 빅카인즈. [www.bigkinds.or.kr](http://www.bigkinds.or.kr) (2020년 4월 3일 검색).

11) 키워드를 문장의 최소 단위인 형태소 단위로 검색하는 방식(빅카인즈 사용자 매뉴얼 2019).

12) 키워드를 두 글자 단위로 잘라 검색하는 방식(빅카인즈 사용자 매뉴얼 2019).

**그림 5** 빅카인즈의 언론 빅데이터 분석시스템 구성도



출처: 빅카인즈. [www.bigkinds.or.kr](http://www.bigkinds.or.kr) (2020년 4월 3일 검색).

#### 4) 시사점 및 소결

- 4차 산업혁명의 정보통신기술(ICT) 정책 패러다임 변화와 경제·사회 변화에 국정운영 및 계획 수립단계에도 효율적이고 체계적인 마중물 연구가 필요
- 3V를 뛰어넘어 복잡성과 다양성의 특성을 추가되는 빅데이터 시대에 데이터 기반의 정책연구의 계획 및 수립 필요성이 강조되고 있음
- 비정형화된 데이터로부터 정형화된 데이터 정제과정의 중요성 또한 의미 있는 빅데이터 발굴과 함께 강조되고 있지만, 아직까지 많은 응용연구 및 사례가 발굴되지 않음
- 때문에 국토 민생현안 모니터링의 필요성이 높아지고 있고, 이를 뒷받침해주기 위한 빅데이터, 오픈소스 등의 도구 및 기술 활용에 대한 필요성이 증대되고 있음

## 03 국토 민생현안 모니터링 사례분석

### 1) 사례분석을 위한 방법론 설계

#### ■ 모니터링을 위한 범위설정

- **(매체 범위)** 국토 전반에서의 유의미한 민생현안을 지속적으로 파악하기 위해서 공신력 있는 언론매체들을 대상으로 함
- **(공간적 범위)** 국토연구를 위해 전국을 대상으로 하고, 분석과정에서는 구득 및 처리 가능한 행정구역 단위만큼 세분화한 후 추가 분석
  - 전국 또는 중앙지, 인천·경기권, 강원권, 충청권, 영남권, 호남권, 제주권
- **(시간적 범위)** 3개월 단위로 설정하되, 가능한 자료의 특성에 따라 탄력적으로 조정
  - 모니터링의 연계성과 현실성을 위해 가장 최근의 범위를 대상으로 고려
  - 민생현안 분석을 위한 시간적 범위가 지나치게 길 경우, 뉴스기자들이 일반화되어 중요한 미시적 이슈를 간과하게 될 위험성이 높은 반면, 지나치게 짧게 범위가 설정될 경우에는 계절별 또는 현안별 큰 맥락을 놓칠 우려가 있어 국토정책에 필요한 유의미한 트렌드 분석에 어려움이 예상

#### ■ 분석방법

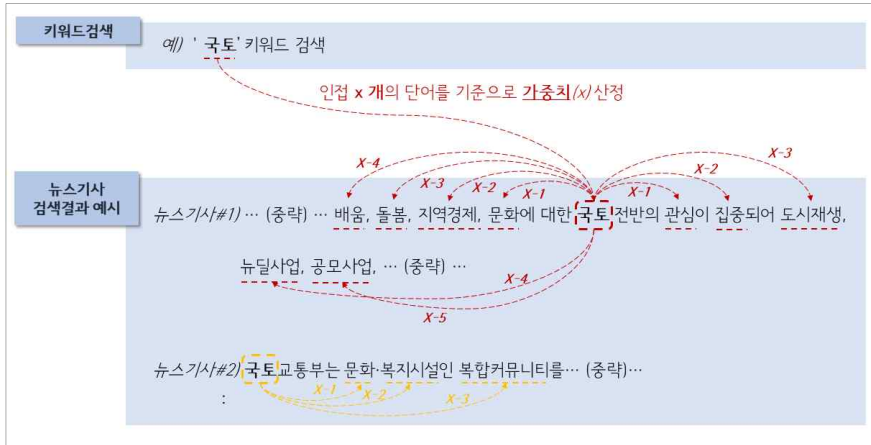
- 본 연구의 사례분석을 위해, 키워드 선별에서부터 키워드 기반의 단어위계 설정 및 세부 분석 방법들을 간략하게 설계함
- **(대상 키워드 선별)** 국토 전반의 전문가 의견수렴을 통해 국토 분야 민생현안과 가장 개연성이 높다고 판단되는 키워드군 선별
- **(키워드 위계 및 관계도 추정)** 「국민심문고」 및 국토 관련 연구보고서 등을 중심으로, 민생과 밀접한 관계가 있다고 판단되는 자료들을 이용하여 선별된 키워드들 간의 개략적인 위계 및 연관 관계 파악

- 선별된 키워드를 대상으로 시간적 범위 내 언론데이터들을 검색 및 수집 후, 각 키워드 언급빈도의 언론매체별 지역별 분포, 언론데이터의 특성 및 분포 파악
- **(뉴스기사특성 파악)** <그림 6>은 복합키워드 분석과정에서 작동되는 검색 키워드와의 의미적 상관도 및 가중치 산출과정을 도식화
  - 검색키워드를 통해 집계되는 뉴스기사들의 각 문장에서 차용된 인접 단어들을 중심으로 가중치 산정
  - 산정된 키워드들의 가중치를 모두 취합한 뒤, 해당 키워드와 관련된 뉴스기사들의 전반적인 성향 및 특성을 파악
- **(키워드군 설정)** 키워드 검색단계에서 각 키워드별로 집계된 뉴스기사 수를 바탕으로 키워드별 뉴스기사 빈도수 및 의미적 상관도<sup>13)</sup>를 파악하고, 유사 빈도 키워드들의 의미적 유사도에 따라 키워드군을 설정
- 대상 키워드군의 전반적 특성뿐만 아니라 키워드별 시·공간적 특성 파악을 위해, 대상 키워드를 개별 및 복합조합으로 구분하고 시·공간적 동향 및 특성 분석을 설계(<그림 7> 참조)
- **(개별키워드 동향 분석)** 해당 기간 동안 특정키워드의 뉴스보도 추이 및 일별 보도 특징을 추적하여 키워드별로 뉴스기사의 보도가 어떻게 다루어져 왔는지를 시계열적으로 파악
- **(키워드군 특성 분석)** 집계되는 뉴스기사의 시멘틱 클러스터링을 통해 뉴스기사 내 차용 어휘들의 특성을 분류한 뒤, 정치, 경제, 사건·사고 등의 특징으로 분류
- **(복합키워드 특성 분석)** 뉴스매체의 발표에는 다양한 키워드들이 복합적으로 조화되어 있고, 각 키워드들 간의 연관성 또한 밀접하게 연결되어 있기 때문에 1차적 검색을 통해 도출된 키워드와 연관성이 높은 키워드들을 n차 재검색을 통해 정제 및 구체화
  - 복합키워드를 통해 산출되는 단어군은 연속적으로 확인되는 n차 관심 키워드 그룹과 함께 뉴스기사 문장 내에서 가장 근접하게 위치되는 특징이 있고,
  - 빈번하게 집계되는 해당 단어군은 뉴스기사 전반에 걸쳐 상대적으로 높은 활용 빈도수를 가지고 있음을 의미
  - 이는 언론매체에서 민생의 관심에 대한 표현을 대신하기 위해 나타난 현상으로 간주될 수 있으며, 민생현안 파악의 상대적인 기준이 될 수 있을 것으로 기대

13) 시멘틱 클러스터링(Semantic Clustering): 단어 기반 특성분석을 통한 집단화(빅카인즈).

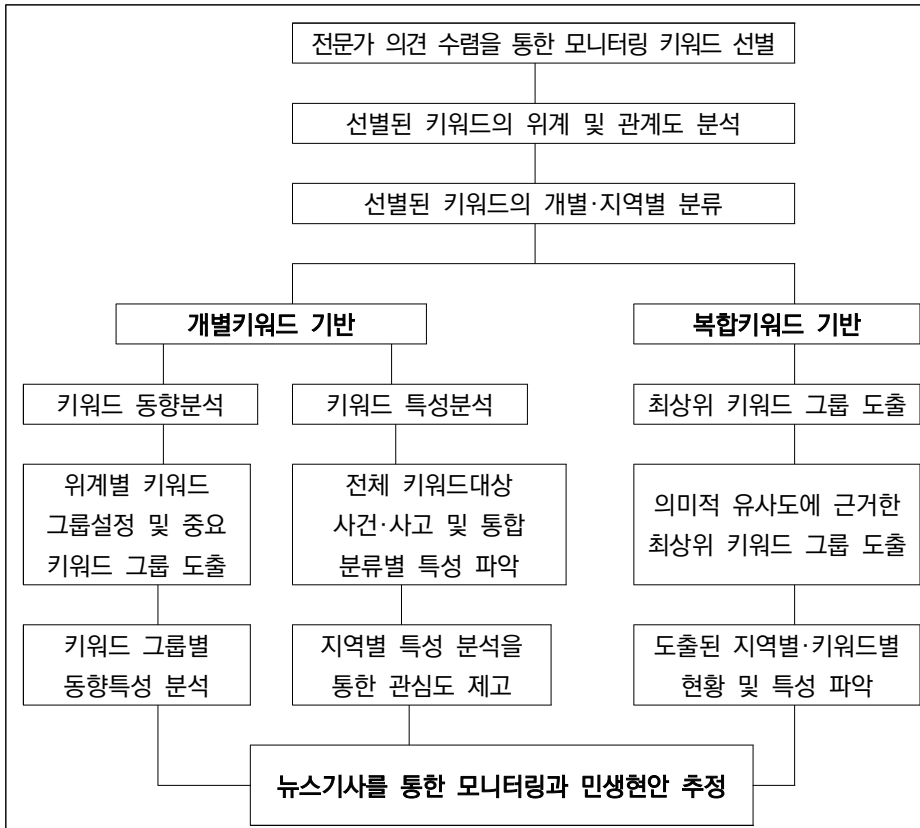


**그림 6** 키워드를 통한 의미적 상관도 및 가중치 산출과정 예시



주: 검색 예시 '국토'를 바탕으로 집계되는 각 뉴스기사에서 인접한 단어들을 중심으로 가중치가 산정되고, 추후에 이들 단어들을 집계하고 최종적으로 최상위 키워드 그룹을 산출하는 방식. 해당 예시에서는 '국토'키워드 검색에 '문화'의 가중치가 가장 높음.  
출처: 빅인즈의 내용을 바탕으로 저자 작성.

**그림 7** 언론 빅데이터 분석을 통한 키워드 분석방법 설정



출처: 저자 작성.

## 2) 사례분석을 위한 분석범위 및 키워드 설정

- 빅카인즈에서 제공하는 전국 54개의 검증된 국내 각종 언론매체들을 활용하고, 시간적 범위는 2019년 11월 1일부터 2020년 1월 31일까지 3개월을 대상으로 함\*
  - \* 분기별 분석대상기간 동안 국내 코로나19(COVID-19)의 급속 전파로, 뉴스기사의 내용이 지나치게 편중될 수 있는 점을 감안하여 시간적 범위를 재조정
- 빅카인즈에서 제공하는 통합분류\*와 사건·사고분류\*\*를 이용하여 검색결과 분석
  - \* (통합분류) 정치, 경제, 사회, 문화, 국제, 지역, 스포츠, IT과학
  - \*\* (사건·사고분류) 정치, 경제, 사회, 문화
- 국토연구원 각 연구본부들의 내부회의를 거쳐 선별된 본부별 중요 키워드를 취합했고, 취합된 국토 분야 민생현안 키워드 총 39개의 항목은 <표 4>와 같음
- 취합된 39개 키워드를 바탕으로 국민신문고와 기타 연구관련 간행물\*의 키워드 분석을 통해 키워드 간 관계도 및 위계관계 재정립(<그림 8>과 <그림 9> 참조)
  - \* 국가정책연구포털(NKIS), 온-나라 정책연구(PRISM), 국토연구원 보고서 등

**표 4** 전문가 설문을 통해 선정된 39개의 키워드 주제

지역계획, 지역정책, 균형, 개발, 도시계획, 토지, 용도변경, 정비, 건축허가, 산업, 특화, 인프라, 교통, 대중교통, 버스, 택시, 전철/지하철, 도로, 철도, 주택, 지역사회, 환경, 수질, 대기, 먼지, 농업용수, 전기, 가스, 에너지, 폐기물, 쓰레기, 상수도, 하수도, 도서관, 체육관, 문화관, 복지관, 어린이집
--

출처: 국토연구원의 내부회의를 거쳐 저자 작성.





### 3) 사례 분석결과

#### ① 키워드 위계 및 관계 분석결과

- 분석된 내용을 키워드별 위계에 따라 차등 비교(〈표 6〉 참조)
- 개별 키워드 검색을 통해 분석된 내용으로, **상위위계**에서는 ‘환경’, ‘지역사회’, ‘주택’ 세 가지 키워드에 대한 관심이 전국에 걸쳐 가장 높았던 것으로 집계 (부록 〈그림 1〉 참조)
  - 국토 전반에 걸쳐 ‘환경’ 및 ‘지역사회’와 관련된 뉴스기사가 압도적으로 높았고, 수도권 지역을 중심으로 ‘주택’과 관련된 언급이 상대적으로 높았던 것으로 확인
  - 수도권 지역으로 집중된 인구의 영향으로 인해 이처럼 높은 주거 관련 관심이 부각되었던 것으로 해석될 수 있는데, 실지로 국가통계포털의 주민등록인구는 수도권 인구가 2019년 12월 기준 대한민국 전체인구의 50%를 넘어서고 있었음을 증명하고 있음<sup>15)</sup>
- **중위위계**에서는 ‘개발’과 ‘산업’ 관련 키워드에 대한 관심이 가장 높게 나타났고, 지역에 따라서 ‘에너지’, ‘정비’, ‘도로’ 및 ‘교통’ 등으로 확인(부록 〈그림 2〉 참조)
  - ‘개발’과 관련된 키워드는 중앙지에서 압도적으로 높았으며, ‘산업’과 관련된 뉴스기사는 경상 및 호남 지역으로 갈수록 점진적으로 높아졌던 것으로 확인
  - 그밖에, ‘교통’ 관련 뉴스기사에 대한 관심은 인천·경기지방이 가장 높았으며, ‘도로’ 관련 뉴스기사는 강원, 경상 및 호남 지역을 중심으로 높게 나타남
- **하위위계**에서는 버스, 택시 등 ‘대중교통’과 관련된 키워드들이 전국에 걸쳐 주를 이뤘고, 그밖에 ‘어린이집’ 및 ‘도서관’ 등 생활 SOC와 관련된 키워드들이 주를 이룸(부록 〈그림 3〉 참조)
  - ‘버스’와 관련된 뉴스기사는 강원지역이 가장 높았고 인천·경기 지역을 중심으로 ‘지하철’에 대한 뉴스기사가 높게 집계
  - ‘어린이집’ 등 생활SOC 시설들에 대한 관심은 충청지역에서 상대적으로 높게 확인됐지만, 그밖에 인천·경기, 강원 및 경상 지역에서도 유사한 비율을 차지하고 있는 것으로 확인
- 종합적으로, 본 연구의 시간적 범위(지난 3개월) 동안 ‘환경’과 관련된 뉴스동향과 더불어, ‘지역사회’ 및 ‘주택’에 대한 관심이 눈에 띄었던 것으로 확인되었고, 국민의 소득과 직결된 ‘개발’ 및 ‘산업’, 이와 함께 ‘교통’ 및 생활 SOC 등의 주거복지 관련 키워드들의 관심이 복합적으로 두드러졌던 것으로 분석됨

14) 서울과 전국 언론매체에 대한 구분이 모호하기 때문에 이 연구에서는 ‘중앙지’의 개념을 사용함.

15) 국가통계포털 (<https://kosis.kr>, 2020년 11월 10일 검색).

**표 6 지역별 키워드 분석결과(상위위계 키워드 3개, 항목(%))**

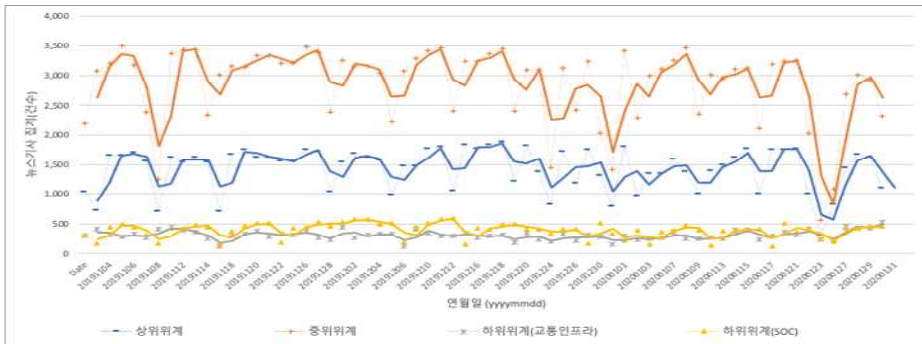
위계별 순위	중앙지 (서울)	인천·경기	강원	충청	경상	호남	제주
상위1	환경(43.8)	환경(44.7)	환경(37.7)	환경(42.7)	환경(42.4)	환경(43.8)	환경(47.2)
상위2	지역사회 (26.4)	지역사회 (21.7)	지역사회 (36.8)	지역사회 (34.8)	지역사회 (32.1)	지역사회 (33.8)	지역사회 (25.6)
상위3	주택(17.0)	주택(19.5)	주택(14.3)	주택(12.6)	주택(13.6)	주택(11.4)	주택(15.5)
중위1	개발(44.7)	개발(24.9)	개발(22.7)	개발(24.1)	개발(26.3)	개발(25.0)	개발(23.1)
중위2	산업(18.3)	산업(15.4)	산업(19.6)	산업(18.6)	산업(22.8)	산업(22.4)	산업(20.8)
중위3	에너지(6.2)	교통(13.6)	도로(10.8)	정비(9.4)	도로(7.6)	도로(7.7)	정비(9.3)
하위1	전기/버스 (14.1)	버스(15.0)	버스(19.5)	어린이집(13.2)	버스(14.6)	버스(14.6)	버스(14.6)
하위2	어린이집(10.5)	어린이집(12.0)	어린이집(11.0)	버스(13.0)	어린이집(12.4)	전기(13.5)	도서관(14.6)
하위3	대중교통(8.4)	지하철(10.3)	도서관(9.6)	도서관(12.3)	전기(12.0)	도서관(12.3)	전기(10.9)

출처: 저자 작성.

**② 개별 키워드 분석결과: 시계열 기반 동향분석**

- 위계별 키워드를 바탕으로 분석기간 동안의 시계열 동향분석을 한 결과, 2019년 12월 중순 (2019.12.18)을 기점으로 관련 뉴스보도가 점진적으로 줄어드는 경향이 확인(〈그림 10〉 참조)
  - 2019년의 마지막인 12월 31일과 연휴를 제외하고, 전반적으로 일정한 양의 보도가 다뤄졌던 것으로 확인
  - 중위위계 그룹의 키워드 관련 뉴스보도가 가장 많았고, 다음으로는 상위위계 키워드 그룹이 많은 양의 뉴스보도를 발표한 것으로 확인
  - 하위위계 키워드 그룹 분석은 키워드의 특성을 고려하여 교통인프라와 연관된 대중교통 및 택시 등의 키워드와 생활 SOC와 연관성이 높은 어린이집 및 도서관 등의 키워드를 구별하여 진행

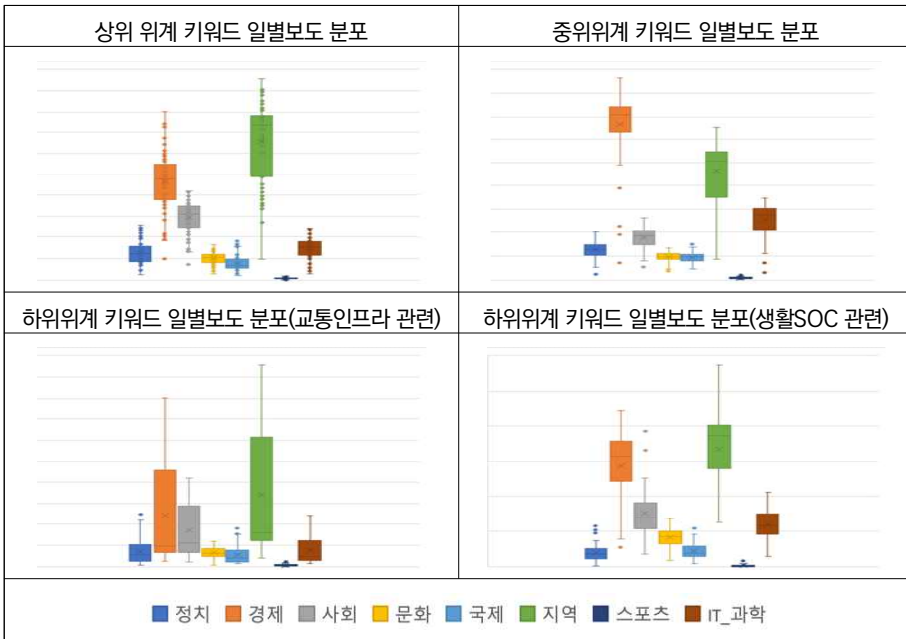
**그림 10 각 위계 키워드 그룹 시계열 동향분석결과**



주: 일별뉴스기사를 구간평균으로 재구성 / 출처: 저자 작성.

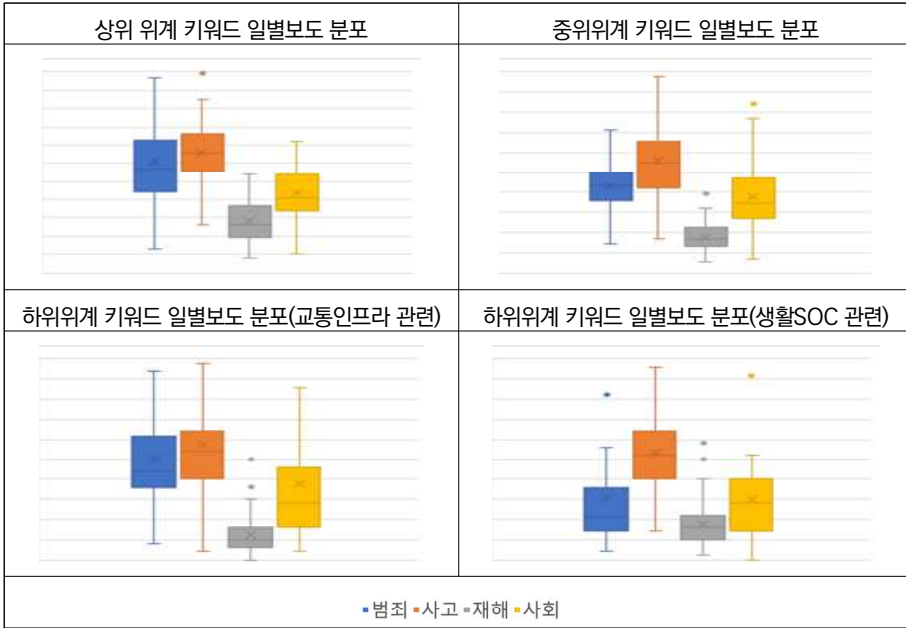
- 단어기반 특성화 및 집단화 과정을 통해 확인된 통합분류와 사건·사고 분석의 영역별 시계열분석을 확인
  - **통합분류 분석결과**, 지역사회, 경제, 그리고 사회 및 IT\_과학과 관련된 뉴스기사 상대적으로 눈에 띄는 것으로 확인(〈그림 11〉 참조)
  - 특히, 상위위계에서 뉴스보도는 지역>경제>사회 및 IT\_과학 순으로 집계되었고, 중위 위계에서는 경제>지역>IT\_과학, 하위위계에서는 지역>경제>사회 순으로 집계
  - **사건·사고 분석**의 분류결과, 상위위계 및 하위위계의 교통인프라 관련 보도에서 범죄 및 사고와 유관한 내용들이 다수 확인됐던 반면, 중위위계와 하위위계의 생활SOC 관련 보도에서는 사고와 연관된 내용들이 다수 확인이 되었던 것으로 집계(〈그림 12〉 참조)

**그림 11** 각 위계 키워드군 별 통합분류 일별보도 분포



주: 주말 제외.  
출처: 저자 작성.

**그림 12** 각 위계 키워드군 별 사건·사고 분류 일별보도 분포



주: 주말 제외.  
출처: 저자 작성.

- 키워드의 시계열 분석결과 사회의 특정 이슈에 따라 뉴스보도의 양이 눈에 띄게 나타나고, 그 뉴스보도에서 차용된 단어들의 특성에 따라 다양한 범주로 분류될 수 있음을 볼 수 있음
- 일례로, 상위위계 키워드군의 일별보도 중, 사건·사고 분류 중 범죄 분류에서 특정일에 눈에 띄는 뉴스기사보도가 확인(〈그림 13〉 참조)
  - 해당일(2020.12.16-17) 뉴스보도의 워드클라우드(Word Cloud) 확인결과, 정부의 부동산대책 전격 마련 발표 시기로 확인
  - 동 시기에 국가통계포털의 2019년 12월 기준 주민등록인구가 대한민국 전체 인구의 50%를 육박하는 시기로,
  - 각 뉴스기사 내용에서 차용된 단어들의 전반적인 성격 및 특성 등이 ‘투기’, ‘과열’ 등 강경한 어조의 단어선택이 잦아, 일반화 분류과정에서 ‘범죄’ 특성이 강한 사건·사고로 분류된 이유로 해석
  - 다른 한편으로는, 부동산 관련 사회 현안이 얼마만큼 민생에 민감하게 인식되는지를 확인할 수 있는 현상
- 그밖에도 중위 및 하위위계 키워드 시계열 분석에서도 범죄 및 사고 등으로 눈에 띄는 현상들이 확인(부록 〈그림 4~6〉 참조)
  - 중위위계 시계열 분석에서는 독도 부근 헬기추락 사건 및 산불 관련 사고가 두드러지게





- **(사건·사고 분류에 따른 검색결과)** 설정된 키워드들에 대해 ‘정치’ 및 ‘경제’에 해당하는 뉴스기사들이 전국 대상의 집계 모형에서는 높게 분포되었던 것으로 나타났고, 농업용수에 대한 사회적인 사건·사고가 두드러졌던 것으로 집계

### 지역별 매체의 검색결과 특성

- **(인천·경기)** 인천·경기 지역 중심의 언론매체 검색결과, 대기 및 먼지와 관련된 사회문제에 상대적으로 높은 관심이 있는 것으로 확인됐고, 상·하수도 및 대중교통, 문화관광과 복지관 관련 지역뉴스가 상대적으로 높은 관심을 가짐
- 사건·사고 분류에서는 농업용수와 사회의 관계, 복지관과 경제의 관계 및 지역계획과 정치의 관계가 상대적으로 유의하게 집계
- **(강원)** 강원지역 중심의 언론매체 검색결과, 키워드별 통합분류에서는 서울에 비해 개발, 용도변경, 대중교통, 버스, 농업용수, 상·하수도와 관련된 지역뉴스에 대한 관심이 높게 집계
- 키워드별 사건·사고에서는 복지관과 관련된 문화이슈에서부터 지하철과 관련된 경제이슈, 지역정책과 관련된 정치이슈에 관한 관심이 두드러짐
- **(충청)** 충청지역 중심의 언론매체 검색결과, 통합분류 분포는 강원지역 경우와 유사
- 사건·사고 분류 분포는 전기, 가스, 지역정책과 관련된 경제이슈와 농업용수 관련 사회 이슈가 기타 지역에 비해 두드러졌던 것으로 분석
- **(경상)** 경상지역 중심의 언론매체 검색결과, 통합분류 분포는 정치·경제·사회·문화 부분에서 경기의 분포와 유사하지만, 지역에 대한 관심은 도시계획, 토지, 용도변경, 상·하수도 등에 더 밀집된 분포를 가짐
  - 하지만 타 지역들에 비해서 두드러지게 집중되는 뉴스 분포는 나타나고 있지 않음
- 사건·사고 분류 분포는 전기, 가스, 상수도 및 문화관광과 관련된 경제이슈와 하수도, 농업용수와 관련된 사회이슈가 도드라지는 뉴스 분포를 나타냄
- 정치부분에서는 도시계획과 관련하여 다소 집중된 관심을 보임
- **(호남)** 호남지역 중심의 언론매체 검색결과, 통합분류 분포는 정치·경제·사회·문화 등의 분류에 도드라진 뉴스 분포는 나타나고 있지 않음
  - 하지만 상·하수도, 교통·대중교통, 토지, 용도변경 및 정비 등 지역과 관련된 뉴스에 다소 집중적인 뉴스 분포를 가짐

- 사건·사고 분류별 분포에서는, 경제이슈와 관련된 분포가 도드라짐
  - 특히, 도시계획, 용도변경, 건축허가, 체육관, 전기·가스 및 상수도에 관한 이슈들이 많았음
- 문화이슈와 관련해서는 문화관이 가장 눈의 띄는 뉴스 기록 수를 나타냄
- **(제주)** 제주 지역 중심의 언론매체 검색결과, 통합분류 분포는 국제, 문화와 관련하여 전철과 관련된 뉴스기사에서 다소 집중된 기록을 보임
- 지역과 관련된 분류에서는 지역정책, 도시계획, 농업용수, 쓰레기와 관련하여 상대적으로 높은 뉴스기록을 보임
- 사건·사고 분류별 분포에서는 정치이슈와 관련해서는 하수도와 지하철이 가장 도드라졌고, 경제이슈와 관련해서는 폐기물, 전기·가스, 수질, 지하철, 전철 및 특화와 관련된 뉴스 기사들이 상대적으로 많았던 것으로 집계
- 그밖에 사회이슈로는 농업용수가 있었고, 문화이슈로는 건축허가가 높은 뉴스의 관심을 받았던 것으로 집계

#### ④ 복합 키워드 분석결과

- 초기 39개 키워드로부터 산출된 최상위 키워드를 통해, 지역별로 n차 복합키워드 분석을 실시
- **(전국)** ‘개발·산업·환경’을 통해 산출된 2차 집중 키워드는 ‘IoT, 경쟁력, AI’ 등으로 집계, 이를 추가해 복합검색으로 집계된 집중 키워드는 ‘IoT, AI, 빅데이터’ 등으로 집계
  - 2020년 1월 6일 발표된 과학기술정보통신부의 발표에 따르면, 우리나라는 2020년의 목표로 데이터경제에 대한 본격적인 발돋움을 개시하였고<sup>16)</sup>,
  - 빅데이터, 데이터개방 및 활용에 대한 높은 확대계획을 발표한 바 있음
  - 이러한 정부발표에 대한 뉴스보도는, 2차 최상위 키워드군을 비롯한 4차 최상위 키워드군에서도 지속적으로 두드러지는 현상
- **(서울)** 서울 지역만을 따로 집계하면, AI 및 일자리에 대한 관심도가 높았고, 복합검색의 결과로는 사물인터넷과 미래 주거문화 모습과 관련된 미래 주거플랫폼에 대한 기사가 두드러짐
  - 중앙지 대부분의 이목이 집중되는 서울지역에서도 데이터경제의 주력산업으로 지목된 통신데이터(KT/SK/LG 통신사), 빅데이터 및 IoT, 인공지능과 관련된 키워드들이 다수 도출되었던 것으로 확인
- **(인천·경기)** 인공지능, 빅데이터 등을 시작으로, IoT 및 미세먼지에 관한 관심이 높게 집계

16) 과학기술정보통신부 보도자료 (2020년 1월 3일 배포).

- 교통에 대한 관심이 높았던 개별검색결과와 함께 인구 10만으로 집계되는 인천·경기 지역의 복합키워드 분석결과, 집중되는 인구유입으로 인한 환경에 대한 관심과 미세먼지 등의 우려가 높았던 것으로 확인
- **(강원)** 일련의 복합검색어 추이로 확인해 보았을 때, 일자리와 관련된 지역경제, 관광자원 개발에 관한 관심이 높았던 것으로 집계
  - 개별 키워드 검색을 통해 확인된 바대로, 개발 및 산업 등의 지역경제와 생활 SOC에 대한 높은 관심을 보인 지역
  - 지역경제 활성화를 위한 관광자원과 관련된 키워드 및 관광객 유치를 위한 대중교통의 접근성 관련 키워드들이 눈에 띄게 집계
- **(충청)** 기후변화와 관련하여 친환경 관련 뉴스에 대한 기록이 집계
  - 개별키워드 분석에서 다른 지역에서 집계된 관심의 정도와 유사한 비율을 보인 '환경' 관련 키워드가 복합키워드에서는 상대적으로 높은 비중을 차지
  - 환경과 관련된 지자체의 관심이 다른 지역에 비해 상대적으로 구체적이고 집중된 형태이기 때문으로 유추
- **(경상)** 조선업계와 관련된 산업체 일반과 ICT 및 인공지능에 대한 관심이 높게 집계
  - 개발 및 산업과 관련된 관심이 높게 확인이 되었던 개별키워드 분석과 유관한 복합키워드 분석결과가 도출
  - 지역경제 및 일자리와 직접적인 연관성이 높은 키워드들이 산출된 것으로 추정
- **(호남)** AI와 함께 일자리, 스마트에너지 등에 관한 뉴스집계가 높았던 것으로 집계
  - 개별키워드에서 관심을 보였던 '전기' 등의 에너지와 함께 '스마트 에너지'에 관한 뉴스보도가 두드러졌던 것으로 확인
  - 경상지역과 유사한 지역경제 및 일자리 관련 관심이 높게 분포
- **(제주)** 민생경제와 관련된 지자체의 지역경제 활성화 및 일자리 창출에 대한 관심이 높음
  - 본 사례분석의 지역별 구분 중 가장 높은 비중을 차지했던 제주지역의 '환경'에 대한 관심이, 복합키워드분석결과에서도 제주 제2공항 및 이착륙 활주로의 조류이동경로 우려와 같은 생태계 측면에 대한 관심이 높게 확인
  - 또한 제주 신공항과 함께 지역경제 활성화와 민생경제에 대한 우려의 목소리도 높았던 것으로 확인

### 3) 사례 분석결과 소결

- 키워드분석결과, 뉴스기사에서 차용된 단어의 어휘 및 어조 분석에 따라 전반적인 민심의 흐름과 언론보도에 대한 방향성이 확인
  - 예를 들어, 급격한 인구유입과 정부의 부동산대책 전격 마련시기가 맞물렸던 2019년 12월 중순의 경우, 뉴스보도 및 민심의 반응에 눈에 띄는 변화가 확인되었고,
  - 데이터경제에 대한 국가 차원의 계획발표가 있었던 2019년과 2020년 시기에는 빅데이터와 AI 등과 관련된 키워드들이 전국에 걸쳐 집중적으로 확인됨
- 개별 키워드 검색을 통해 분석된 내용과 복합키워드를 통해 분석된 내용 간 키워드의 결과에는 다소 차이가 있지만, 전반적인 지역별 우려 및 관심에 대한 흐름은 개별과 복합 분석 모두 유사한 것으로 확인
  - 개별 키워드 분석결과, 상위위계에서는 환경, 지역사회, 주택과 관련된 뉴스기사의 비중이 가장 높았고, 중위위계에서는 개발과 연관성이 높은 산업 및 정비 관련 언급이 주를 이뤘으며, 하위위계에서는 교통인프라 및 생활 SOC 등에 대한 관심이 높은 것으로 확인
  - 복합키워드 분석결과에서는 전국적으로는 환경과 주거 관련 관심과 함께, 새롭게 도약하는 정부의 데이터경제 관련 관심이 두드러지게 확인되었고,
  - 각 지자체에서는 지역경제와 일자리, 환경, 인공지능 관련 관심이 눈에 띄게 확인
- 키워드별 통합분류와 사건·사고 분류의 등고선 형태의 시각화를 통해서 지역별로 관심을 가지고 있는 키워드와 그에 따른 지역사회의 사회적인 인식도 정도를 함께 가늠해볼 수 있음
  - 제시된 키워드들의 통합분류 및 사건·사고 분류기준 집계에서는 대체적으로 지역과 관련된 뉴스분류와 함께 도시계획 및 상·하수도와 관련된 기사들이 다양하게 집계됨
  - 특히 사건·사고 분류기준 집계를 통하여 '농업용수' 등과 관련된 키워드들이 지방지역을 중심으로 사회 및 경제 분야에 큰 이슈였음을 확인
- 복합 키워드 검색을 통해 도출된 키워드를 분석해보면 개별 키워드와는 다소 구체적인 지역별 관심사들이 도출되었음을 확인
  - 전국적으로 IoT, AI, 빅데이터에 대한 관심이 높았지만, 지역적으로는 주거문화와 지역경제, 기후변화, 일자리 등 다양한 관심사가 포출

**표 7** 복합키워드 검색으로 분석된 각 차별 최상위 키워드군 결과

지역 구분	2차 최상위 키워드군	가중치	3차 최상위 키워드군	가중치	4차 최상위 키워드군	가중치
전국 (중앙지)	자동차	5.42	IoT	6.48	사물인터넷	20
	중소기업	4.71	경쟁력	6.39	AI	16.77
	업무협약	4.29	인공지능	5.86	빅데이터	6.67
서울	AI	4.5	인공지능	49.36	사물인터넷	15.56
	업무협약	4.24	KT/SK/LG 통신사	15.95	건설회사 (아파트 등)	14.55
	일자리	3.27	IoT/빅데이터	7.79	아파트 AI 플랫폼 (주거문화 모습)	7.62
인천, 경기	AI	6.09	인공지능	10.68	IoT	10.52
	업무협약	3.33	빅데이터	5.68	인구 10만	4.71
	경쟁력	3	지속가능	3.53	미세먼지	4.36
강원	사회복지	3.2	일자리창출	7.33	지역경제	3.91
	본격화	3.2	도시재생	4.8	비전 선편식	3.91
	활성화	3.08	관광객	4.36	관광자원	2.4
충청	지속가능	6.48	타운홀 미팅	11.82	지자체	12.77
	미세먼지	4.8	대응 계절관리제	7.28	기후변화	9.6
	활성화	4.42	온실가스 감축	7.03	전기버스	6.43
경상	자동차	6	조선	10.84	정보통신기술	9.02
	자율운항선박	4.32	고도화	5.05	석유화학	6.77
	일자리	4.07	ICT	4.05	인공지능	5.74
호남	인공지능/AI	14.37	자동차	10.62	도시철도 2호	5.68
	일자리	7.73	경제자유구역	8.67	스마트에너지	3.53
	상용화/R&D	6	조직개편	5.54	중심도시	3.47
제주	제주테크노파크	4.62	지역경제 활성화	5.09	관광객	3.2
	일자리 창출	4.36	JDC (제주국제자유도시개발센터)	3.53	제2공항	3.11
	지속가능	3.33	지자체	2.86	민생경제	3

주: \* 분석 뉴스 내에서 의미적 유사도가 높은 키워드 순(빅카인즈),  
출처: 저자 작성.

## 04 정책제언 및 향후 연구방향

- 이 연구는 국정운영과 국민 소통방식의 패러다임에 대응하기 위해 시도
  - 현재 운영 중인 전문가 자문 위주의 국정운영 연구과제 개발방법과 함께 국민의 소리가 보다 세밀하고 지속가능하게 환류할 수 있게 하기 위해 이 연구에서는 빅데이터와 ICT를 이용하여 ‘국민의 소리’ 모니터링을 시도
- 4차 산업혁명의 ICT 정책 패러다임 변화에 따라 대내·외 환경이 다양하게 변화 중
  - 시민의식의 변화와 체감형 정책수립을 위한 지속적인 정책과제 발굴이 요구
  - 국민의 소리 환류 지원정책과 더불어 지속가능하고 참신한 수단의 국토연구 아이디어 발굴의 필요성 증대
  - 이 연구에서는 국토와 관련된 키워드들을 중심으로 언론 빅데이터를 분석했고, 지역별로 지난 기간 동안 관심이 표출됐던 다양한 분야에 키워드를 도출
  - 개별 키워드 분석부터 복합 키워드 분석까지 도출된 지역별 키워드들을 중심으로 민생현안들과 각 연구 분야와의 접목 및 시기별·지역별 연구과제 발굴 시도가 필요
- 다양한 관점에서 언론데이터 키워드 추출과 머신러닝 등의 첨단기법 이용 또한 필요
  - 이 연구에서는 전문가 자문을 통해 집계된 39개 키워드를 이용하여 뉴스기사 분석을 실시했으나 코로나19, 주택시장, 경기침체 등 특정 민생현안 키워드를 중심으로 키워드를 확장 분석해나가는 방법 또한 정책연구과제 발굴에 필요한 접근
  - 이 연구에서는 특정 의미 범주 단어(개체명) 추출기법으로 머신러닝 기법 중 하나인 SSVM(Structured Support Vector Machine) 알고리즘을 활용(빅카인즈)
  - 이 연구에서 응용된 개체명 추출기법뿐만 아니라, 각 키워드별 트렌드 분석과 언론데이터

내용의 긍정·부정 분석 등 다양한 부분에 머신러닝 등의 최첨단 기술 접목이 후속 연구에서 시도될 필요

- 분석된 내용의 다양한 시각화 접근 또한 분석결과를 활용할 사용자들의 편의성 제고를 위해 지속적으로 시도돼야 함

#### ■ 연구의 한계점: 다양한 오픈소스 빅데이터의 활용이 요구

- 이 연구에서 사례분석으로 시도한 빅카인즈의 언론 빅데이터에서는 황명화(2014)에서 언급한 빅데이터의 관계망과 개별 매체 이용자 간의 사회관계망 등이 구별되지 못함
- 키워드 검색구문 및 키워드 분류방법의 변환을 통한 지역별 맞춤형 뉴스기사 도출방법 또한 보다 직관적이고 섬세한 지역관심사 도출을 위해 향후 연구에서 고려되어야 할 사항
- 국민신문고나 서울시 설문조사 등에서 다뤄지고 있는 분석방법에 비해 신문의 답변 또는 게시들을 통한 시민의식 반영 등은 데이터의 분포 특성상 표본집단 반영의 한계점 때문에 다뤄지지 못함
  - 이는 언론데이터가 가지고 있는 대중성과 이용자들에 대한 익명성 보장 등의 한계 때문이지만, 사회관계망(SNS)의 개념이 함께 접목 가능하다면, 이 연구에서 도출해낸 지역별 키워드 등의 민생현안에 어떠한 국민의 정서·관심이 집중되고 있는지 알 수 있을 뿐만 아니라 국가정책 발굴을 위해 후속 연구로 심도 있게 접근해봐야 함

32

#### ■ 정책제언

- 지역별 언론매체 모니터링을 통한 사회이슈에 대한 국민의 지속적인 현안 및 관심도 파악과 선제적 정책방안 발굴
- 현재 국토교통부 및 경제·인문사회연구회 등에서 개방형 정책연구 발굴 시스템으로 제안되는 키워드들을 이용한 모니터링 시도, 지역별 민생현안 파악을 통한 정책 필요 및 중요성 반영
- 최근 늘어나고 있는 국토·도시계획 관련 민원 시스템과 리빙랩(Living lab)<sup>17)</sup> 등의 민원참여 시스템의 비교분석을 통한 연계·확장형 민생현안 모니터링 발굴 필요

17) 사용자가 적극적으로 혁신활동에 참여 가능한 사용자 주도 개방형 혁신 행태계(성지는 외 2013).

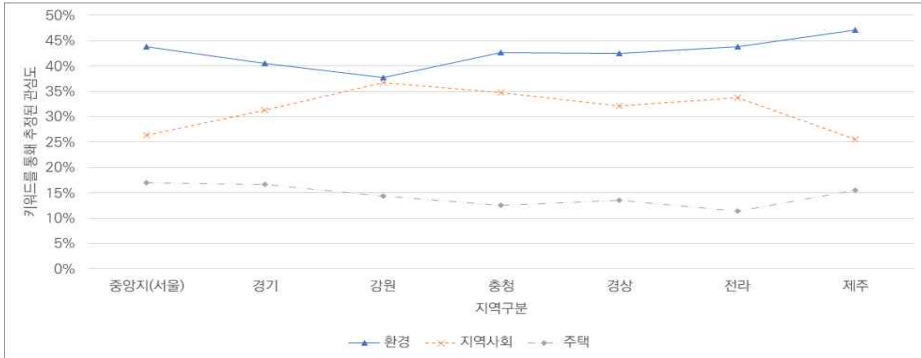


## 참고문헌

- 국가지표체계. [www.index.go.kr](http://www.index.go.kr) (2020년 3월 5일 검색).
- 국립국어원. [www.korean.go.kr](http://www.korean.go.kr) (2020년 4월 1일 검색).
- 국민권익위원회. <http://www.acrc.go.kr/acrc/board.do?command=searchDetail&menuId=05040705> (2020년 4월 3일 검색).
- 김대중·윤서연. 2013. 빅데이터를 활용한 국토정책반응 모니터링 및 정책수요 예측방안. 안양: 국토연구원.
- 김동한·강혜경·김준기·한우석·박준·강민규·성해정·김은빈. 2014. 공간 빅데이터를 활용한 국토도시 정책방안 연구. 안양: 국토연구원.
- 김미정·김대중·이영주. 2013. 과학적 국토정책을 위한 공간빅데이터 활용방안. 안양: 국토연구원.
- 빅카인즈. <https://www.bigkinds.or.kr> (2020년 4월 3일 검색).
- \_\_\_\_\_. 2019. 빅카인즈 사용자 매뉴얼. 서울: 한국언론진흥재단.
- 서울특별시 여론조사. <http://research.seoul.go.kr/surveyWebGuide.do> (2020년 4월 1일 검색).
- 서울특별시 응답소. [http://eungdapso.seoul.go.kr/Gud/minwon\\_status.jsp](http://eungdapso.seoul.go.kr/Gud/minwon_status.jsp). (2020년 4월 3일 접속)
- 성지은·송위진·박인용. 2013. 리빙랩의 운영 체계와 사례. STEPI Insight 제127호. 서울: STEPI.
- 이영주·황명화, Ryosuke Shibasaki, 임은선·박재희·김동한·김종학·이현아, Yuki Akiyama, Yuka Nishimoto. 2016. 인구감소시대의 지역문제 진단을 위한 공간 빅데이터 활용방안 연구. 안양: 국토연구원
- 정영철·이기호·배하석·신철민. 2012. 보건복지 부문의 소셜미디어 활용 현황 및 정책과제. 한국보건사회연구원
- 황명화. 2014. 여론 모니터링을 위한 비정형 빅데이터 시공간분석방법론 연구. 안양: 국토연구원.
- 국가통계포털. <https://kosis.kr> (2020년 11월 10일 검색).
- 과학기술정보통신부. 2020. 우리나라 데이터 경제가 본격적으로 열립니다. 1월 3일, 보도자료.
- Gouvernement UK. <https://www.gov.uk/government/publications/uk-digital-strategy> (2020년 4월 3일 검색).
- Groups Google. <https://groups.google.com/forum/#!forum/us-open-government> (2020년 4월 3일 검색).
- Laney, Doug. 2001. 3D data management: Controlling data volume, velocity and variety. *META group research note* 6, no.70: 1.
- Literature and Latte, Scapple. <http://www.literatureandlatte.com> (2020년 4월 3일 검색).
- Merriam-Webster. <https://www.merriam-webster.com> (2020년 4월 1일 검색).
- Obama white house. <https://obamawhitehouse.archives.gov/open> (2020년 4월 3일 검색).
- Petitions, Obama white house. <https://petitions.obamawhitehouse.archives.gov/> (2020년 4월 3일 검색).

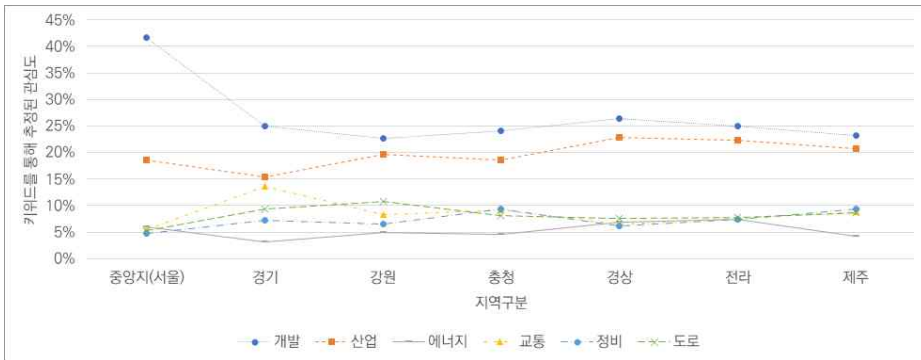


**부록 그림 1** 상위 위계 키워드 중 최상위 키워드 그룹의 지역별 차이



출처: 저자 작성.

**부록 그림 2** 중위 위계 키워드 중 최상위 키워드 그룹의 지역별 차이



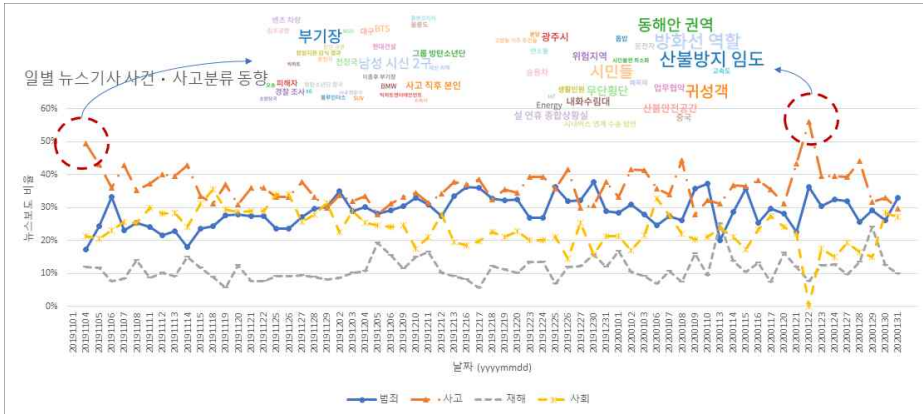
출처: 저자 작성.

**부록 그림 3** 하위 위계 키워드 중 최상위 키워드 그룹의 지역별 차이



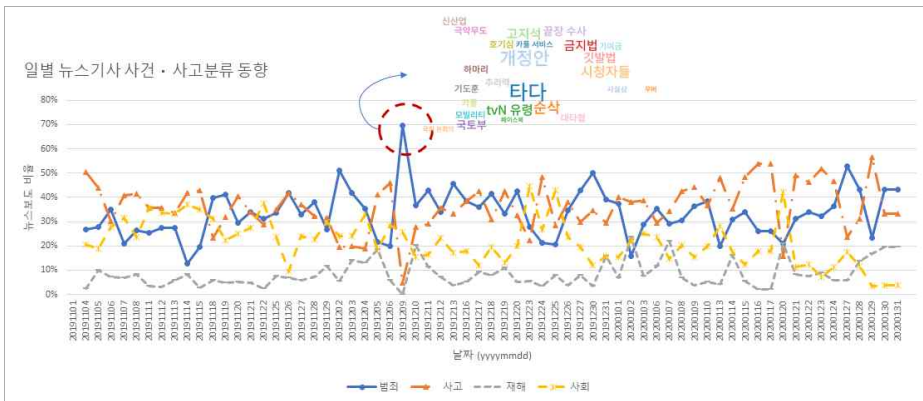
출처: 저자 작성.

**부록 그림 4 중위위계 키워드군 사건·사고 분류 일별보도**



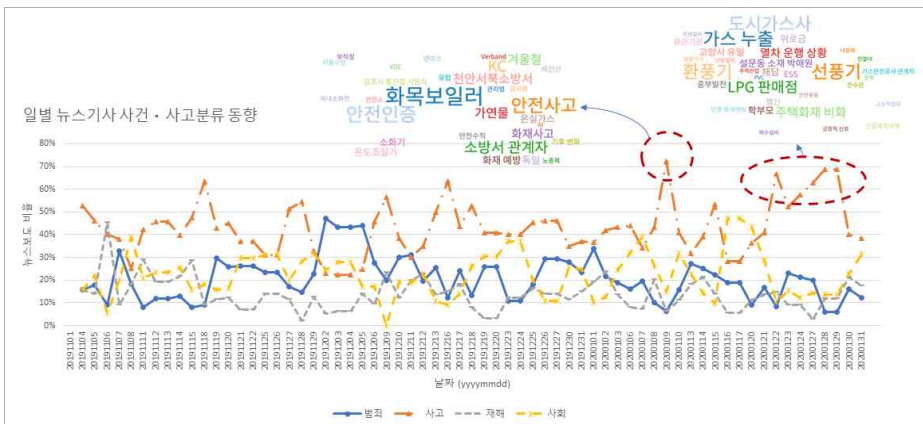
주: 주말 제외; 출처: 저자 작성.

**부록 그림 5 하위위계 키워드군 사건·사고 분류 일별보도(교통 인프라 관련)**



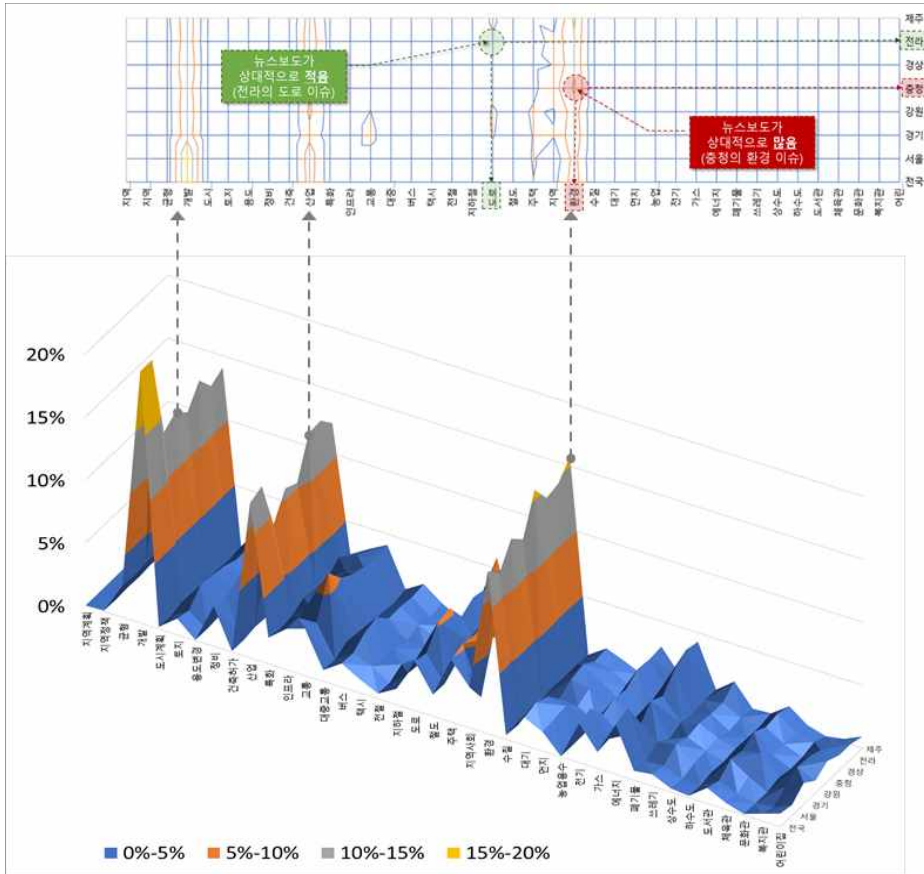
주: 주말 제외; 출처: 저자 작성.

**부록 그림 6 하위위계 키워드군 사건·사고 분류 일별보도(생활 SOC 관련)**



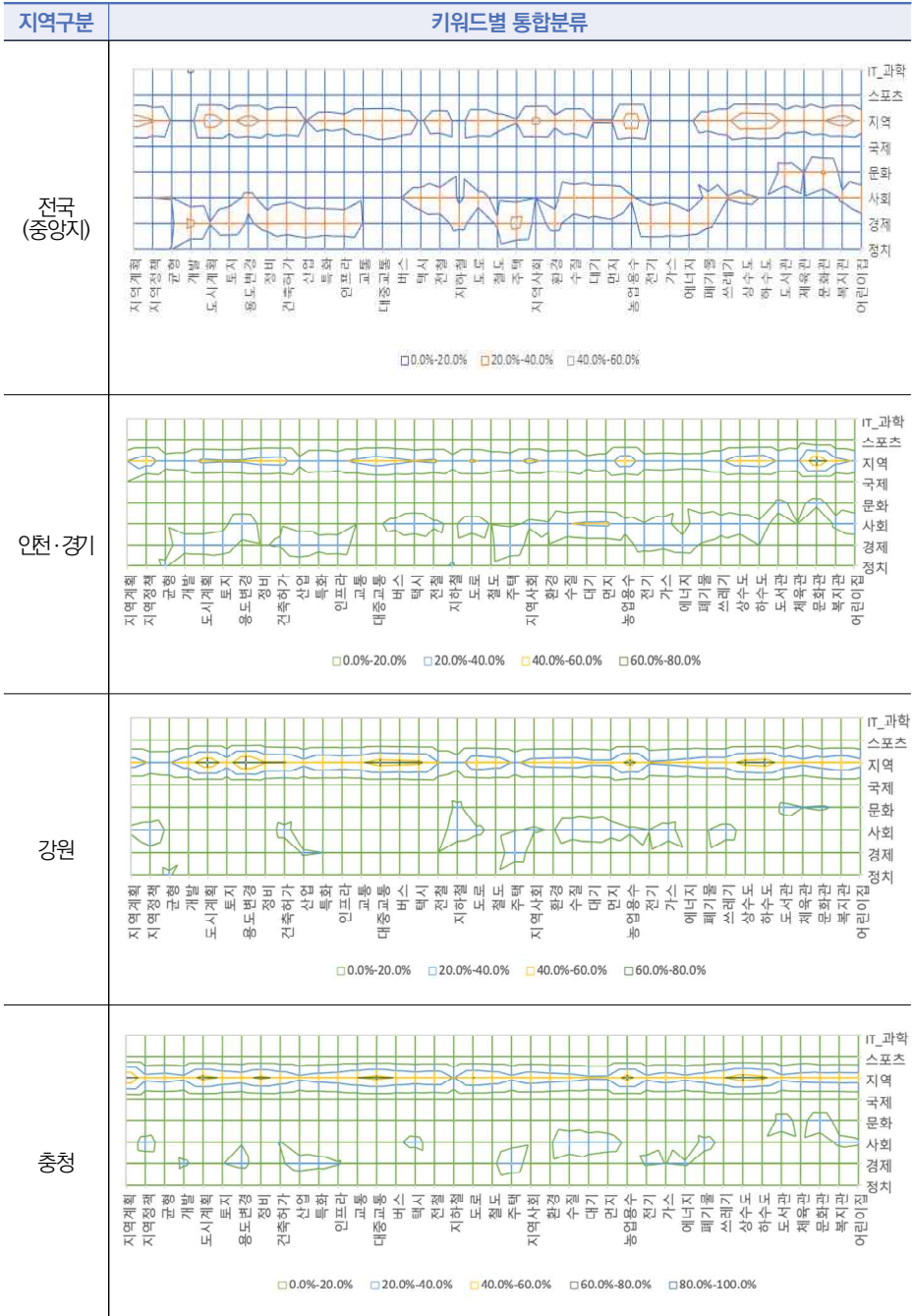
주: 주말 제외; 출처: 저자 작성.

부록 그림 7 뉴스기사의 키워드별 분포에 따른 시각화 예시



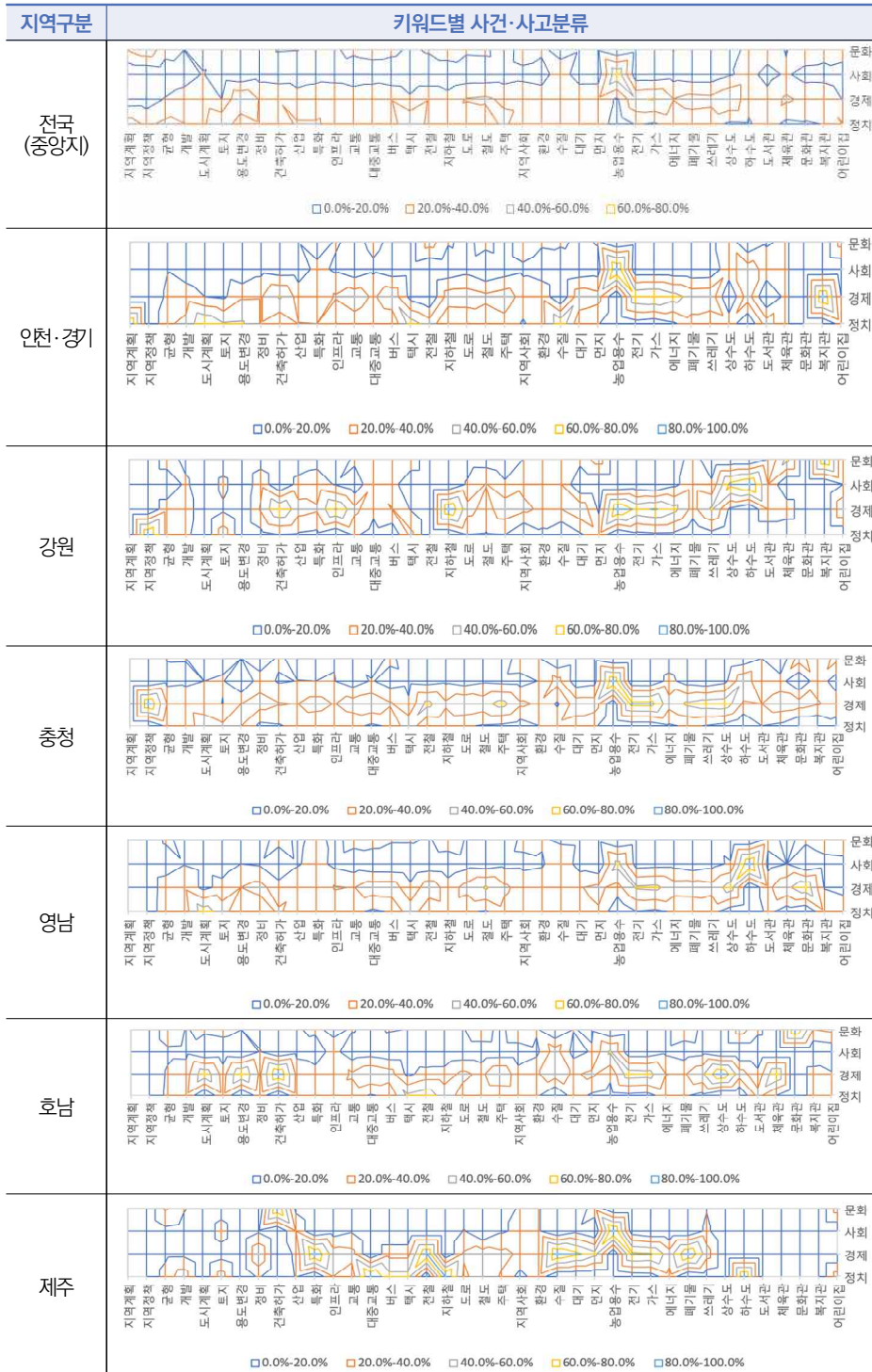
주: 종축(세로축)은 각 키워드를 의미하고, 횡축(가로축)은 각 지역을 의미, 각 횡축과 종축이 교차되는 지점이 해당 지역에 해당 키워드의 빈출빈도를 의미.  
출처: 저자 작성.

**부록 그림 8** 지역 매체별 키워드 검색결과 - 통합분류에 따른 분포 (별첨 그림 8 참조)





부록 그림 9 지역 매체별 키워드 검색 결과 - 사건·사고 분류에 따른 분포 (별첨 그림 8 참조)



주: 종축은 각 키워드, 횡축은 각 지역을 의미, 각 횡축과 종축이 교차되는 지점이 해당 지역에 해당 키워드의 빈출빈도를 의미함.  
출처: 저자 작성.



국토연구원 Working Paper는 다양한 국토 현안에 대하여 시의성 있고 활용도 높은 대안을 제시할 목적으로 실험정신을 가지고 작성한 짧은 연구물입니다. 투고된 원고는 정해진 절차를 거쳐 발간되며, 외부 연구자의 투고도 가능합니다. 공유하고 싶은 새로운 이론이나 연구방법론, 국토 현안이나 정책에 대한 찬반 논의, 국내외 사례 연구나 비교연구, 창의적 제안 등 국토분야 이론과 정책에 도움이 될 어떠한 연구도 환영합니다.

투고를 원하시는 분은 국토연구원 연구기획·평가팀(044-960-0582, jhkim@krihs.re.kr)으로 연락주시시오. 채택된 원고에 대해서는 소정의 원고료를 드립니다.

## WP 20-09

### 빅데이터를 이용한 국토 민생현안 모니터링 연구

**연 구 진**      장요한  
**발 행 일**      2020년 12월 15일  
**발 행 인**      강현수  
**발 행 처**      국토연구원  
**홈페이지**    <http://www.krihs.re.kr>

---

© 2020, 국토연구원

---

이 연구보고서의 내용은 국토연구원의 자체 연구물로서 정부의 정책이나 견해와는 상관없습니다.

---

이 연구보고서는 한국출판인협회에서 제공한 KoPub 서체와 대한인쇄문화협회가 제공한 바른바탕체가 적용되어 있습니다.

