

정책자료 2016-02

2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석



송태민 · 이기호 · 진달래 · 천미경 · 서동철 · 박현애

【책임연구자】

송태민 한국보건사회연구원 선임연구위원

【주요 저서】

R을 활용한 소셜 빅데이터 연구방법론
한나레아카데미, 2016(공저)

빅데이터 분석 방법론
한나레아카데미, 2013(공저)

【공동연구진】

이기호 한국보건사회연구원 부연구위원

진달래 한국보건사회연구원 연구원

천미경 한국보건사회연구원 연구원

서동철 인디애나주립대학교

박현애 서울대학교 간호대학

정책자료 2016-02

2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

발행일 2016년 12월 31일

저자 송태민

발행인 김상호

발행처 한국보건사회연구원

주소 [30147] 세종특별자치시 시청대로 370
세종국책연구단지 사회정책동(1~5층)

전화 대표전화: 044)287-8000

홈페이지 <http://www.kihasa.re.kr>

등록 1994년 7월 1일 (제8-142호)

인쇄처 (주)법신사

정가 비매품

© 한국보건사회연구원 2016
ISBN 978-89-6827-333-9 93510

발간사 <<

최근 스마트폰, 스마트TV, RFID, 센서 등의 급속한 보급과 모바일 인터넷 및 소셜미디어의 확산으로 데이터양이 기하급수적으로 증가하고 데이터의 생산, 유통, 소비 체계에 큰 변화를 주면서 데이터가 경제적 자산이 될 수 있는 빅데이터 시대를 맞이하게 되었다. 세계 각국의 정부와 기업들은 빅데이터가 향후 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하고 있으며, The Economist, Gartner, McKinsey 등은 빅데이터를 활용한 시장 변동 예측과 신사업 발굴 등 경제적 가치 창출 사례 및 효과를 제시하고 있다. 특히 빅데이터는 미래 국가 경쟁력에도 큰 영향을 미칠 것으로 기대돼 각 국가는 안전을 위협하는 글로벌 요인이나 테러, 재난·재해, 질병, 위기 등에 선제적으로 대응하기 위해 우선적으로 도입하고 있다.

한국에서는 최근 정부3.0과 창조경제를 추진·실현하고 현 정부의 주요 정책과제를 지원하기 위해 다양한 분야에서 빅데이터의 활용 가치가 강조되고 있다. 빅데이터는 데이터의 형식이 다양하고 방대할 뿐만 아니라 그 생성 속도가 매우 빨라 기존의 데이터를 처리하던 방식이 아닌 새로운 관리 및 분석 방법을 요구한다. 또한 트위터, 페이스북 등 소셜미디어에 남긴 정치, 경제, 사회, 문화에 대한 메시지는 그 시대의 감성과 정서를 파악할 수 있는 원천으로 등장함에 따라, 대중매체에 의해 수립된 정책의 제는 이제 소셜미디어로부터 파악할 수 있으며 개인이 주고받은 수많은 댓글과 소셜 로그 정보는 공공정책을 위한 공공재로서 진화 중이다.

이와 같이 많은 국가와 기업에서는 SNS로 인해 생산되는 소셜 빅데이

터의 활용과 분석을 통하여 새로운 경제적 효과와 일자리 창출은 물론 사회적 문제의 해결을 위해 적극적으로 노력하고 있다.

이에 본 연구는 소셜 빅데이터를 기반으로 우리나라 보건복지 주요 이슈에 대한 동향을 분석하기 위한 것으로, 2016년의 주요 이슈로 보건복지 정책 미래신호 예측, 담배, 통일, 어린이 예방접종 온톨로지 개발, 지카바이러스 온톨로지 개발의 5개 주제에 대해 소셜 빅데이터를 수집하여 동향 분석과 수요 예측 및 위험 예측 모형을 제시하였다. 이와 같은 연구는 보건복지 주요 이슈를 적시에 분석하고 수요를 예측할 수 있다는 점에서 정책적·분석방법론적으로 의의가 있으며, 실제적으로 내용을 더욱 빠르게 효과적으로 파악하여 사회조사가 지닌 한계를 보완할 수 있는 새로운 조사 방법으로서의 소셜 빅데이터 가치를 확인하였다는 점에서 조사 방법론적 의의를 가진다고 할 수 있다. 본 연구에 귀중한 조언을 아끼지 않으신 많은 전문가분들과 주요 이슈에 대한 논문 집필에 참여해 주신 교수님들께 감사드린다. 끝으로 본 보고서에 수록된 모든 내용은 우리 연구원의 공식적인 견해는 아니며 어디까지나 참여한 연구진의 의견임을 밝힌다.

2016년 12월

한국보건사회연구원 원장

김 상 호

목 차

| | |
|---|-----------|
| Abstract | 1 |
| 요 약 | 3 |
| 제1장 소셜 빅데이터를 활용한 보건복지 정책 미래신호 예측 | 7 |
| 제1절 서론 | 9 |
| 제2절 소셜 빅데이터 기반 미래신호 예측 방법론 | 10 |
| 제3절 소셜 빅데이터 기반 보건복지 정책 미래신호 예측 | 11 |
| 참고문헌 | 27 |
| 제2장 2016 소셜 빅데이터를 활용한 통일인식 동향 분석 및 예측 29 | |
| 제1절 서론 | 31 |
| 제2절 연구 방법 | 33 |
| 제3절 연구 결과 | 38 |
| 제4절 결론 | 58 |
| 참고문헌 | 63 |
| 제3장 2016 소셜 빅데이터를 활용한 담배 위험 예측 | 65 |
| 제1절 서론 | 67 |
| 제2절 연구 방법 | 70 |
| 제3절 연구 결과 | 75 |
| 제4절 결론 | 87 |
| 참고문헌 | 91 |

제4장 지카바이러스 감염증 소셜 빅데이터 분류를 위한 온톨로지 개발 ... 95

| | |
|------------------------------------|-----|
| 제1절 서론 | 97 |
| 제2절 지카바이러스 소셜 빅데이터 온라인 문서 현황 | 109 |
| 제3절 연구 결과 | 123 |
| 제4절 결론 | 125 |
| 참고문헌 | 131 |

제5장 어린이 예방접종 소셜 빅데이터 분류를 위한 온톨로지 개발 ... 135

| | |
|----------------------------|-----|
| 제1절 서론 | 137 |
| 제2절 어린이 예방접종 온톨로지 개발 | 141 |
| 제3절 연구 결과 | 154 |
| 제4절 논의 | 185 |
| 제5절 결론 | 187 |
| 참고문헌 | 189 |

표 목차

| | |
|--|-----|
| 〈표 1-1〉 온라인 채널의 보건복지 정책, 이슈의 키워드 분석 | 13 |
| 〈표 1-2〉 보건복지 정책, 이슈의 DoV 평균증가율과 평균단어빈도 | 15 |
| 〈표 1-3〉 보건복지 정책, 이슈의 DoD 평균증가율과 평균단어빈도 | 16 |
| 〈표 1-4〉 보건복지 관련 정책, 이슈의 미래신호 | 19 |
| 〈표 1-5〉 보건복지 수요에 영향을 주는 주요 정책 요인 | 21 |
| 〈표 1-6〉 보건복지 주요 정책의 연관규칙 | 23 |
| 〈표 2-1〉 통일인식의 감정 키워드 연관성 예측 | 42 |
| 〈표 2-2〉 통일 관련 문서(버즈) 현황 | 46 |
| 〈표 2-3〉 연도별 통일 관련 국민 인식 | 48 |
| 〈표 2-4〉 통일 관련 안보이슈의 국가별 버즈 현황 | 49 |
| 〈표 2-5〉 통일 관련 안보이슈의 기관별 버즈 현황 | 50 |
| 〈표 2-6〉 안보와 이슈 요인에 대한 통일인식의 연관성 예측 | 51 |
| 〈표 2-7〉 주변 국가(4국)의 통일인식의 연관성 예측 | 53 |
| 〈표 2-8〉 통일인식에 영향을 미치는 안보이슈 요인 | 54 |
| 〈표 2-9〉 통일인식 관련 안보이슈 요인의 예측 모형에 대한 이익도표 | 57 |
| 〈표 3-1〉 담배 관련 온라인 문서 현황 | 77 |
| 〈표 3-2〉 연도별 담배 관련 감정 변화 | 80 |
| 〈표 3-3〉 요인별 담배 위험 예측 | 81 |
| 〈표 3-4〉 담배의 위험군과 잠재군에 영향을 미치는 정책 및 도구 요인 | 84 |
| 〈표 3-5〉 정책 요인의 예측 모형에 대한 이익도표 | 86 |
| 〈표 4-1〉 국내 지카바이러스 확진자 현황 | 104 |
| 〈표 4-2〉 발생 구분별 지카바이러스 감염 국가 | 105 |
| 〈표 4-3〉 성관계를 통한 지카바이러스 감염 국가 | 106 |
| 〈표 4-4〉 지카바이러스 소셜 빅데이터 수집을 위한 분류 | 110 |
| 〈표 4-5〉 지카바이러스 소셜 빅데이터 수집을 위한 분류 | 123 |
| 〈표 5-1〉 Competency questions | 144 |

| | |
|---|-----|
| 〈표 5-2〉 연구 개념 틀에서 제시한 관련 요인들의 내용 분류 | 147 |
| 〈표 5-3〉 예방접종 관련 게시 글의 주제 분류 | 151 |
| 〈표 5-4〉 인적요인의 하위분류 및 용어체계 | 157 |
| 〈표 5-5〉 인적요인의 데이터모델링 예시 | 158 |
| 〈표 5-6〉 구조사회문화요인의 하위분류 및 용어체계 | 160 |
| 〈표 5-7〉 구조사회문화요인의 데이터모델링 예시 | 161 |
| 〈표 5-8〉 정보요인의 하위분류 및 용어체계 | 162 |
| 〈표 5-9〉 정보요인의 데이터모델링 예시 | 165 |
| 〈표 5-10〉 환경요인의 하위분류 및 용어체계 | 166 |
| 〈표 5-11〉 환경요인의 데이터모델링 예시 | 168 |
| 〈표 5-12〉 예방접종 의도의 하위분류 및 용어체계 | 169 |
| 〈표 5-13〉 예방접종 의도의 데이터모델링 | 170 |
| 〈표 5-14〉 예방접종 행위의 하위분류 및 용어체계 | 171 |
| 〈표 5-15〉 예방접종 행위의 데이터모델링 예시 | 172 |
| 〈표 5-16〉 행동반응요인의 하위분류 및 용어체계 | 173 |
| 〈표 5-17〉 행동반응요인의 데이터모델링 예시 | 175 |
| 〈표 5-18〉 Competency questions와 DL-Query | 176 |
| 〈표 5-19〉 어린이 예방접종 빅데이터 연구 도구(키워드) | 179 |
| 〈표 5-20〉 어린이 예방접종 관련 온라인 문서 현황 | 183 |

그림 목차

| | |
|---|----|
| [그림 1-1] 보건복지 관련 정책, 이슈의 KEM(Keyword Emergence Map) | 18 |
| [그림 1-2] 보건복지 관련 정책, 이슈의 KIM(Keyword Issue Map) | 18 |
| [그림 1-3] 랜덤 포리스트 모델의 보건복지 주요 정책의 중요도 | 20 |
| [그림 1-4] 보건복지 정책 수요 예측의 의사결정나무 모형 | 22 |
| [그림 2-1] 통일 관련 문서(버즈)량 일별 추이: 2016년 1~3월 | 39 |

| | |
|--|-----|
| [그림 2-2] 통일 연도별 감정 변화 | 40 |
| [그림 2-3] 통일인식 감정의 연관규칙에 대한 병렬좌표 시각화 | 45 |
| [그림 2-4] 통일인식 관련 안보·이슈 요인의 예측 모형 | 56 |
| [그림 3-1] 월별, 시간별 담배 관련 온라인 문서 현황 | 76 |
| [그림 3-2] 담배 관련 요일별 온라인 문서 현황 | 76 |
| [그림 3-3] 월별 담배 관련 질환별 온라인 문서 현황 | 77 |
| [그림 3-4] 월별 담배 위험군(담배를 애호적으로 생각하는 감정 집단) 감정 변화 | 79 |
| [그림 3-5] 월별 담배 일반군(담배를 혐오적으로 생각하는 감정 집단) 감정 변화 | 79 |
| [그림 3-6] 정책 요인의 예측 모형 | 86 |
| [그림 4-1] SIR 모형 | 101 |
| [그림 4-2] SEIR 모형 | 102 |
| [그림 4-3] 모기에 의한 바이러스 이동 경로 | 103 |
| [그림 4-4] 지카바이러스 감염 경로 | 104 |
| [그림 4-5] WHO 지역 구분별 지카바이러스 감염 누적 국가 수 | 107 |
| [그림 4-6] 지카바이러스 확산 현황 | 108 |
| [그림 4-7] 지카바이러스 온톨로지 분류 | 127 |
| [그림 4-8] 생태학적 관점의 지카바이러스 온톨로지 분류체계 | 127 |
| [그림 5-1] 어린이 예방접종 온톨로지 개발 과정 | 142 |
| [그림 5-2] 예방접종 관련 게시 글에 포함된 용어의 word cloud | 152 |
| [그림 5-3] 어린이 예방접종 온톨로지 개념 틀 | 155 |
| [그림 5-4] 어린이 예방접종 온톨로지 | 175 |
| [그림 5-5] Competency questions 평가 과정 예시 | 178 |



Abstract ‹‹

Social big data trend analysis based on health and welfare issues in 2016

Big data consists of various forms of data in large volumes, which are rapidly created. Therefore, it requires a new management and analysis methodology. Moreover, as social media platforms emerge as the source of information about the feelings and sentiments of the current times with messages on politics, economics, society, and culture, policy agendas set up in public spheres can be identified from social media. Countless comments between individuals and society evolve as a log of information, which continues to evolve as an asset for public policy. As such, many nations and businesses actively strive for new economic effects, job creation, as well as solving social problems through the utilization and analysis of social big data created through social networking services(SNS). This study proposes study methodologies and utilization strategies of social big data that can create value and predict the future by gathering and analyzing social big data from various fields.



1. 연구의 배경 및 목적

스마트폰, 센서, 모바일 인터넷 그리고 소셜미디어 등의 급속한 보급과 확산으로 데이터양이 비약적으로 증가하면서 데이터가 경제적 자산이 될 수 있는 빅데이터가 도래하였다. 빅데이터는 미래 국가 경쟁력에도 큰 영향을 미칠 것으로 기대돼 각 국가는 안전을 위협하는 글로벌 요인이나 테러, 재난·재해, 질병, 위기 등에 선제적으로 대응하기 위해 우선적으로 도입하고 있다. 빅데이터는 데이터의 형식이 다양하고 방대할 뿐만 아니라 그 생성 속도가 매우 빨라 기존의 데이터를 처리하던 방식이 아닌 새로운 관리 및 분석 방법을 요구한다. 또한 트위터, 페이스북 등 소셜미디어에 남긴 정치, 경제, 사회, 문화에 대한 메시지는 그 시대의 감성과 정서를 파악할 수 있는 원천으로 등장함에 따라, 대중매체에 의해 수립된 정책의 제는 이제 소셜미디어로부터 파악할 수 있으며 개인이 주고받은 수많은 댓글과 소셜 로그 정보는 공공정책을 위한 공공재로서 진화 중이다. 이와 같이 많은 국가와 기업에서는 SNS로 인해 생산되는 소셜 빅데이터의 활용과 분석을 통하여 새로운 경제적 효과와 일자리 창출은 물론 사회적 문제를 해결하기 위해 적극적으로 노력하고 있다.

본 연구에서는 다양한 분야의 소셜 빅데이터를 수집·분석하여 가치를 창출하고 미래를 예측할 수 있는 소셜 빅데이터 연구 방법과 활용 방안을 제시하였다. 그리고 2016년의 주요 보건복지 이슈로 보건복지 정책, 통일, 담배, 지카바이러스, 어린이 예방접종에 대한 소셜 빅데이터를 수집하여 동향 분석과 수요 예측 및 위험 예측 모형을 제시하였다.

2. 주요 연구 결과

제1장 ‘소셜 빅데이터를 활용한 보건복지 정책 미래신호 예측’에서는 소셜 빅데이터 기반 미래 예측 방법론을 사용하여 보건복지와 관련하여 나타나는 주요 정책과 이슈에 대한 미래신호 예측 모형을 제시하였다. 우리나라 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 기반으로 보건복지 분야의 각 정책 및 이슈별로 강신호와 약신호로 분류하였다. 특히 복지급여와 일자리가 동시에 언급된 문서의 찬성 비율이 매우 높은 것으로 나타나 능동적 복지체계 구축을 통한 일자리 창출이 필요한 것으로 보이며, 건강보험 및 중증질환이 동시에 언급된 문서 또한 찬성 비율이 높은 것으로 나타나 건강보험 혜택 확대로 4대 중증질환의 보장성 강화가 국민 의료비를 줄임으로써 정부 정책에 대한 좋은 평가 결과로 나타난 것으로 확인되었다.

제2장 ‘2016 소셜 빅데이터를 활용한 통일인식 동향 분석 및 예측’에서는 우리나라 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 바탕으로 우리나라 국민의 통일인식에 대한 동향을 분석하고 통일인식의 예측 모형과 연관규칙을 파악하였다. 우리나라 안보와 이슈에 대한 통일인식의 연관성 예측 분석 결과 통일대박이 언급되었을 때 언급되지 않은 문서보다 통일에 찬성할 확률이 0.47배 높아지는 것으로 나타났으며 핵무기, 천안함이 동시에 언급되었을 경우 언급되지 않은 문서보다 통일에 반대할 확률이 5.92배 높아지는 것으로 나타났다.

제3장 ‘2016년 소셜 빅데이터를 활용한 담배 위험 예측’에서는 우리나라 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 바탕으로 우리나라 국민의 담배에 대한 위험 예측 모형과 연

관규칙을 파악하였다. ‘담뱃값’과 ‘흡연 규제’가 동시에 온라인에서 언급 되면 일반군이 될 확률이 증가하며, ‘담뱃값 인상’만 언급되어도 위험군을 감소시키는 것으로 나타났다.

제4장 ‘지카바이러스 감염증 소셜 빅데이터 분류를 위한 온톨로지 개발’에서는 우리나라 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 수집하기에 앞서 용어 추출에 대한 다수준적 접근을 위하여 생태학적 체계 이론을 바탕으로 주제 분류를 하여 온톨로지 용어를 구축하였다. 추출된 용어를 기반으로 2016년 1월 1일부터 4월 15일까지 빅데이터를 수집하였으며, 지카바이러스와 관련하여 긍정적인 감정을 나타내는 온라인 문서는 36.1%, 보통의 감정을 나타내는 문서는 12.2%, 부정적인 감정을 나타내는 문서는 51.6%로 나타났으며, 지카바이러스 대처 방법에 대해서는 초기 대응, 정부 대응, 감염 검사 등의 순으로 확인되었다.

제5장 ‘어린이 예방접종 소셜 빅데이터 분류를 위한 온톨로지 개발’에서는 어린이 예방접종에 대한 부모의 감정, 경험, 태도, 지식과 관련된 용어를 수집하고, 용어로부터 관련 개념을 추출하여 추출된 개념의 데이터 모델을 개발하였으며, 각 영역 간의 관계를 나타내는 온톨로지와 유의어 및 동의어를 포함한 용어체계를 개발·구축하였다. 예방접종 온톨로지는 부모와 자녀 간의 개인 요인에서부터 구조적·사회적·문화적·환경적 요인을 모두 포함하고 있으며, 예방접종 이전부터 예방접종을 하는 시점, 예방접종을 하고 난 후의 모든 시기에 관련된 요인을 포함하고 있다.

3. 결론 및 시사점

소셜 빅데이터는 다양한 분야에 활용할 수 있다.

첫째, 본 연구의 적용과 같이 정부 정책과 새로운 기술에 대한 미래 신호를 사전에 예측하여 대비할 수 있다. 분야별 주요 정책에 대한 미래 신호를 탐색하고 예측함으로써 정책의 수요 예측과 평가가 가능할 수 있다. 또한 빅데이터, 사물인터넷, 머신러닝, 인공지능 등 신기술에 대한 미래 신호를 예측하여 적시 대응체계를 구축할 수 있다.

둘째, 조사를 통한 기존 정보수집체계의 한계를 보완할 수 있는 새로운 자료 수집 방법으로 활용할 수 있다. 통일인식 조사, 청소년 우울 조사, 담배인식 조사, 식품안전 조사, 인터넷 중독 조사 등 다양한 분야의 조사에 활용할 수 있을 것이다.

셋째, 사회적 위기 상황에 대한 모니터링과 예측으로 위험에 대한 사전 대응체계를 구축할 수 있다. 자살 위험 예측, 사이버폭력 위험 예측, 메르스 정보 확산 위험 예측, 지카바이러스 정보 확산 위험 예측, 북한 위험 예측 등 다양한 사회 위험에 대한 즉시 대응체계를 구축할 수 있다.

끝으로 빅데이터를 분석하여 인과성과 연관성을 발견하고 미래를 예측하기 위해서는 데이터사이언티스트 양성을 위한 정부 차원의 노력이 필요할 것으로 본다.

*주요 용어: 소셜 빅데이터, 보건복지, 데이터마이닝, 다변량분석, 시각화, R

제 1 장

소셜 빅데이터를 활용한 보건복지 정책 미래신호 예측

제1절 서론

제2절 소셜 빅데이터 기반 미래신호 예측 방법론

제3절 소셜 빅데이터 기반 보건복지 정책 미래신호 예측

참고문헌



1

소셜 빅데이터를 활용한 << 보건복지 정책 미래신호 예측>>

제1절 서론

미래변화의 트렌드를 파악하고 미래의 핵심 기술을 선별하기 위해 주요 선진국들은 주기적으로 국가의 미래 트렌드를 분석하고 그 결과를 발표하고 있다(정근하, 2010, p.6)²⁾. 그동안 미래 트렌드를 예측하기 위한 다양한 연구가 시도되어 왔으나 대부분의 전문가는 지식과 의견에 따라 미래를 전망하는 방법을 사용해 왔다. 최근 소셜네트워크서비스(SNS)를 비롯한 온라인 채널에서 생산되는 텍스트 형태의 비정형 데이터가 실제 경제 및 사회에 미치는 영향력이 매우 커짐에 따라 소셜 빅데이터를 활용한 미래 예측 연구가 진행되고 있으나 수집기술과 분석기술의 어려움으로 활발히 확산되지 못하고 있는 실정이다.

한편 현 정부는 국민이 행복한 사회를 이루기 위한 사회보장 정책 방향으로 '생애주기별 맞춤형 복지'를 제시하고 이를 실현하기 위한 맞춤형 복지정책을 도입 및 확대하고 있다. 2016년도에는 국민이 체감하는 맞춤형 복지 확산을 목표로 맞춤형 기초생활보호제도 정착, 4대 중증질환 등 의료보장 지속, 맞춤형 보육 개편, 기초연금 및 장기요양 지원 확대, 복지안내 강화, 취약계층 필수 서비스 지속, 국민 체감형 원격의료 확산, 신약 개발 및 제약산업 육성 등을 중점적으로 추진해 왔다. 정부에서 추진하고

-
- 1) 본 연구는 미래창조과학부/정보통신기술진흥센터의 R&D 프로그램(R7117-16-0219, 점진적 기계학습 기반 자가진화(Self-Evolving) 에이전트 시뮬레이션을 이용한 사회변화 예측분석 기술 개발의 지원을 받아 작성되었으며, '송태민(2016). 소셜 빅데이터를 활용한 미래신호예측: 보건복지 주요 정책과 이슈를 중심으로. 보건복지포럼, 통권 제238호, 한국보건사회연구원'에 게재된 것임을 밝힘.
 - 2) 정근하. (2010). 텍스트마이닝과 네트워크분석을 활용한 미래예측 방법 연구. 한국과학기술기획평가원.

있는 보건복지 정책을 성공적으로 추진하여 예상하는 성과를 얻기 위해서는 다양한 보건복지 욕구와 이해집단의 갈등을 최소화하기 위한 정책 동향 및 수요를 예측해 적시에 대응할 수 있는 체계가 구축되어야 한다. 본 연구는 우리나라에서 수집 가능한 모든 온라인 채널에서 언급된 보건복지 관련 문서를 수집하여 주제분석과 감성분석을 통하여 보건복지 주요 키워드를 분류하고 보건복지와 관련하여 나타나는 주요 정책과 이슈에 대한 미래신호를 탐지해 예측 모형을 제시하고자 한다.

제2절 소셜 빅데이터 기반 미래신호 예측 방법론

오늘날 미래의 환경 변화를 감지하기 위한 다양한 연구가 시도되고 있으며, 여러 연구 중에서 가장 많은 주목을 받고 있는 것은 미래의 변화를 예감할 수 있는 약신호(weak signal)를 탐지하는 것이다(Yoon, 2012; 박찬국, 김현제, 2015)³⁾. 약신호는 ‘미래에 가능한 변화의 징후’(Ansoff, 1975)⁴⁾로 약신호는 시간이 흐르면서 강신호(strong signal)로, 강신호는 다시 트렌드(trend)나 메가 트렌드(mega trend)로 발전할 수 있다. Hiltunen(2008)⁵⁾은 약신호를 미래신호(future sign)라는 개념을 이용하여 미래신호를 신호(signal), 이슈(issue), 이해(interpretation)와 같

3) Yoon, J. (2012). "Detecting weak signals for long-term business opportunities using text mining of Web news", *Journal Expert Systems with Applications* 39(16), pp.12543-12550; 박찬국, 김현제. (2015). 사물인터넷을 통한 에너지 산업 발전방향 연구-텍스트마이닝을 이용한 미래 신호 탐색. 에너지경제연구원.

4) Ansoff, H.I. (1975). Managing strategic surprise by response to weak signals, *Californian Management Review* 18(2), pp.21-33.

5) Hiltunen, E. (2008). "The future sign and its three dimensions", *Futures* 40, pp.247-260.

이 3차원의 미래신호 공간으로 설명하였다. Yoon(2012)은 웹 뉴스의 문서를 수집하여 텍스트마이닝 분석을 통해 생성된 단어빈도(Term Frequency, TF)와 문서빈도(Document Frequency, DF)를 Hiltunen(2008)의 신호와 이슈로 각각 연계하였다. Yoon(2012)은 단어빈도, 문서빈도, 발생빈도 증가율을 이용하여 KEM(Keyword Emergence Map)과 KIM(Keyword Issue Map)의 키워드 포트폴리오를 작성하고 작성된 키워드 포트폴리오를 이용하여 약신호를 선별하였다. KEM은 가시성을 보여 주는 것으로 DoV(degree of visibility)를 산출하고, KIM은 확산 정도를 보여 주는 것으로 DoD(degree of diffusion)를 산출할 수 있다.

$$DoV_{ij} = \left(\frac{TF_{ij}}{NN_j} \right) \times \{1 - tw \times (n - j)\}$$

$$DoD_{ij} = \left(\frac{DF_{ij}}{NN_j} \right) \times \{1 - tw \times (n - j)\}$$

여기서 NN 은 전체 문서 수를 의미하고, TF 는 단어빈도, DF 는 문서빈도, tw 는 시간가중치(본 연구에서 시간가중치는 0.05를 적용), n 은 전체시간구간, j 는 시점을 의미한다.

제3절 소셜 빅데이터 기반 보건복지 정책 미래신호 예측

1. 연구 대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜네트워크서비스(SNS), 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 대상으로 하였다. 본 분석에서는 149개의 온라인 뉴스 사이트, 4개의 블로그(네이

버·네이트·다음·티스토리), 2개의 카페(네이버·다음), 1개의 SNS(트위터), 15개의 게시판(네이버지식인·네이트지식·네이트톡·네이트판 등) 등 총 171개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서(버즈)를 소셜 빅데이터로 정의하였다. 보건복지 관련 토픽은 2016. 1. 1.~2016. 3. 31. 해당 채널에서 요일별, 주말, 휴일을 고려하지 않고 매시간 단위로 수집하였으며, 수집된 총 20만 1849건(1월: 8만 7567건, 2월: 6만 5278건, 3월: 4만 9004건)의 텍스트(Text) 문서를 본 연구의 분석에 포함시켰다. 본 연구를 위한 소셜 빅데이터의 수집⁶⁾은 크롤러(Crawler)를 사용하였고, 토픽의 분류는 주제분석 기법을 사용하였다. 보건복지 토픽은 모든 관련 문서를 수집하기 위해 ‘보건’, ‘복지’, 그리고 ‘보건복지’를 사용하였다.

2. 단어 및 문서 빈도 분석⁷⁾

온라인 채널에서 수집된 텍스트 형태의 문서를 분석하기 위해서는 텍스트마이닝을 통하여 우선적으로 문서 내에서 출현하는 단어별 빈도를 산출해야 한다. 텍스트마이닝 분석을 위해서는 단어빈도와 문서빈도를 산출해야 한다. 그리고 중요한 정보의 추출을 위해서 TF-IDF(Term Frequency-Inverse Document Frequency) 방법을 사용하고 있다.

Spärck(1972)⁸⁾는 희귀한 단어일수록 더 높은 가중치를 부여하기 위

6) 본 연구를 위한 소셜 빅데이터의 수집 및 토픽 분류는 '(주)SK텔레콤 스마트인사이트'에서 수행함.

7) 본 단어빈도 및 문서빈도의 분석 절차는 박찬국, 김현제(2015)의 분석 방법을 참고한 것임을 밝힌다.

8) Spärck Jones, K. (1972). "A Statistical Interpretation of Term Specificity and Its Application in Retrieval". *Journal of Documentation* 28: 11-21. doi:10.1108/eb026526.

해 역문서빈도[Inverse Document Frequency, $IDF_j = \log_{10}(\frac{N}{DF_j})$]를 제안하였다. 따라서 단어빈도 분석에 희귀한 단어일수록 더 높은 가중치를 부여할 필요가 있다면 단어빈도와 역문서빈도를 결합하여 'TF-IDF= $TF_{ij} \times IDF_j$ '를 산출하여 가중치(단어의 중요도 지수)를 적용한다.

상기 분석방법론에 따라 단어빈도, 문서빈도, 단어의 중요도 지수를 고려한 문서의 빈도 분석을 통하여 보건복지 관련 정책과 주요 이슈에 대한 인식 변화를 살펴보았다(〈표 1-1〉 참조). 단어빈도에서는 일자리, 증세, 세금, 복지급여, 결혼출산, 건강증진, 치료 등의 순위로 나타나고 있어 정책은 일자리, 복지급여, 결혼출산이 우선이고 주요 이슈는 증세, 세금, 치료가 우선인 것으로 나타났다. 문서빈도는 단어빈도와 비슷한 추이를 나타내고 있으나 결혼출산이 단어빈도에서는 5위인 반면 문서빈도에서는 7위로 나타났다. 키워드의 중요성을 나타내는 단어빈도에서는 결혼출산이 중요하나 주제의 확산을 나타내는 문서빈도에서는 다소 떨어져 결혼출산 정책의 확산에 대한 노력이 필요할 것으로 본다. 중요도 지수를 고려한 단어빈도에서 정책은 일자리, 복지급여, 결혼출산이 우선이고 주요 이슈는 증세, 세금이 우선인 것으로 나타났다. 그리고 키워드의 월별 순위 변화는 2016년 2월까지 증세, 일자리, 세금, 복지급여, 치료가 중요한 키워드로 나타나다 3월에는 건강증진이 강조되기 시작하여 건강에 대한 관심이 확산되고 있는 것으로 나타났다.

〈표 1-1〉 온라인 채널의 보건복지 정책, 이슈의 키워드 분석

| 순위 | TF | | DF | | TF-IDF | |
|----|------|-------|------|-------|--------|-------|
| | 키워드 | 빈도 | 키워드 | 빈도 | 키워드 | 빈도 |
| 1 | 일자리 | 8,212 | 증세 | 8,059 | 일자리 | 6,328 |
| 2 | 증세 | 8,059 | 일자리 | 7,459 | 증세 | 5,940 |
| 3 | 세금 | 5,339 | 세금 | 5,339 | 복지급여 | 4,955 |
| 4 | 복지급여 | 4,520 | 복지급여 | 3,524 | 세금 | 4,890 |

14 2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

| 순위 | TF | | DF | | TF-IDF | |
|----|--------|--------|--------|--------|--------|--------|
| | 키워드 | 빈도 | 키워드 | 빈도 | 키워드 | 빈도 |
| 5 | 결혼출산 | 3,419 | 건강증진 | 3,352 | 결혼출산 | 4,267 |
| 6 | 건강증진 | 3,352 | 치료 | 2,637 | 건강증진 | 3,748 |
| 7 | 치료 | 2,938 | 결혼출산 | 2,485 | 치료 | 3,591 |
| 8 | 건강보험 | 1,307 | 무상정책 | 1,116 | 건강보험 | 2,114 |
| 9 | 무상정책 | 1,156 | 건강보험 | 1,062 | 무상정책 | 1,845 |
| 10 | 기초연금 | 922 | 기초연금 | 922 | 기초연금 | 1,548 |
| 11 | 개인정보 | 872 | 개인정보 | 872 | 개인정보 | 1,485 |
| 12 | 미래세대육성 | 817 | 미래세대육성 | 817 | 미래세대육성 | 1,414 |
| 13 | 의료비 | 754 | 의료비 | 754 | 의료비 | 1,332 |
| 14 | 국민연금 | 738 | 국민연금 | 738 | 국민연금 | 1,310 |
| 15 | 의료민영화 | 686 | 의료민영화 | 662 | 보건산업 | 1,263 |
| 16 | 보건산업 | 672 | 자살 | 634 | 의료민영화 | 1,250 |
| 17 | 자살 | 634 | 보건산업 | 581 | 자살 | 1,167 |
| 18 | 부동산 | 483 | 부동산 | 483 | 부동산 | 946 |
| 19 | 등록금 | 471 | 등록금 | 459 | 보육 | 939 |
| 20 | 보육 | 460 | 보육 | 400 | 등록금 | 933 |
| 21 | 담배 | 383 | 담배 | 383 | 담배 | 789 |
| 22 | 가족친화 | 348 | 양극화 | 306 | 가족친화 | 761 |
| 23 | 양극화 | 306 | 가족친화 | 287 | 양극화 | 660 |
| 24 | 원격의료 | 237 | 중증질환 | 223 | 원격의료 | 546 |
| 25 | 중증질환 | 233 | 원격의료 | 219 | 중증질환 | 535 |
| 26 | 환자안전 | 229 | 환자안전 | 212 | 환자안전 | 531 |
| 합계 | | 47,547 | 합계 | 43,985 | 합계 | 55,084 |

3. 보건복지 관련 키워드의 미래신호 탐색⁹⁾

미래신호 탐지방법론에 따라 분석한 결과는 <표 1-2>, <표 1-3>과 같다. 보건복지 관련(정책, 이슈) 키워드에 대한 DoV 증가율과 평균단어빈도를 산출한 결과 일자리와 복지급여는 높은 빈도를 보이고 있으나 DoV 증가율은 중앙값보다 낮게 나타나 시간이 지날수록 신호가 약해지는 것으로 나타났다. 결혼출산, 건강증진은 평균단어빈도가 높게 나타났으며, DoV 증가율은 중앙값보다 높게 나타나 시간이 지날수록 빠르게 신호가

9) 본 미래신호 탐색 절차는 박찬국, 김현제(2015)의 분석 방법을 참고한 것임을 밝힌다.

강해지는 것으로 나타났다.

미래신호 탐색을 위해 DoV의 평균단어빈도와 DoD의 평균문서빈도를 X축으로 설정하고 DoV와 DoD의 평균증가율을 Y축으로 설정한 후, 각 값의 중앙값을 사분면으로 나누면 2사분면에 해당하는 영역의 키워드는 약신호가 되고 1사분면에 해당하는 키워드는 강신호가 된다. 빈도수 측면에서는 상위 10위에 DoV는 일자리, 증세, 세금, 복지급여, 결혼출산, 건강증진, 치료, 건강보험, 무상정책, 기초연금 순으로 포함되었고 DoD에는 증세, 일자리, 세금, 복지급여, 건강증진, 치료, 결혼출산, 무상정책, 건강보험, 기초연금 순으로 포함되었다. DoV 증가율의 중앙값(0.23)보다 높은 증가율을 보이는 키워드는 결혼출산, 건강증진, 건강보험으로 나타났으며 DoD 증가율의 중앙값(0.23)보다 높은 증가율을 보이는 키워드는 건강증진, 건강보험으로 나타났다. 특히 결혼출산의 DoV 증가율은 중앙값보다 높은 반면 DoD 증가율은 중앙값보다 낮게 나타나 결혼출산 정책의 확산을 위한 방안이 필요할 것이다.

〈표 1-2〉 보건복지 정책, 이슈의 DoV 평균증가율과 평균단어빈도

| 키워드 | DoV | | | 평균증가율 | 평균 단어빈도 |
|--------|-------|-------|-------|--------|---------|
| | 1월 | 2월 | 3월 | | |
| 일자리 | 0.142 | 0.168 | 0.186 | 0.147 | 2737 |
| 증세 | 0.158 | 0.237 | 0.063 | -0.116 | 2686 |
| 세금 | 0.136 | 0.098 | 0.077 | -0.246 | 1780 |
| 복지급여 | 0.11 | 0.067 | 0.094 | 0.008 | 1507 |
| 결혼출산 | 0.064 | 0.049 | 0.097 | 0.369 | 1140 |
| 건강증진 | 0.062 | 0.051 | 0.094 | 0.339 | 1117 |
| 치료 | 0.045 | 0.07 | 0.061 | 0.208 | 979 |
| 건강보험 | 0.026 | 0.02 | 0.034 | 0.245 | 436 |
| 무상정책 | 0.024 | 0.021 | 0.024 | 0.014 | 385 |
| 기초연금 | 0.013 | 0.02 | 0.023 | 0.33 | 307 |
| 개인정보 | 0.005 | 0.04 | 0.004 | 3.394 | 291 |
| 미래세대육성 | 0.01 | 0.005 | 0.04 | 3.498 | 272 |
| 의료비 | 0.015 | 0.01 | 0.022 | 0.415 | 251 |
| 국민연금 | 0.007 | 0.014 | 0.026 | 0.886 | 246 |

16 2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

| 키워드 | DoV | | | 평균증가율 | 평균 단어빈도 |
|-------|-------|-------|-------|--------|---------|
| | 1월 | 2월 | 3월 | | |
| 의료민영화 | 0.003 | 0.004 | 0.04 | 5.242 | 229 |
| 보건산업 | 0.011 | 0.013 | 0.017 | 0.216 | 224 |
| 자살 | 0.006 | 0.005 | 0.03 | 2.167 | 211 |
| 부동산 | 0.009 | 0.012 | 0.007 | -0.034 | 161 |
| 등록금 | 0.011 | 0.007 | 0.011 | 0.1 | 157 |
| 보육 | 0.008 | 0.012 | 0.006 | 0.015 | 153 |
| 담배 | 0.005 | 0.005 | 0.014 | 0.827 | 128 |
| 가족친화 | 0.004 | 0.004 | 0.014 | 1.275 | 116 |
| 양극화 | 0.01 | 0.003 | 0.004 | -0.226 | 102 |
| 원격의료 | 0.005 | 0.005 | 0.004 | -0.103 | 79 |
| 중증질환 | 0.007 | 0.004 | 0.003 | -0.307 | 78 |
| 환자안전 | 0.002 | 0.008 | 0.003 | 1.062 | 76 |
| 중앙값 | | | | 0.23 | 249 |

〈표 1-3〉 보건복지 정책, 이슈의 DoD 평균증가율과 평균단어빈도

| 키워드 | DoD | | | 평균증가율 | 평균 단어빈도 |
|--------|-------|-------|-------|--------|---------|
| | 1월 | 2월 | 3월 | | |
| 증세 | 0.172 | 0.251 | 0.07 | -0.131 | 2686 |
| 일자리 | 0.138 | 0.163 | 0.186 | 0.159 | 2486 |
| 세금 | 0.148 | 0.104 | 0.085 | -0.238 | 1780 |
| 복지급여 | 0.092 | 0.055 | 0.082 | 0.047 | 1175 |
| 건강증진 | 0.067 | 0.054 | 0.104 | 0.369 | 1117 |
| 치료 | 0.043 | 0.068 | 0.06 | 0.221 | 879 |
| 결혼출산 | 0.054 | 0.042 | 0.069 | 0.224 | 828 |
| 무상정책 | 0.026 | 0.021 | 0.026 | 0.015 | 372 |
| 건강보험 | 0.022 | 0.017 | 0.032 | 0.309 | 354 |
| 기초연금 | 0.014 | 0.021 | 0.026 | 0.337 | 307 |
| 개인정보 | 0.005 | 0.042 | 0.005 | 3.278 | 291 |
| 미래세대육성 | 0.011 | 0.005 | 0.044 | 3.683 | 272 |
| 의료비 | 0.016 | 0.011 | 0.024 | 0.453 | 251 |
| 국민연금 | 0.008 | 0.015 | 0.028 | 0.902 | 246 |
| 의료민영화 | 0.003 | 0.003 | 0.044 | 6.347 | 221 |
| 자살 | 0.007 | 0.006 | 0.033 | 2.279 | 211 |
| 보건산업 | 0.01 | 0.013 | 0.015 | 0.229 | 194 |
| 부동산 | 0.01 | 0.013 | 0.008 | -0.04 | 161 |
| 등록금 | 0.011 | 0.007 | 0.012 | 0.12 | 153 |
| 보육 | 0.008 | 0.012 | 0.005 | 0.015 | 133 |
| 담배 | 0.006 | 0.005 | 0.015 | 0.875 | 128 |
| 양극화 | 0.011 | 0.003 | 0.004 | -0.203 | 102 |
| 가족친화 | 0.004 | 0.004 | 0.012 | 1.047 | 96 |

| 키워드 | DoD | | | 평균증가율 | 평균 단어빈도 |
|------|-------|-------|-------|--------|---------|
| | 1월 | 2월 | 3월 | | |
| 중증질환 | 0.007 | 0.003 | 0.003 | -0.265 | 74 |
| 원격의료 | 0.005 | 0.005 | 0.004 | -0.119 | 73 |
| 환자안전 | 0.002 | 0.008 | 0.003 | 1.27 | 71 |
| 중앙값 | | | | 0.23 | 249 |

[그림 1-1], [그림 1-2]와 같이 보건복지 관련 주요 키워드는 복지급여와 일자리는 KEM에서는 강신호로 나타난 반면 KIM에서는 강하지는 않지만 잘 알려진 신호로 나타났다. KEM과 KIM에 공통적으로 나타나는 강신호(1사분면)에는 미래세대육성, 개인정보, 국민연금, 의료비, 기초연금, 건강보험, 결혼출산, 치료, 건강증진이 포함되었고 약신호(2사분면)에는 의료민영화, 자살, 환자안전, 가족친화, 담배, 보건산업이 포함된 것으로 나타났다. KIM의 4사분면에만 나타난, 강하지는 않지만 잘 알려진 신호는 무상정책, 복지급여, 세금, 일자리, 증세이며 KIM의 3사분면에만 나타난 잠재신호는 등록금, 보육, 부동산, 원격의료, 양극화, 중증질환으로 나타났다.

〈표 1-4〉 보건복지 관련 정책, 이슈의 미래신호

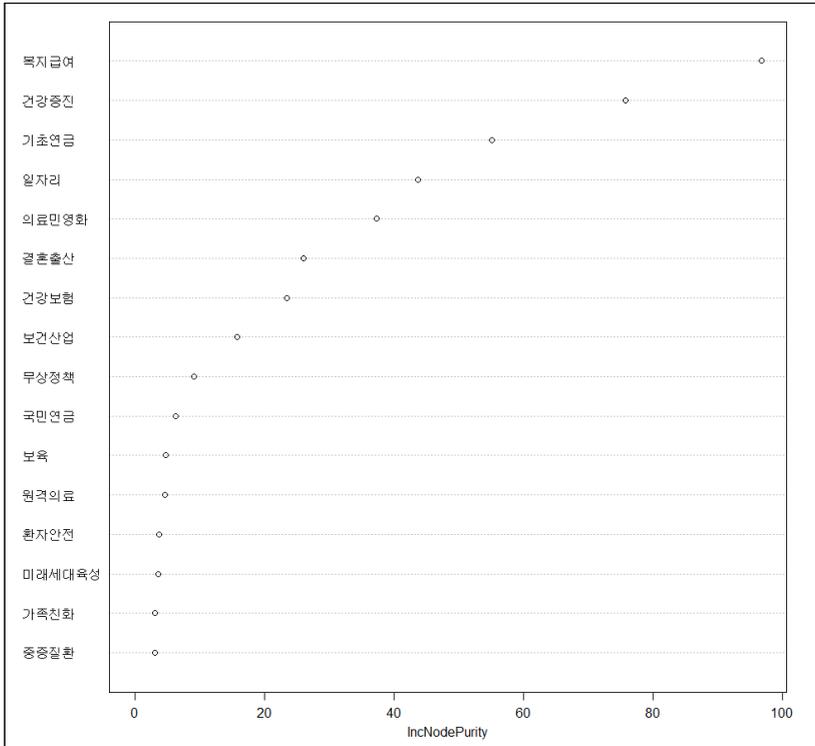
| 구분 | 잠재신호 (Latent Signal) | 약신호 (Weak Signal) | 강신호 (Strong Signal) | 강하지 않지만 잘 알려진 신호 (Not Strong but well known Signal) |
|-------|-------------------------------|---------------------------------------|---|--|
| KEM | 등록금, 보육, 부동산, 원격의료, 양극화, 증증질환 | 의료민영화, 자살, 가족친화, 환자안전, 담배, 보건산업, 국민연금 | 미래세대육성, 개인정보, 의료비, 기초연금, 건강보험, 결혼출산, 치료, 건강증진 | 무상정책, 복지급여, 세금, 일자리, 증세 |
| KIM | 등록금, 보육, 부동산, 원격의료, 양극화, 증증질환 | 의료민영화, 자살, 환자안전, 가족친화, 담배, 보건산업, 국민연금 | 미래세대육성, 개인정보, 의료비, 기초연금, 건강보험 치료, 건강증진 | 무상정책, 결혼출산, 복지급여, 세금, 일자리, 증세 |
| 주요 신호 | 등록금, 보육, 부동산, 원격의료, 양극화, 증증질환 | 의료민영화, 자살, 환자안전, 가족친화, 담배, 보건산업, 국민연금 | 미래세대육성, 개인정보, 의료비, 기초연금, 건강보험, 치료, 건강증진 | 무상정책, 복지급여, 세금, 일자리, 증세 |

4. 보건복지 정책 관련 미래신호 예측

가. 랜덤 포리스트 분석을 통한 주요 보건복지 정책 요인 예측

본 연구의 랜덤 포리스트(Random Forest) 분석을 활용하여 보건복지 수요(태도)에 영향을 주는 주요 정책 요인을 살펴보면 [그림 1-3]과 같다. 보건복지 수요(찬성, 반대)에 가장 큰 영향을 미치는(연관성이 높은) 정책 요인은 ‘복지급여’ 정책으로 나타났으며 그 뒤를 이어 건강증진, 기초연금, 일자리, 의료민영화, 결혼출산, 건강보험, 보건산업 정책 등의 순으로 나타났다.

[그림 1-3] 랜덤 포리스트 모델의 보건복지 주요 정책의 중요도



랜덤 포리스트의 중요도로 나타난 정책 요인들이 보건복지 수요에 미치는 영향을 로지스틱회귀분석을 통하여 살펴본 결과 기초노령연금, 의료민영화는 반대의 확률이 높으며, 그 외 국민연금($p < .1$), 보육, 출산양육, 가족친화, 건강보험, 원격의료 등은 찬성의 확률이 높은 것으로 나타났다(〈표 1-5〉 참조).

〈표 1-5〉 보건복지 수요에 영향을 주는 주요 정책 요인^{주)}

| 정책 | 찬성 | | | |
|--------|----------------|-------------------|-----------------|------|
| | b [†] | S.E. [‡] | OR [§] | P |
| 국민연금 | .254 | .153 | 1.289 | .098 |
| 기초연금 | -1.483 | .100 | .227 | .000 |
| 보육 | 1.440 | .318 | 4.221 | .000 |
| 결혼출산 | .629 | .070 | 1.875 | .000 |
| 가족친화 | 1.119 | .300 | 3.061 | .000 |
| 미래세대육성 | .796 | .212 | 2.216 | .000 |
| 무상정책 | .405 | .120 | 1.499 | .001 |
| 의료민영화 | -1.634 | .121 | .195 | .000 |
| 건강보험 | 1.180 | .131 | 3.256 | .000 |
| 원격의료 | 1.133 | .316 | 3.106 | .000 |
| 중증질환 | 1.037 | .342 | 2.821 | .002 |
| 환자안전 | 1.303 | .351 | 3.681 | .000 |
| 보건산업 | 1.625 | .232 | 5.079 | .000 |
| 복지급여 | 1.225 | .068 | 3.405 | .000 |
| 건강증진 | 1.041 | .065 | 2.831 | .000 |
| 일자리 | .501 | .045 | 1.650 | .000 |

주: 기본범주=반대, †Standardized coefficients, ‡Standard error, § odds ratio.

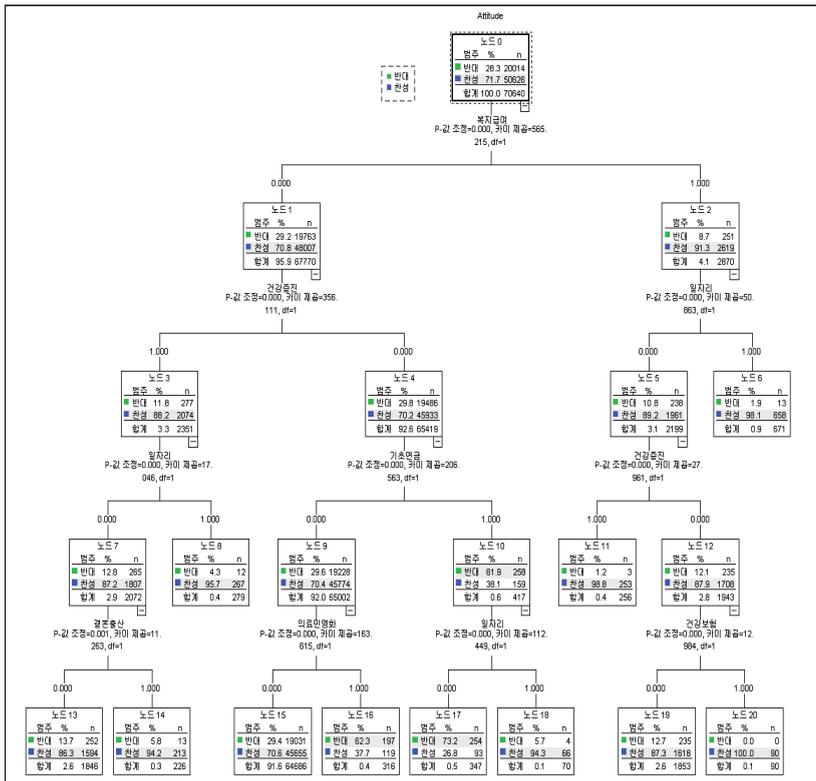
나. 의사결정나무 분석을 통한 보건복지 정책 수요 예측

보건복지 정책 수요 예측 모형에 대한 의사결정나무는 [그림 1-4]와 같다. 나무 구조의 최상위에 있는 뿌리나무는 예측변수(독립변수)가 투입되지 않은 종속변수의 빈도를 나타낸다. 뿌리마디의 보건복지에 대한 감정의 비율을 보면 보건복지에 대해 찬성은 71.7%, 반대는 28.3%로 나타

22 2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

났다. 뿌리마디 하단의 가장 상위에 위치하는 정책 요인이 종속변수에 대한 영향력이 가장 높은 요인(관련성이 깊은)으로, ‘복지급여’ 정책의 영향력이 가장 큰 것으로 나타났다. 즉, 온라인 문서에 ‘복지급여’ 정책이 있는 경우 찬성은 이전의 71.1%에서 91.3%로 증가하였다. ‘복지급여’ 정책이 있고 ‘일자리’ 정책이 있는 경우 찬성은 이전의 91.3%에서 98.1%로 증가하는 것으로 나타났다.

[그림 1-4] 보건복지 정책 수요 예측의 의사결정나무 모형



다. 연관분석을 통한 정책 요인 예측

소셜 빅데이터 분석에서 연관분석은 하나의 온라인 문서에 포함된 둘 이상의 단어들에 대한 상호 관련성을 발견하는 것이다. 본 연구에서는 <표 1-6>과 같이 하나의 문서에 나타난 정책 요인의 수요에 대한 연관규칙을 분석하였다. {의료민영화}⇒{반대} 두 변인의 연관성은 지지도가 0.003, 신뢰도는 0.579, 향상도는 2.044로 나타났다. 이는 온라인 문서에서 ‘의료민영화’ 정책이 언급되면 반대할 확률이 57.9%이며, ‘의료민영화’ 정책이 언급되지 않은 문서보다 반대할 확률이 약 2.04배 높아지는 것을 나타낸다. {건강보험, 중증질환}⇒{찬성}으로 세 변인의 향상도는 1.40으로 온라인 문서에서 건강보험과 중증질환 정책이 언급되지 않은 문서보다 찬성할 확률이 1.40배 높은 것으로 나타났다.

<표 1-6> 보건복지 주요 정책의 연관규칙

| 규칙 | 지지도 | 신뢰도 | 향상도 |
|-------------------------|-------------|------------|-----------|
| {의료민영화}⇒{반대} | 0.002845413 | 0.57925072 | 2.0444824 |
| {기초연금}⇒{반대} | 0.003680634 | 0.44444444 | 1.5686797 |
| {건강보험, 중증질환}⇒{찬성} | 0.001670442 | 1.00000000 | 1.3953305 |
| {출산양육, 건강증진}⇒{찬성} | 0.001528879 | 1.00000000 | 1.3953305 |
| {출산양육, 복지급여}⇒{찬성} | 0.002180068 | 1.00000000 | 1.3953305 |
| {보건산업, 복지급여}⇒{찬성} | 0.001005096 | 1.00000000 | 1.3953305 |
| {국민연금, 복지급여}⇒{찬성} | 0.002109287 | 1.00000000 | 1.3953305 |
| {건강보험, 복지급여}⇒{찬성} | 0.002831257 | 1.00000000 | 1.3953305 |
| {출산양육, 복지급여, 건강증진}⇒{찬성} | 0.001090034 | 1.00000000 | 1.3953305 |
| {출산양육, 건강증진, 일자리}⇒{찬성} | 0.001005096 | 1.00000000 | 1.3953305 |
| {출산양육, 복지급여, 일자리}⇒{찬성} | 0.001373160 | 1.00000000 | 1.3953305 |
| {국민연금, 건강보험, 복지급여}⇒{찬성} | 0.001061721 | 1.00000000 | 1.3953305 |
| {국민연금, 복지급여, 일자리}⇒{찬성} | 0.001160815 | 1.00000000 | 1.3953305 |
| {건강보험, 복지급여, 건강증진}⇒{찬성} | 0.001061721 | 1.00000000 | 1.3953305 |
| {건강보험, 복지급여, 일자리}⇒{찬성} | 0.001443941 | 1.00000000 | 1.3953305 |

| 규칙 | 지지도 | 신뢰도 | 향상도 |
|------------------------|-------------|------------|-----------|
| {중증질환, 복지급여}⇒{찬성} | 0.001613817 | 0.99130435 | 1.3831972 |
| {기초연금, 복지급여, 일자리}⇒{찬성} | 0.001359003 | 0.98969072 | 1.3809456 |
| {복지급여, 건강증진, 일자리}⇒{찬성} | 0.002548131 | 0.98901099 | 1.3799972 |
| {복지급여, 건강증진}⇒{찬성} | 0.006129672 | 0.98858447 | 1.3794020 |
| {국민연금, 건강보험, 일자리}⇒{찬성} | 0.001160815 | 0.98795181 | 1.3785193 |

5. 결론

보건복지 정책과 이슈의 미래신호를 예측하기 위해서는 연도별 시계열 정보가 필요하나 본 연구는 2016. 1.~2016. 3. 3개월간의 정보를 수집하여 우선적으로 미래신호 예측 방법론을 적용해 보았다. 본 연구의 정보 분석 기간의 한계에도 불구하고 보건복지 정책과 이슈에 대한 미래신호 예측 결과를 살펴보면 다음과 같다. 첫째, 본 연구의 보건복지 정책과 이슈의 미래신호 분석에서 미래세대육성, 국민연금, 기초연금, 건강보험, 결혼출산, 건강증진, 개인정보, 의료비, 치료가 강신호로 분류되어 미래세대육성 및 건강증진과 관련된 의료비와 치료 등이 강조되고 있는 것으로 나타났다. 특히 미래세대육성과 개인정보는 강신호이면서 높은 증가율을 보이고 있어 미래세대육성 정책에 포함된 아동학대 문제의 해결과 개인정보 보호와 관련된 제도 개선에 대한 논의가 지속적으로 이루어져야 할 것으로 본다. 둘째, 의료민영화, 자살, 환자안전, 가족친화, 담배, 보건산업은 약신호로 분류되었다. 특히 약신호인 의료민영화와 자살은 높은 증가율을 보이고 있어 이들 키워드는 시간이 지날수록 강신호로 발전할 수 있기 때문에 이에 대한 대응책이 마련되어야 할 것으로 본다. 보건복지 정책의 미래신호 예측에서 중요한 정책이면서 찬성하는 정책은 복지급여, 건강증진, 일자리, 결혼출산, 건강보험, 보건산업 등의 순으로 나타났다. 특히 복지급여와 일자리가 동시에 언급된 문서의 찬성이 매우

높은 것으로 나타나 능동적 복지체계 구축을 통한 일자리 창출이 필요할 것으로 본다. 이는 정책의 연관분석 결과와 같이 {기초연금} 정책만 언급된 문서에는 반대하는 것으로 나타났으나 {기초연금, 복지급여, 일자리} 정책이 동시에 언급된 문서에는 찬성하는 것으로 나타나 노인의 능동적 자활과 근로를 통한 복지체계 구축에 대한 국민의 요구가 높은 것으로 본다. 또한 {건강보험, 중증질환}이 동시에 언급된 문서의 찬성이 높은 것으로 나타나 건강보험 혜택 확대로 4대 중증질환의 보장성 강화가 국민의 의료비 부담을 줄임으로써 정부의 정책에 대한 좋은 평가 결과로 나타난 것으로 본다.

소셜 빅데이터는 다양한 분야에 활용할 수 있다. 첫째, 본 연구의 적용과 같이 정부의 정책과 새로운 기술에 대한 미래신호를 사전에 예측하여 대비할 수 있다. 둘째, 조사를 통한 기존 정보수집체계의 한계를 보완할 수 있는 새로운 자료 수집 방법으로 활용할 수 있다. 셋째, 사회적 위기 상황에 대한 모니터링과 예측으로 위험에 대비한 사전 대응체계를 구축할 수 있다. 끝으로 빅데이터를 분석하여 인과성과 연관성을 발견하고 미래를 예측하기 위해서는 데이터사이언티스트 양성을 위한 정부 차원의 노력이 필요하다.



참고문헌 <<

- 정근하. (2010). 텍스트마이닝과 네트워크분석을 활용한 미래예측 방법 연구. 한국과학기술기획평가원.
- 박찬국, 김현제. (2015). 사물인터넷을 통한 에너지 신산업 발전방향 연구-텍스트 마이닝을 이용한 미래 신호 탐색. 에너지경제연구원.
- Ansoff, H.I. (1975). *Managing strategic surprise by response to weak signals*, Californian Management Review 18(2), pp.21-33.
- Hiltunen, E. (2008). *The future sign and its three dimensions*, Futures 40, pp.247-260.
- Spärck Jones, K. (1972). *A Statistical Interpretation of Term Specificity and Its Application in Retrieval*. Journal of Documentation 28: 11-21. doi:10.1108/eb026526.
- Yoon, J. (2012). *Detecting weak signals for long-term business opportunities using text mining of Web news*, Journal Expert Systems with Applications 39(16), pp.12543-12550.



제 2 장

2016 소셜 빅데이터를 활용한
통일인식 동향 분석 및 예측

제1절 서론

제2절 연구 방법

제3절 연구 결과

제4절 결론

참고문헌



2

2016년 소셜 빅데이터를 활용한 << 통일인식 동향 분석 및 예측

제1절 서론

최근 ‘한반도 정세’가 북한의 로켓 발사와 유엔(UN)의 북한에 대한 강력한 제재 결정, 각종 군사훈련(최윤철, 2016, p.51), 개성공단 잠정 폐쇄 등으로 매우 긴박하게 전개되고 있다. 현 정부는 ‘통일’을 핵심적 국정 과제 중 하나로 설정한 데 이어 통일에 대한 준비와 세부 정책을 추진해 오고 있다. ‘통일대박론’으로 통일된 한반도에 대한 비전을 제시했으며, 북한의 불안정한 정치 상황으로 인한 실질적 흡수통일 준비라는 해석부터 경제적 효과에 중점을 둔 전략적 판단이라는 설까지 다양하게 제기되고 있다(최영준, 2016, p.6). 통일에 대한 한국인의 전반적인 관심도는 2010년 52.6%에서 2014년 82.6%로 증가하였으나 20대의 통일에 대한 무관심은 다른 세대에 비해 높은 상태로 지속되고 있는 것으로 나타났다(아산정책연구원, 2015, pp.25-26). 김병조(2015)의 연구에서도 최근의 남북 관계가 모든 세대의 통일인식에 강하게 영향을 미치는 것으로 나타났다지만, 신자유주의통일세대(19~30세)는 통일 필요성에 대한 인식이 낮은 것으로 나타났다(김병조, 2015, p.15, p.26). 남북통일을 선도하고 통일 이후 시대를 살아가야 할 젊은 층의 통일에 대한 인식을 제고하고 관련 현안에 대한 관심을 환기시키는 것은 향후 정부의 통일 관련 커뮤니케이션의 핵심적인 과업이라고 할 수 있다(최재원, 김성철, 2016). 최근 미디어 환경은 다양한 플랫폼, 이를 통해 제공되는 다양한 콘텐츠들로 확장됨에 따라(홍중윤, 2016) 통일에 대한 공감대 형성에서 미디어는 중요

한 역할을 수행한다. 통일 관련 정보나 뉴스를 접할 창구가 제한된 현실에서 미디어는 통일에 대한 가장 폭넓은 정보 및 뉴스를 쉽게 얻을 수 있는 통로로 작용함과 동시에 담론이 형성·공유되는 곳으로 이에 국민들이 통일인식의 질을 결정한다고 해도 과언이 아니다(연세대학교 산학협력단, 2014). 이에 세계 각국의 정부와 기업들은 SNS를 통해 생산되는 소셜 빅데이터의 활용과 분석을 통하여 사회적 문제의 해결은 물론 새로운 경제적 효과와 일자리 창출을 위해 적극적으로 노력하고 있으며, 우리나라는 정부3.0과 창조경제의 추진과 실현을 위하여 다양한 분야에 대한 빅데이터의 효율적 활용을 적극적으로 모색하고 있다. 그뿐만 아니라 인터넷 또는 온라인 공간에서의 정치 토론과 정치 담론의 특성에 관한 연구는 많은 연구자들에 의해 이뤄졌으나, 최근 SNS나 소셜미디어와 같은 새로운 웹 기술이 온라인 정치 토론과 정치 담론에 미치는 영향에 관한 연구는 아직까지 미흡한 상태이다. 소셜미디어의 급격한 성장과 확산에도 불구하고 정치 담론의 본질과 특성에 대한 관심이 여전히 높지 않은 실정이기 때문에 소셜미디어에 기반을 둔 온라인 정치 담론의 특성에 대한 분석으로 관련 연구 영역을 확장시켜야 할 필요성이 제기된다(이원태, 2010).

우리 국민의 통일에 대한 다양한 인식을 살펴보기 위하여 그동안 실시 하던 설문조사는 정해진 변인에 대한 개인과 집단의 관계를 보는 데는 유용하나, 사이버상에 언급된 개인별 담론(buzz)과 사회적 현상들이 얼마나 어떻게 연관되어 있는지 밝히고 원인을 파악하는 데는 한계가 있다(송주영, 송태민, 2014). 이에 반해 소셜 빅데이터의 분석은 훨씬 방대한 양의 데이터를 활용하여 다양한 참여자의 생각과 의견을 확인할 수 있기 때문에 기존의 오프라인 조사와 함께 활용하면 통일인식의 예측을 더욱 정확하게 할 수 있다. 본 장에서는 우리나라 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 바탕으로 우리나라

국민의 통일인식에 대한 동향을 분석하고 통일인식의 예측 모형과 연관 규칙을 파악한다. 소셜 빅데이터의 분석은 사용자가 남긴 문서의 의미를 분석하는 것으로 자연어 처리 기술인 주제분석(text mining)과 감성분석 기술인 오피니언 마이닝(opinion mining)을 실시한 후 네트워크분석(network analysis)과 통계분석(statistical analysis)을 해야 한다.

제2절 연구 방법

1. 연구 대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 대상으로 하였다. 본 분석에서는 160개의 온라인 뉴스 사이트, 4개의 블로그(네이버·네이트·다음·티스토리), 2개의 카페(네이버·다음), 1개의 SNS(트위터), 9개의 게시판(네이버 지식인·네이트지식·네이트톡·다음지식인·다음아고라·SLR클럽 등) 등 총 176개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 온라인 문서(버즈)를 소셜 빅데이터로 정의하였다. 통일 관련 토픽(topic)¹⁰⁾은 2016년 1월부터 3월까지 총 3개월 동안¹¹⁾ 해당 채널에서 요일, 주말, 휴일을 고려하지 않고 매시간 단위로 수집하였으며, 수집된 총 9만 5507건의 텍스트(Text) 문서¹²⁾를 본 연구의 분석에 포함시켰다. 통일 토픽은 모든 관

10) 토픽은 소셜 분석 및 모니터링의 '대상이 되는 주제어'를 의미하며, 문서 내에 관련 토픽이 포함된 문서를 수집함.

11) 본 연구는 2014년 1월 2일 통일 대박 발언 이후 통일 담론이 확산됨에 따라 연도별 통일인식 비교를 위해 1/4분기를 분석 대상으로 결정함.

12) 수집된 문서는 SNS(73.8%, 7만 774건), 뉴스(6.0%, 2만 4693건), 블로그(6.9%, 6641건), 카페(9.6%, 9205건), 게시판(4.4%, 5111건)의 순으로 나타남.

련 문서를 수집하기 위해 ‘통일’을 사용하였으며, 토픽과 같은 의미로 사용되는 토픽 유사어로는 ‘남북통일, 한반도통일’ 용어를 사용하였다. 그리고 불용어로는 ‘통일신라, 통일교 등’을 사용하였다. 본 연구를 위한 소셜 빅데이터의 수집¹³⁾은 크롤러(Crawler)를 사용하였고, 이후 주제분석을 통해 분류된 명사형 어휘를 유목화(categorization)하여 분석요인으로 설정하였다.

2. 연구 도구

통일과 관련하여 수집된 문서는 주제분석¹⁴⁾의 과정을 거쳐 다음과 같이 정형화 데이터로 코드화하여 사용하였다.

가. 통일 관련 감정

본 연구의 통일 감정 키워드는 문서 수집 이후 주제분석을 통하여 총 56개(대박, 문제, 평화, 필요, 통일대박, 애국심, 중요, 노력, 반대, 족박, 관심, 희망, 신뢰, 우려, 불가능, 잘못, 사랑, 성공, 부정, 최고, 재앙, 위협, 행복, 공감, 빨갱이, 평화적, 천박, 위험, 염원, 포기, 비난, 고통, 강력, 걱정, 갈등, 분열, 혼란, 위대, 환영, 경박, 압박, 비아냥, 조롱, 든든, 불신, 환장, 친절, 한심, 다행, 응원, 긴장, 아픔, 천박한, 소중, 부담, 충격) 키워드로 분류하였다. 본 연구에서는 56개의 통일 감정 키워드(변수)

13) 본 연구를 위한 소셜 빅데이터의 수집 및 토픽 분류는 ‘(주)SK텔레콤 스마트인사이트’에서 수행함.

14) 주제분석에 사용되는 사전은 ‘21세기 세종계획’과 같은 범용사전도 있지만 대부분 분석의 목적에 맞게 사용자가 설계한 사전이다. 본 연구의 통일 관련 주제분석은 SKT에서 관련 문서 수집 후 원시자료(raw data)에서 나타난 상위 2000개의 키워드를 대상으로 유목화하여 사용자 사전을 구축하였다.

가 가지는 통일인식의 정도를 판단하기 위해 2차 요인분석을 통하여 14개의 요인(38개 변수)으로 축약한 다음 감성분석을 실시하였다. 요인분석에서 결정된 2개의 요인에 대한 주제어의 의미를 파악하여 ‘찬성, 보통, 반대’로 감성분석을 실시하였다. 일반적으로 감성분석은 긍정과 부정의 감성어 사전으로 분석해야 하나, 본 연구에서는 요인분석의 결과로 분류된 주제어의 의미를 파악하여 감성분석을 실시하였다. 따라서 본 연구에서는 찬성(13개: 관심, 필요, 중요, 노력, 신뢰, 평화(평화적), 든든, 다행, 공감, 행복, 사랑, 위대, 대박), 반대(18개: 쪽박, 천박, 경박, 빨갱이, 환장, 친절, 한심, 문제, 위협, 압박, 포기, 재앙, 혼란, 분열, 갈등, 아픔, 고통, 불신)로 분류하였다. 그리고 찬성과 반대의 감정을 동일한 횟수로 표현한 문서는 보통의 감정으로 분류하였다.

나. 통일에 대한 정책

통일에 대한 정책의 정의는 주제분석의 과정을 거쳐 ‘진보, 보수, 공산주의, 사회주의, 대북정책(대북정책, 정책, 통일정책), 창조경제(창조경제, 박근혜정권), 민주주의(민주정부, 민주주의), 햇볕정책’의 8개 정책으로 정책이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

다. 통일에 대한 안보

통일에 대한 안보의 정의는 주제분석 과정을 거쳐 ‘핵무기(미사일, 핵무기, 도발, 전쟁), 정상회담(회담, 정상회담, 비핵화), 휴전선(휴전선, DMZ, 판문점), 남북공동선언(10·4선언, 남북공동선언), 간첩(고정간첩, 간첩)’의 5개 안보로 안보가 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

라. 통일에 대한 이슈

통일에 대한 이슈의 정의는 주제분석을 거쳐 ‘천안함(연평도, 천안함), 민영화(민영화, 특검), 이산가족 상봉(이산가족 상봉, 이산가족), 선거(대선, 부정선거, 지방선거, 선거), 통일대박(통일대박론, 대박, 대박론)’의 5개 이슈로 이슈가 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

마. 통일에 대한 주변 국가

통일에 대한 주변 국가의 정의는 주제분석 과정을 거쳐 ‘중국, 독일(독일, 동독, 서독), 일본, 미국, 러시아(소련, 러시아), 유럽(유럽, 영국, 스위스, 네덜란드)’의 6개로 해당 국가가 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

바. 통일에 대한 관련 기관

통일에 대한 관련 기관의 정의는 요인분석과 주제분석의 과정을 거쳐 ‘범민련(범청학련, 범민련, 조국통일범민족연합, 한총련), 시민단체(실천연대, 시민단체, 전교조), 정부(외교부, 국방부, 통일부, 정부, 법무부, 국가보훈처, 해수부), 청와대(청와대, 박근혜정부), 국정원, 통일준비위원회(통일준비위원회, 통일준비위), 조국통일평화위원회, 자유통일포럼, 국회(국회, 외교통일위원회)’의 9개 관련 기관으로 해당 관련 기관이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

사. 통일에 대한 정당

통일에 대한 정당의 정의는 주제분석의 과정을 거쳐 ‘새정치민주연합(민주당, 새정치, 새천년민주, 새정치민주연합), 새누리당(새누리, 보수당, 새누리당), 통합진보당(통진당, 통합진보당), 자유선진당, 노동당’의 5개 정당으로 해당 정당이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

아. 통일에 대한 통일 방식

통일에 대한 통일 방식의 정의는 주제분석의 과정을 거쳐 ‘흡수통일(흡수, 흡수통일), 자유통일(자유통일, 자유북진통일, 북진통일), 평화통일’의 3개 통일 방식으로 해당 통일 방식이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

3. 분석 방법

본 연구에서는 대국민 통일인식을 설명하는 가장 효율적인 예측 모형을 구축하기 위해 특별한 통계적 가정이 필요하지 않은 데이터마이닝(data mining)의 연관분석(association analysis)과 의사결정나무(decision tree) 방법을 사용하였다. 소셜 빅데이터 분석에서 연관분석은 하나의 온라인 문서(transaction)에 포함된 둘 이상의 단어들에 대한 상호 관련성을 발견하는 것으로 동시에 발생한 어떤 단어들의 집합에 대해 조건과 연관규칙을 찾는 분석 방법이다. 전체 문서에서 연관규칙의 평가 척도는 지지도(support), 신뢰도(confidence), 향상도(lift)로 나타낼 수 있다. 연관분석 과정은 연구자가 지정한 최소 지지도를 만족시키는 빈발항목집합(frequent itemset)을 생성한 후 이들에 대해 최저 신뢰도 기준을 마련하

고 향상도가 1 이상인 것을 규칙으로 채택한다(박희창, 2010). 본 연구의 연관분석은 선형적 규칙(apriori principle)¹⁵⁾ 알고리즘을 사용하였다. 본 연구의 통일 감정에 사용된 연관분석의 측도는 지지도 0.001, 신뢰도 0.15를 기준으로 시뮬레이션하였다. 데이터마이닝의 의사결정나무 분석은 방대한 자료 속에서 종속변인을 가장 잘 설명하는 예측 모형을 자동적으로 산출해 줌으로써 각기 다른 속성을 가진 통일인식에 대한 요인을 쉽게 파악할 수 있다. 본 연구의 의사결정나무 형성을 위한 분석 알고리즘은 CHAID(Chi-squared Automatic Interaction Detection)를 사용하였다. CHAID(Kass, 1980)는 이산형인 종속변수의 분리 기준으로 카이제곱(χ^2 -검정)을 사용하며, 모든 가능한 조합을 탐색하여 최적분리를 찾는다. 정지규칙(stopping rule)으로 관찰치가 충분하여 상위 노드(부모마디)의 최소 케이스 수는 100, 하위 노드(자식마디)의 최소 케이스 수는 50으로 설정하였고, 나무 깊이는 3수준으로 정하였다. 본 연구의 기술분석, 다중응답분석, 로지스틱회귀분석, 의사결정나무분석은 SPSS v. 22.0을 사용하였고, 연관분석과 시각화분석은 R version 3.3.1을 사용하였다.

제3절 연구 결과

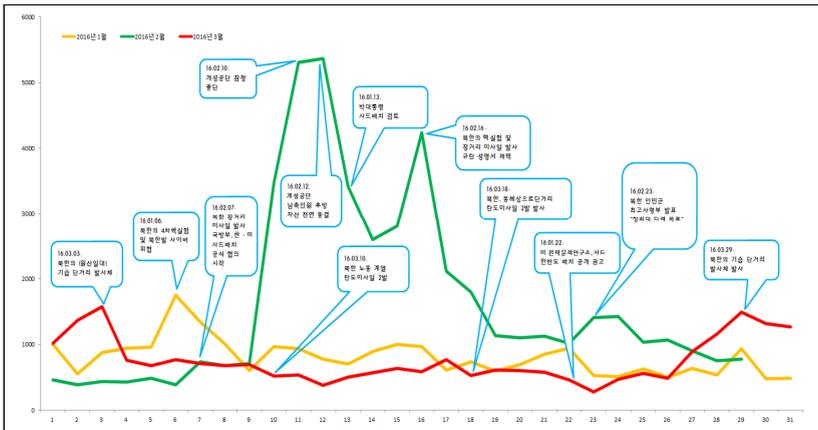
1. 통일 관련 문서(버즈) 현황

2016년 1월부터 3월까지 통일 관련 문서 버즈양 추이를 분석한 결과 북한과 관련된 이슈 발생 시에 커뮤니케이션이 급증하는 양상을 보이고 있다. 1월

15) 한 항목집합이 빈발하다면 이 항목집합의 모든 부분집합 역시 빈발항목집합이며, 한 항목집합이 비빈발하다면 이 항목집합을 포함하는 모든 집합은 비빈발항목 집합이다(이정진, 2011, p.123).

북한의 4차 핵실험으로 인해 문서량이 증가하였으며, 특히 2월 북한의 개성공단 잠정 폐쇄 발표 후 문서량이 급증한 것으로 나타났다(그림 2-1 참조).

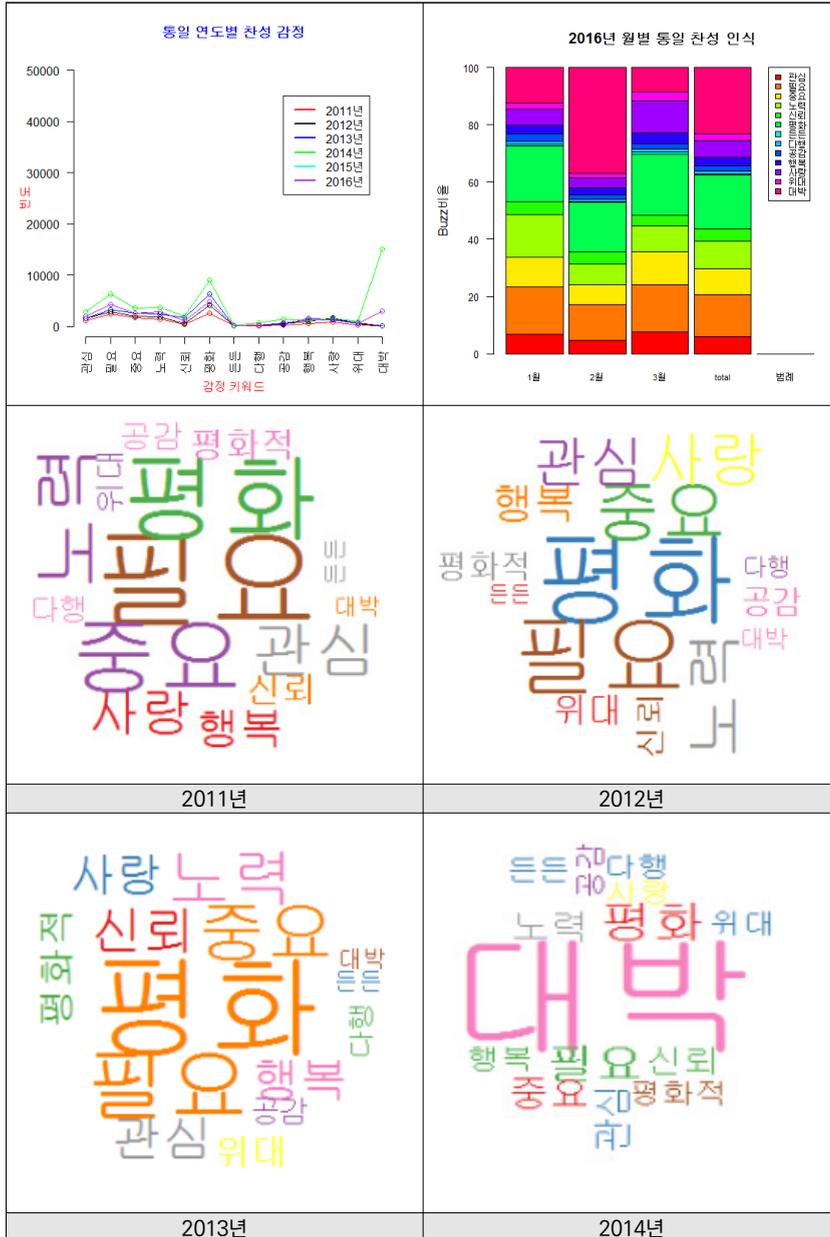
[그림 2-1] 통일 관련 문서(버즈)량 일별 추이: 2016년 1~3월

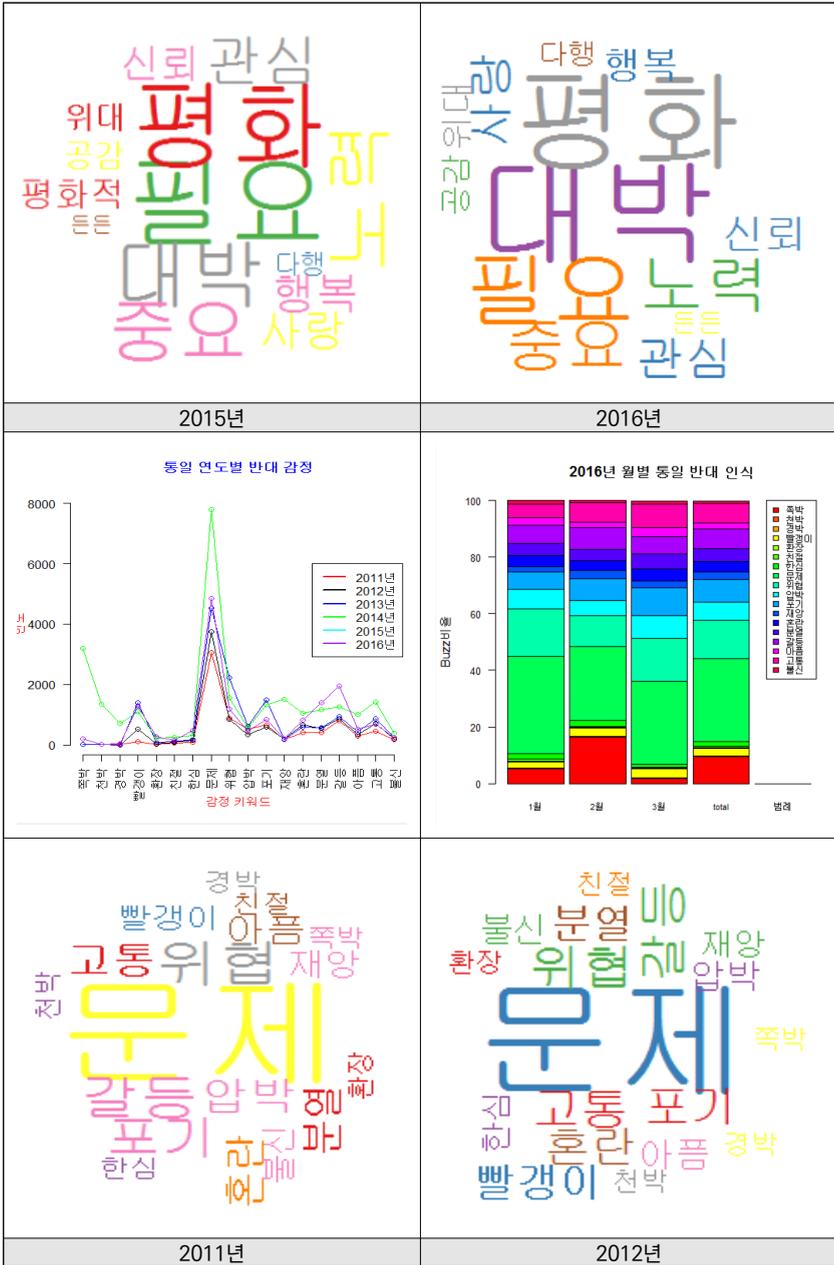


[그림 2-2]와 같이 연도별 통일 찬성의 감정 변화는 대박을 제외하고 2011년 대비 평균 2.23배 증가(평화 2.42배, 필요 1.68배, 중요 1.54배, 노력 1.95배, 관심 1.75배 등) 하였으며 찬성 감정의 표현 단어는 평화, 필요, 중요, 노력, 관심 등의 순으로 집중된 것으로 나타났다. 특히 대박은 2011년 56건에서 2014년 4만 7480건¹⁶⁾으로 크게 증가한 것으로 나타났다. 연도별 통일 반대의 감정 변화는 쪽박, 천박, 경박을 제외하고 2011년 대비 평균 3.25배 증가(문제 1.71배, 위협 1.64배, 갈등 1.56배, 포기 1.58배, 분열 2.29배 등) 하였으며, 반대 감정의 표현 단어는 문제, 위협, 갈등, 포기, 분열 등의 순으로 집중된 것으로 나타났다. 특히 쪽박, 천박, 경박은 2011년 30건에서 2014년 5254건으로 증가한 것으로 나타났다.

16) [그림 2-2]의 통일 찬성 인식에서 '대박'은 1만 5000건으로 표기함.

[그림 2-2] 통일 연도별 감정 변화







〈표 2-1〉, [그림 2-3]과 같이 통일인식에 대한 감정 키워드의 연관성 예측에서 찬성 감정의 경우 중요, 필요의 감정에 강하게 연결되어 있는 것으로 나타났으며 반대 감정의 경우 압박, 위협, 문제 감정에 강하게 연결되어 있는 것으로 나타났다.

〈표 2-1〉 통일인식의 감정 키워드 연관성 예측

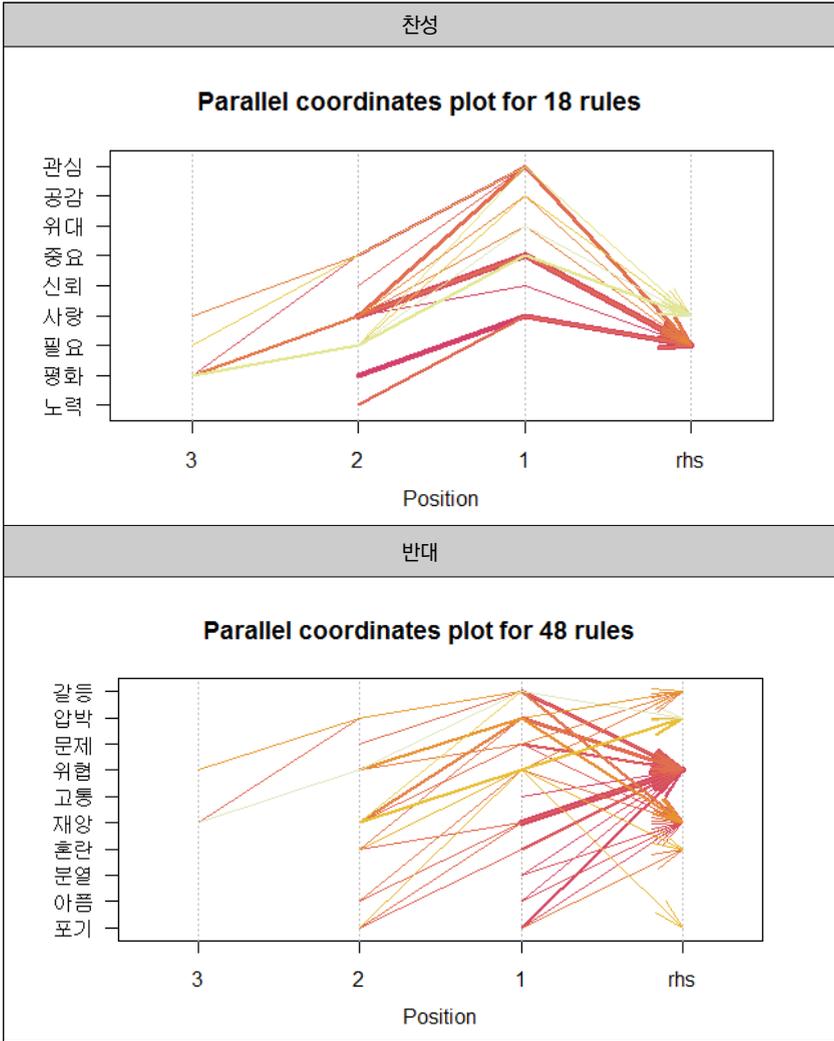
| | 규칙 | 지지도 | 신뢰도 | 향상도 |
|----|-------------------|---------|---------|----------|
| 찬성 | {필요, 위대}⇒{중요} | 0.00115 | 0.66265 | 26.69166 |
| | {필요, 노력, 평화}⇒{중요} | 0.00255 | 0.66038 | 26.60009 |
| | {관심, 필요, 평화}⇒{중요} | 0.00135 | 0.64824 | 26.11124 |

| | 규칙 | 지지도 | 신뢰도 | 향상도 |
|----|-------------------|----------|----------|----------|
| | {관심, 필요, 노력}⇒{중요} | 0.00149 | 0.61111 | 24.61564 |
| | {필요, 공감}⇒{중요} | 0.00109 | 0.60000 | 24.16808 |
| | {중요, 공감}⇒{필요} | 0.00109 | 0.79545 | 19.40719 |
| | {관심, 중요, 평화}⇒{필요} | 0.00135 | 0.78659 | 19.19080 |
| | {중요, 위대}⇒{필요} | 0.00115 | 0.78571 | 19.16955 |
| | {관심, 중요, 노력}⇒{필요} | 0.00149 | 0.77297 | 18.85869 |
| | {중요, 노력, 평화}⇒{필요} | 0.00255 | 0.72059 | 17.58063 |
| | {중요, 사랑}⇒{필요} | 0.00203 | 0.71168 | 17.36326 |
| | {관심, 신뢰}⇒{필요} | 0.00107 | 0.70068 | 17.09492 |
| | {관심, 중요}⇒{필요} | 0.00301 | 0.69976 | 17.07242 |
| | {관심, 노력, 평화}⇒{필요} | 0.00109 | 0.65625 | 16.01093 |
| | {중요, 노력}⇒{필요} | 0.00445 | 0.63731 | 15.54892 |
| | {중요, 신뢰}⇒{필요} | 0.00164 | 0.62550 | 15.26066 |
| | {관심, 노력}⇒{필요} | 0.00244 | 0.62069 | 15.14334 |
| | {중요, 평화}⇒{필요} | 0.00435 | 0.60876 | 14.85227 |
| 반대 | {문제, 위협, 포기}⇒{압박} | 0.001387 | 0.525692 | 45.87581 |
| | {위협, 포기}⇒{압박} | 0.001783 | 0.453581 | 39.58288 |
| | {문제, 위협}⇒{압박} | 0.003795 | 0.397814 | 34.71626 |
| | {문제, 분열}⇒{갈등} | 0.001335 | 0.400000 | 31.94238 |
| | {문제, 갈등}⇒{분열} | 0.001335 | 0.252964 | 31.75531 |
| | {문제, 포기}⇒{압박} | 0.002221 | 0.353234 | 30.82584 |
| | {위협, 압박}⇒{포기} | 0.001783 | 0.375824 | 26.38665 |
| | {문제, 위협, 압박}⇒{포기} | 0.001387 | 0.365385 | 25.65369 |
| | {문제, 압박, 포기}⇒{위협} | 0.001387 | 0.624413 | 25.61403 |
| | {압박, 포기}⇒{위협} | 0.001783 | 0.595819 | 24.44106 |
| | {문제, 압박}⇒{포기} | 0.002221 | 0.326187 | 22.90161 |
| | {문제, 압박}⇒{위협} | 0.003795 | 0.557427 | 22.8662 |
| | {문제, 갈등}⇒{압박} | 0.001199 | 0.227273 | 19.83353 |
| | {문제, 위협}⇒{포기} | 0.002638 | 0.276503 | 19.41329 |
| | {문제, 고통}⇒{포기} | 0.001022 | 0.276056 | 19.38194 |
| | {분열}⇒{갈등} | 0.001898 | 0.238220 | 19.02328 |
| | {압박}⇒{포기} | 0.002992 | 0.261147 | 18.33512 |
| | {포기}⇒{압박} | 0.002992 | 0.210103 | 18.33512 |
| | {문제, 혼란}⇒{위협} | 0.001147 | 0.421456 | 17.28853 |
| | {문제, 포기}⇒{위협} | 0.002638 | 0.419569 | 17.21112 |
| | {압박}⇒{위협} | 0.004744 | 0.414013 | 16.9832 |
| | {압박, 갈등}⇒{문제} | 0.001199 | 0.864662 | 16.61866 |

44 2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

| 규칙 | 지지도 | 신뢰도 | 향상도 |
|-------------------|----------|----------|----------|
| {문제, 분열}⇒{위협} | 0.001272 | 0.381250 | 15.63924 |
| {위협, 압박}⇒{문제} | 0.003795 | 0.800000 | 15.37587 |
| {위협, 고통}⇒{문제} | 0.001126 | 0.794118 | 15.26281 |
| {포기, 고통}⇒{문제} | 0.001022 | 0.790323 | 15.18987 |
| {위협, 압박, 포기}⇒{문제} | 0.001387 | 0.777778 | 14.94876 |
| {위협, 갈등}⇒{문제} | 0.001637 | 0.762136 | 14.64813 |
| {위협, 분열}⇒{문제} | 0.001272 | 0.748466 | 14.3854 |
| {압박, 포기}⇒{문제} | 0.002221 | 0.74216 | 14.2642 |
| {위협, 혼란}⇒{문제} | 0.001147 | 0.714286 | 13.72846 |
| {분열, 갈등}⇒{문제} | 0.001335 | 0.703297 | 13.51725 |
| {위협, 포기}⇒{문제} | 0.002638 | 0.671088 | 12.8982 |
| {문제, 갈등}⇒{위협} | 0.001637 | 0.310277 | 12.72785 |
| {문제, 고통}⇒{위협} | 0.001126 | 0.304225 | 12.47962 |
| {압박}⇒{문제} | 0.006809 | 0.594177 | 11.41998 |
| {포기}⇒{위협} | 0.003931 | 0.275988 | 11.32130 |
| {재앙}⇒{위협} | 0.001126 | 0.250580 | 10.27903 |

[그림 2-3] 통일인식 감정의 연관규칙에 대한 병렬좌표 시각화



〈표 2-2〉와 같이 통일과 관련하여 통일대박 감정을 포함하는 경우 통일을 찬성하는 문서는 59.8%, 중립의 문서는 12.6%, 반대하는 문서는 27.6%로 나타났다. 통일 관련 정책의 문서는 대북정책(36.1%), 민주주

의(18.9%), 창조경제(12.9%), 햇볕정책(8.8%), 보수(8.7%), 진보(6.1%) 등의 순으로 나타났다. 통일 관련 안보의 문서는 핵무기(72.7%), 정상회담(13.9%), 휴전선(6.3%) 등의 순으로 나타났다. 통일 관련 이슈의 문서는 통일대박(72.8%), 선거(20.0%), 천안함(3.5%), 이산가족 상봉(3.4%) 등의 순으로 나타났다. 통일 관련 국가의 문서는 중국(31.5%), 미국(25.3%), 일본(18.4%) 등의 순으로 나타났다. 통일 관련 기관의 문서는 정부(53.1%), 청와대(22.0%), 국회(12.8%), 국정원(7.9%) 등의 순으로 나타났다. 통일 관련 정당의 문서는 새누리당(64.1%), 새정치민주연합(19.1%), 노동당(11.2%), 통합진보당(4.8%) 순으로 나타났다. 통일 방식 관련 문서는 평화통일(58.2%), 자유통일(22.0%), 흡수통일(19.8%) 순으로 나타났다.

〈표 2-2〉 통일 관련 문서(버즈) 현황

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) |
|----|-------|---------------|--------------------|-------------|--------------|
| 연도 | 2011년 | 10,211(2.0) | 감정 [통일대박 포함] | 찬성 | 14,691(59.8) |
| | 2012년 | 53,884(10.6) | | | |
| | 2013년 | 83,268(16.4) | | | |
| | 2014년 | 166,952(32.9) | | 중립 | 3,105(12.6) |
| | 2015년 | 96,820(19.1) | | | |
| | 2016년 | 95,907(18.9) | | | |
| | 계 | 507,042 | | 반대 | 6,788(27.6) |
| 채널 | 카페 | 6,641(6.9) | 국가 | 계 | 24,584 |
| | SNS | 9,205(9.6) | | | |
| | 블로그 | 70,744(73.8) | | 중국 | 6,558(31.5) |
| | 게시판 | 5,111(5.3) | | | |
| | 뉴스 | 4,206(4.4) | | | |
| | 계 | 95,907 | | | |
| 정책 | 진보 | 659(6.1) | 러시아 | 2,308(11.1) | |
| | 보수 | 945(8.7) | | | |
| | 공산주의 | 515(4.7) | | | |
| | 사회주의 | 416(3.8) | | | |
| | | | 기관 | 범민련 | 141(0.7) |
| | | | | 미국 | 5,258(25.3) |
| | | | | 독일 | 1,637(7.9) |
| | | | | 일본 | 3,820(18.4) |
| | | | | 유럽 | 1,209(5.8) |
| | | | | 계 | 20,790 |

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) | |
|----|---------|-------------|-------|-----------|--------------|-------------|
| | 대북정책 | 3,917(36.1) | | 시민단체 | 332(1.7) | |
| | 창조경제 | 1,403(12.9) | | 정부 | 10,548(53.1) | |
| | 민주주의 | 2,048(18.9) | | 청와대 | 4,381(22.0) | |
| | 햇볕정책 | 957(8.8) | | 국정원 | 1,571(7.9) | |
| | 계 | 10,860 | | 통일준비위원회 | 149(0.7) | |
| 안보 | 핵무기 | 8,599(72.7) | | 조국평화통일위원회 | 216(1.1) | |
| | 정상회담 | 1,645(13.9) | | 국회 | 2,534(12.8) | |
| | 휴전선 | 741(6.3) | | 계 | 19,872 | |
| | 남북공동선언 | 109(0.9) | | 정당 | 새정치민주연합 | 940(19.9) |
| | 간첩 | 730(6.2) | | | 새누리 | 3,033(64.1) |
| | 계 | 11,824 | 통합진보당 | | 227(4.8) | |
| 이슈 | 천안함 | 342(3.5) | 노동당 | | 532(11.2) | |
| | 민영화 | 36(0.4) | 계 | 4,732 | | |
| | 이산가족 상봉 | 331(3.4) | 통일 방식 | 흡수통일 | 1,494(19.8) | |
| | 선거 | 1,970(20.0) | | 평화통일 | 4,402(58.2) | |
| | 통일대박 | 7,187(72.8) | | 자유통일 | 1,664(22.0) | |
| | 계 | 9,866 | | 계 | 7,560 | |

2. 통일 관련 국민 인식

〈표 2-3〉과 같이 통일의 필요성에 대한 국민의 통일인식은 조사 기관 별로 차이를 보이고 있으나, 2011년부터 2014년까지 통일대박을 제거한 소셜 빅데이터의 분석 결과는 서울대 통일평화연구원의 여론조사 결과와 비슷한 추이를 보이는 것으로 나타났다. 통일대박을 포함한 연도별 통일 관련 찬성 인식은 2011년 55.0%, 2012년 62.2%, 2013년 57.8%, 2014년 77.1%, 2015년 59.8%, 2016년 59.8%로 나타났다. 특히 아산 정책연구원의 2014년 통일인식 여론조사에서 80.5%(통일은 가능한 빨리 되어야+점진적으로 이루어져야)가 통일이 필요하다고 보고 있어 통일대박을 포함한 소셜 빅데이터 분석 결과(77.1%)와 비슷한 것으로 나타났다.

〈표 2-3〉 연도별 통일 관련 국민 인식

(단위: %)

| 연도 | 찬성 | | | 중립 | | | 반대 | | |
|------|------------------|-------------------|-------------------|------------------|-------------------|-------------------|------------------|-------------------|-------------------|
| | 소셜 ¹⁾ | 서울대 ²⁾ | KBS ³⁾ | 소셜 ¹⁾ | 서울대 ²⁾ | KBS ³⁾ | 소셜 ¹⁾ | 서울대 ²⁾ | KBS ³⁾ |
| 2011 | 55.0 | 53.7 | 74.4 | 18.3 | 25.0 | 19.9 | 26.7 | 21.3 | 5.7 |
| 2012 | 62.0 | 57.0 | 68.4 | 10.7 | 21.6 | 24.6 | 27.3 | 21.4 | 7.0 |
| 2013 | 57.7 | 54.8 | 69.1 | 8.7 | 21.5 | 21.2 | 33.6 | 23.7 | 9.7 |
| 2014 | 58.2 | 55.8 | - | 9.0 | 22.5 | - | 32.7 | 21.7 | - |
| 2015 | 56.2 | 51.0 | | 8.3 | 24.3 | | 35.5 | 24.8 | |
| 2016 | 59.8 | | | 12.6 | | | 27.6 | | |

주: 1) 본 연구의 소셜 빅데이터의 감성분석 결과.

2) 서울대학교 통일평화연구원이 매년 7월부터 8월 사이(3주간) 1200명을 대상으로 대면조사한 결과.

3) KBS 방송문화연구소. (2013). 2013년 국민 통일의식 조사.

3. 통일 관련 안보이슈의 동향

〈표 2-4〉와 같이 통일 관련 안보이슈에 대한 주변 국가의 문서는 미국·중국·일본·러시아 등 대부분의 국가에서 핵무기, 정상회담, 통일대박, 천

안함 등의 순으로 많이 언급된 것으로 나타났다.

〈표 2-4〉 통일 관련 안보·이슈의 국가별 버즈 현황

(단위: N, %)

| | 핵무기 | 정상 회담 | 휴전선 | 남북 공동선언 | 천안함 | 통일 대박 | 합계 |
|-----|-----------------|-----------------|----------------|--------------|--------------|----------------|-------------------|
| 중국 | 2,457 (37.5) | 951 (14.5) | 255 (3.9) | 77 (1.2) | 137 (2.1) | 497 (7.6) | 4,374 (27.9) |
| 독일 | 536 (32.7) | 187 (11.4) | 109 (6.7) | 3 (0.2) | 45 (2.7) | 133 (8.1) | 1,013 (6.5) |
| 미국 | 2,538 (48.3) | 959 (18.2) | 305 (5.8) | 77 (1.5) | 163 (3.1) | 344 (6.5) | 4,386 (28.0) |
| 일본 | 1,450 (38.0) | 552 (14.5) | 209 (5.5) | 12 (0.3) | 79 (2.1) | 156 (4.1) | 2,458 (15.7) |
| 러시아 | 1,327 (57.5) | 465 (20.1) | 164 (7.1) | 68 (2.9) | 91 (3.9) | 122 (5.3) | 2,237 (14.3) |
| 유럽 | 621 (51.4) | 285 (23.6) | 114 (9.4) | 74 (6.1) | 41 (3.4) | 67 (5.5) | 1,202 (7.7) |
| 합계 | 8,929 (57.0) | 3,399 (21.7) | 1,156 (7.4) | 311 (2.0) | 556 (3.5) | 1,319 (8.4) | 15,670 (100.0) |

통일과 관련한 기관별 안보·이슈에 대한 버즈는 〈표 2-5〉와 같이 나타났다. 청와대는 핵무기(31.0%), 정상회담(10.1%), 통일대박(9.6%) 등의 순으로 나타났고 정부는 핵무기(33.6%), 통일대박(10.3%), 정상회담(8.8%) 등의 순으로, 국회는 핵무기(24.4%), 선거(13.3%), 정상회담(7.7%), 통일대박(4.4%) 등의 순으로, 국정원은 핵무기(9.2%), 선거(5.2%), 정상회담(3.0%) 등의 순으로 나타났으며, 시민단체의 경우에도 핵무기(27.7%), 선거(17.5%), 정상회담(9.9%) 순으로 나타났다. 기관별로 핵무기, 정상회담 문서가 많이 언급된 것으로 나타났는데 이는 2016년 1월 북한의 4차 핵실험으로 인해 언급량이 증가한 것으로 판단된다.

〈표 2-5〉 통일 관련 안보·이슈의 기관별 버즈 현황

(단위: N, %)

| | 핵무기 | 정상 회담 | 휴전선 | 남북 공동 선언 | 천안함 | 이산가족 상봉 | 선거 | 통일대박 | 합계 |
|-------------|-----------------|-----------------|--------------|--------------|--------------|--------------|-----------------|-----------------|-----------------|
| 범민련 | 10 (7.1) | 5 (3.5) | 4 (2.8) | 1 (0.7) | 1 (0.7) | 3 (2.1) | 7 (5.0) | 1 (0.7) | 32 (0.3) |
| 시민단체 | 92 (27.7) | 33 (9.9) | 17 (5.1) | 0 (0.0) | 12 (3.6) | 10 (3.0) | 58 (17.5) | 24 (7.2) | 246 (2.0) |
| 정부 | 3548 (33.6) | 933 (8.8) | 322 (3.1) | 30 (0.3) | 209 (2.0) | 178 (1.7) | 601 (5.7) | 1,083 (10.3) | 6,904 (56.1) |
| 청와대 | 1358 (31.0) | 441 (10.1) | 126 (2.9) | 68 (1.6) | 95 (2.2) | 70 (1.6) | 334 (7.6) | 419 (9.6) | 2,911 (23.7) |
| 국정원 | 144 (9.2) | 47 (3.0) | 21 (1.3) | 3 (0.2) | 24 (1.5) | 4 (0.3) | 82 (5.2) | 40 (2.5) | 365 (3.0) |
| 통일준비 위원회 | 64 (43.0) | 33 (22.1) | 22 (14.8) | 1 (0.7) | 25 (16.8) | 10 (6.7) | 18 (12.1) | 53 (35.6) | 226 (1.8) |
| 조국 평화통일 위원회 | 75 (34.7) | 23 (10.6) | 28 (13.0) | 1 (0.5) | 6 (2.8) | 4 (1.9) | 13 (6.0) | 1 (0.5) | 151 (1.2) |
| 국회 | 618 (24.4) | 196 (7.7) | 84 (3.3) | 7 (0.3) | 57 (2.2) | 60 (2.4) | 336 (13.3) | 112 (4.4) | 1,470 (11.9) |
| 계 | 5,909 (48.0) | 1,711 (13.9) | 624 (5.1) | 111 (0.9) | 429 (3.5) | 339 (2.8) | 1,449 (11.8) | 1,733 (14.1) | 12,305 (100) |

4. 통일 관련 인식에 대한 연관성 분석

〈표 2-6〉과 같이 안보와 이슈에 대한 통일인식의 연관성 예측에서 가장 신뢰도가 높은 연관규칙은 {통일대박}⇒{찬성}이며, 두 변인의 연관성은 0.05, 신뢰도는 0.7, 향상도는 4.57로 나타났다. 이는 온라인 문서에서 통일대박이 언급되면 통일에 찬성할 확률이 70%이며, 통일대박이 언급되지 않은 문서보다 통일에 대한 찬성 확률이 4.57배 높아지는 것을 나타낸다. 또한 {이산가족 상봉}⇒{찬성} 두 개 변인의 연관성 지지도는 0.001, 신뢰도 0.44, 향상도는 2.88로 온라인 문서상에서 이산가족 상봉

이 언급되면 통일에 찬성할 확률이 44.1%이며, 이산가족 상봉이 언급되지 않은 문서보다 통일에 찬성할 확률이 2.8배 높아지는 것을 의미한다. {핵무기, 천안함}⇒{반대} 세 변인의 연관성 지지도는 0.001, 신뢰도 0.42, 향상도는 5.92로 온라인 문서상에서 핵무기와 천안함이 언급되면 통일에 반대할 확률이 41.9%이며 핵무기, 천안함이 언급되지 않은 문서보다 통일에 대해 반대할 확률이 5.92배 높아지는 것을 나타낸다.

〈표 2-6〉 안보와 이슈 요인에 대한 통일인식의 연관성 예측

| 규칙 | 지지도 | 신뢰도 | 향상도 |
|------------------|----------|----------|----------|
| {통일대박}⇒{찬성} | 0.052488 | 0.700431 | 4.572614 |
| {선거, 통일대박}⇒{찬성} | 0.001595 | 0.642857 | 4.196753 |
| {핵무기, 통일대박}⇒{찬성} | 0.005495 | 0.545549 | 3.561496 |
| {이산가족 상봉}⇒{찬성} | 0.001522 | 0.441088 | 2.879545 |
| {핵무기, 휴전선}⇒{찬성} | 0.001939 | 0.440758 | 2.877395 |
| {핵무기, 정상회담}⇒{찬성} | 0.004817 | 0.428571 | 2.797835 |
| {휴전선}⇒{찬성} | 0.003243 | 0.419703 | 2.73994 |
| {핵무기, 천안함}⇒{반대} | 0.001053 | 0.419087 | 5.921242 |
| {정상회담}⇒{찬성} | 0.00684 | 0.398784 | 2.603376 |
| {천안함}⇒{반대} | 0.001335 | 0.374269 | 5.288011 |
| {핵무기, 선거}⇒{찬성} | 0.001971 | 0.348709 | 2.276468 |
| {핵무기, 정상회담}⇒{반대} | 0.003785 | 0.336735 | 4.757692 |
| {천안함}⇒{찬성} | 0.001053 | 0.295322 | 1.927943 |
| {핵무기, 선거}⇒{반대} | 0.001668 | 0.295203 | 4.170894 |
| {핵무기}⇒{찬성} | 0.02614 | 0.291546 | 1.903292 |
| {정상회담}⇒{반대} | 0.00488 | 0.284499 | 4.019652 |
| {핵무기, 휴전선}⇒{반대} | 0.00122 | 0.277251 | 3.917255 |
| {선거}⇒{찬성} | 0.005234 | 0.254822 | 1.663552 |
| {휴전선}⇒{반대} | 0.001846 | 0.238866 | 3.37492 |
| {핵무기}⇒{반대} | 0.019060 | 0.212583 | 3.003563 |
| {핵무기, 통일대박}⇒{반대} | 0.001574 | 0.156315 | 2.208555 |

〈표 2-7〉과 같이 주요 주변 국가에 대한 통일인식의 연관성 예측에서 가장 신뢰도가 높은 연관규칙은 {중국, 일본, 러시아}⇒{찬성}이며 네 가

지 변인의 연관성 지지도는 0.004, 신뢰도는 0.41, 향상도는 2.69로 나타났다. 이는 온라인 문서에서 중국, 일본, 러시아가 언급되면 통일에 찬성할 확률이 41.2%이며 중국, 일본, 러시아가 언급되지 않은 문서보다 통일에 찬성할 확률이 2.69배 높아지는 것을 나타낸다. {미국, 일본, 러시아}⇒{찬성} 네 변인의 연관성 지지도는 0.003, 신뢰도 0.41, 향상도 2.65로 미국, 일본, 러시아가 언급되지 않은 문서보다 통일에 대한 찬성 확률이 2.65배 높아지는 것으로 나타났다. {중국, 미국, 일본, 러시아}⇒{찬성} 다섯 가지 변인의 연관성 지지도는 0.003, 신뢰도는 0.41, 향상도는 2.69로 중국, 미국, 일본, 러시아가 온라인상에서 언급되면 통일에 찬성할 확률이 41%이며 언급되지 않은 문서보다 통일에 찬성할 확률이 2.69배 높은 것으로 나타났다. 반면 {중국, 미국, 일본, 러시아}⇒{반대}의 연관성 지지도는 0.002, 신뢰도는 0.33, 향상도는 4.64로 온라인상에서 중국, 미국, 일본, 러시아가 언급되면 통일에 반대할 확률이 32%로 나타났으며 언급되지 않은 문서보다 통일에 반대할 확률이 4.64배 높은 것으로 나타났다.

{중국, 미국, 러시아}⇒{찬성}의 연관성 지지도는 0.005, 신뢰도 0.38, 향상도 2.46으로 중국, 미국, 러시아가 언급될 경우 통일에 찬성할 확률이 38%이며 중국, 미국, 러시아가 언급되지 않은 문서보다 통일에 대한 찬성의 확률이 2.64배 높은 것으로 나타났다. 반면 {중국, 미국, 러시아}⇒{반대}의 지지도는 0.005, 신뢰도 0.36, 향상도 5.10으로 중국, 미국, 러시아가 언급될 경우 통일에 반대할 확률이 36%이며 중국, 미국, 러시아가 언급되지 않은 문서보다 통일에 반대할 확률이 5.1배 높은 것으로 나타났다.

〈표 2-7〉 주변 국가(4국)의 통일인식의 연관성 예측

| 규칙 | 지지도 | 신뢰도 | 항상도 |
|------------------------|----------|----------|----------|
| {중국, 일본, 러시아}⇒{찬성} | 0.004004 | 0.41246 | 2.692654 |
| {중국, 미국, 일본, 러시아}⇒{찬성} | 0.003681 | 0.410943 | 2.682752 |
| {일본, 러시아}⇒{찬성} | 0.004348 | 0.407227 | 2.65849 |
| {미국, 일본, 러시아}⇒{찬성} | 0.003910 | 0.405844 | 2.649465 |
| {중국, 미국, 러시아}⇒{찬성} | 0.00538 | 0.376368 | 2.457034 |
| {미국, 러시아}⇒{찬성} | 0.005975 | 0.367072 | 2.396352 |
| {중국, 미국, 러시아}⇒{반대} | 0.005161 | 0.361050 | 5.101245 |
| {미국, 일본}⇒{찬성} | 0.007007 | 0.360322 | 2.352282 |
| {중국, 미국, 일본}⇒{찬성} | 0.005339 | 0.354325 | 2.313135 |
| {중국, 미국, 일본}⇒{반대} | 0.005255 | 0.348789 | 4.928005 |
| {중국, 러시아}⇒{찬성} | 0.006298 | 0.343377 | 2.241662 |
| {중국, 미국}⇒{찬성} | 0.011052 | 0.340727 | 2.224359 |
| {미국, 러시아}⇒{반대} | 0.005505 | 0.338245 | 4.779027 |
| {중국, 일본}⇒{찬성} | 0.006736 | 0.334369 | 2.182852 |
| {중국, 미국, 일본, 러시아}⇒{반대} | 0.00294 | 0.328289 | 4.63836 |
| {미국, 일본, 러시아}⇒{반대} | 0.003138 | 0.325758 | 4.602597 |
| {중국, 러시아}⇒{반대} | 0.00586 | 0.3195 | 4.514181 |
| {중국, 일본, 러시아}⇒{반대} | 0.003065 | 0.31579 | 4.461759 |
| {일본, 러시아}⇒{반대} | 0.003326 | 0.311523 | 4.401485 |
| {미국, 일본}⇒{반대} | 0.006048 | 0.310992 | 4.393975 |
| {러시아}⇒{찬성} | 0.007476 | 0.310659 | 2.028067 |
| {미국}⇒{찬성} | 0.016923 | 0.308673 | 2.015101 |
| {중국, 미국}⇒{반대} | 0.009874 | 0.304404 | 4.300891 |
| {중국, 일본}⇒{반대} | 0.005985 | 0.297101 | 4.197718 |
| {일본}⇒{찬성} | 0.011094 | 0.278534 | 1.818349 |
| {러시아}⇒{반대} | 0.006412 | 0.266465 | 3.764851 |
| {중국}⇒{찬성} | 0.017507 | 0.256023 | 1.671392 |
| {미국}⇒{반대} | 0.013096 | 0.238874 | 3.375029 |
| {중국}⇒{반대} | 0.01564 | 0.228728 | 3.23168 |
| {일본}⇒{반대} | 0.007518 | 0.188744 | 2.666738 |

5. 통일인식에 영향을 미치는 안보·이슈 요인

〈표 2-8〉과 같이 통일대박, 남북공동선언, 민영화, 이산가족 상봉, 휴전선, 선거, 간첩 순으로 통일의 찬성에 정적(+) 영향을 주는 것으로 나타났으나 정상회담, 핵무기, 천안함 순으로 통일의 찬성에 부적(-) 영향을 주는 것으로 나타났다.

〈표 2-8〉 통일인식에 영향을 미치는 안보·이슈 요인¹⁾

| 변수 | 찬성 | | | | 중립 | | | |
|---------|----------------|-------------------|-----------------|-------|----------------|-------------------|-----------------|-------|
| | b [†] | S.E. [‡] | OR [§] | P | b [†] | S.E. [‡] | OR [§] | P |
| 핵무기 | -.429 | .039 | .651 | .000 | .272 | .054 | 1.313 | .000 |
| 정상회담 | -.005 | .069 | .995 | .939 | .380 | .090 | 1.462 | .000 |
| 휴전선 | .263 | .099 | 1.301 | .008 | .584 | .126 | 1.793 | .000 |
| 남북공동선언 | 1.204 | .250 | 3.333 | .000 | -.422 | .470 | .656 | .370 |
| 간첩 | .061 | .198 | 1.063 | .756 | .591 | .237 | 1.806 | .013 |
| 천안함 | -.890 | .147 | .411 | .000 | -.735 | .189 | .479 | .000 |
| 민영화 | .584 | .535 | 1.793 | .275 | 1.493 | .574 | 4.448 | .009 |
| 이산가족 상봉 | .333 | .147 | 1.395 | .024 | .659 | .186 | 1.933 | .000 |
| 선거 | .087 | .082 | 1.091 | .288 | .824 | .096 | 2.279 | .000 |
| 통일대박 | 2.581 | .067 | 13.205 | 0.000 | 3.091 | .075 | 22.004 | 0.000 |

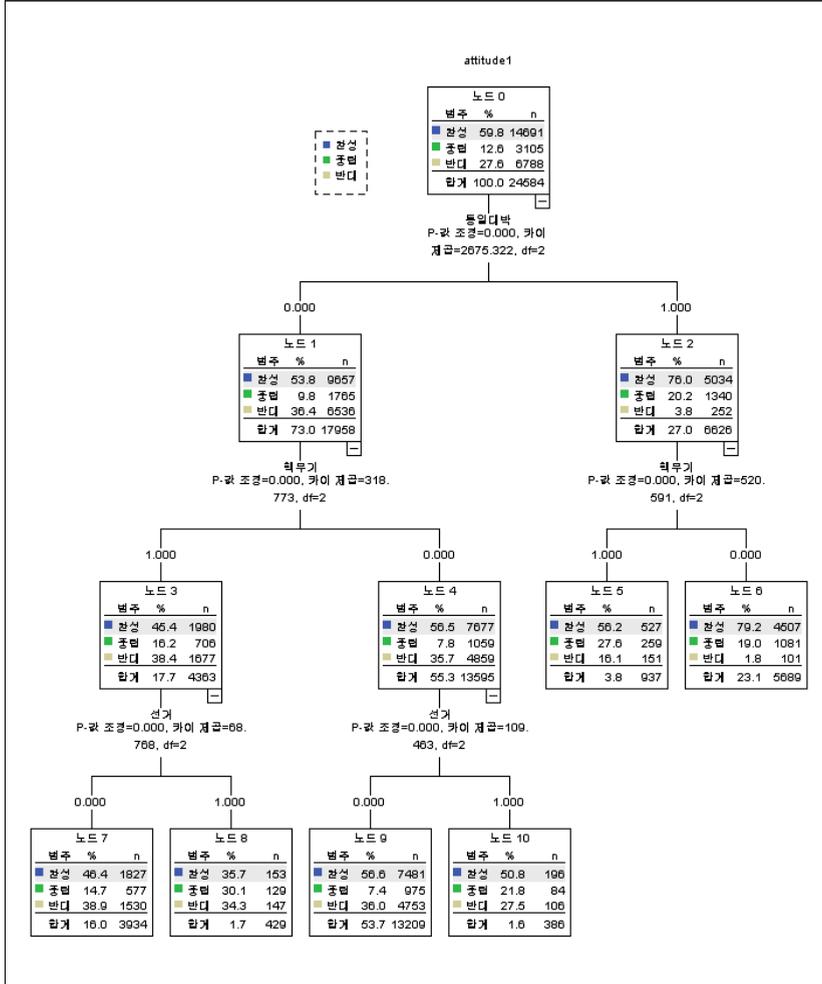
주: 1) 기본범주-반대, † Standardized coefficients, ‡ Standard error, § Adjusted odds ratio.

6. 통일인식 관련 예측 모형

[그림 2-4]와 같이 통일의 안보·이슈 요인이 통일인식에 미치는 영향에서는 ‘통일대박’의 영향력이 가장 큰 것으로 나타났다. ‘통일대박’이 있을 경우 통일에 대한 찬성 인식이 이전의 59.8%에서 76.0%로 증가한 반면, 반대 인식은 27.5%에서 3.8%로 감소하였다. ‘통일대박이 있고 핵무

기가 있는' 경우 통일에 대한 인식은 이전의 76.0%에서 58.2%로 감소한 반면, '통일대박이 있고 핵무기가 없는' 경우는 통일에 대한 인식이 79.2%로 증가한 것으로 나타났다. 반대의 인식은 3.7%에서 16.1%로 12.4% 포인트 증가한 것으로 나타났다. <표 2-9>의 통일인식 안보·이슈 요인의 예측 모형에 대한 이익도표와 같이 통일의 찬성에 가장 영향력이 높은 경우는 '통일대박'이 있고 '핵무기'가 없는 조합으로 나타났다. 즉 6번 노드의 지수(Index)가 132.6%로 뿌리마디와 비교했을 때 5번 노드의 조건을 가진 집단이 통일을 찬성할 확률이 1.32배로 나타났다. 반면 통일 반대에 영향력이 가장 높은 경우는 '통일대박'이 없고, '핵무기'가 있으며 '선거'가 없는 조합으로 나타났다. 즉 7번 노드의 지수(Index)가 140.9%로 뿌리마디와 비교했을 때 7번 노드의 조건을 가진 집단이 통일을 반대할 확률이 1.41배로 나타났다.

[그림 2-4] 통일인식 관련 안보이슈 요인의 예측 모형



〈표 2-9〉 통일인식 관련 안보·이슈 요인의 예측 모형에 대한 이익도표

| 구분 | 노드 | 이익지수 | | | | 누적지수 | | | |
|----|----|--------|-------|-------|-------|--------|-------|-------|-------|
| | | 노드(n) | 노드(%) | 이익(%) | 지수(%) | 노드(n) | 노드(%) | 이익(%) | 지수(%) |
| 찬성 | 6 | 5,689 | 23.1 | 30.7 | 132.6 | 5,689 | 23.1 | 30.7 | 132.6 |
| | 9 | 13,209 | 53.7 | 50.9 | 94.8 | 18,898 | 76.9 | 81.6 | 106.2 |
| | 5 | 937 | 3.8 | 3.6 | 94.1 | 19,835 | 80.7 | 85.2 | 105.6 |
| | 10 | 386 | 1.6 | 1.3 | 85.0 | 20,221 | 82.3 | 86.5 | 105.2 |
| | 7 | 3,934 | 16.0 | 12.4 | 77.7 | 24,155 | 98.3 | 99.0 | 100.7 |
| | 8 | 429 | 1.7 | 1.0 | 59.7 | 24,584 | 100.0 | 100.0 | 100.0 |
| 보통 | 8 | 429 | 1.7 | 4.2 | 238.1 | 429 | 1.7 | 4.2 | 238.1 |
| | 5 | 937 | 3.8 | 8.3 | 218.9 | 1,366 | 5.6 | 12.5 | 224.9 |
| | 10 | 386 | 1.6 | 2.7 | 172.3 | 1,752 | 7.1 | 15.2 | 213.3 |
| | 6 | 5,689 | 23.1 | 34.8 | 150.4 | 7,441 | 30.3 | 50.0 | 165.2 |
| | 7 | 3,934 | 16.0 | 18.6 | 116.1 | 11,375 | 46.3 | 68.6 | 148.3 |
| | 9 | 13,209 | 53.7 | 31.4 | 58.4 | 24,584 | 100.0 | 100.0 | 100.0 |
| 반대 | 7 | 3,934 | 16.0 | 22.5 | 140.9 | 3,934 | 16.0 | 22.5 | 140.9 |
| | 9 | 13,209 | 53.7 | 70.0 | 130.3 | 17,143 | 69.7 | 92.6 | 132.7 |
| | 8 | 429 | 1.7 | 2.2 | 124.1 | 17,572 | 71.5 | 94.7 | 132.5 |
| | 10 | 386 | 1.6 | 1.6 | 99.5 | 17,958 | 73.0 | 96.3 | 131.8 |
| | 5 | 937 | 3.8 | 2.2 | 58.4 | 18,895 | 76.9 | 98.5 | 128.2 |
| | 6 | 5,689 | 23.1 | 1.5 | 6.4 | 24,584 | 100.0 | 100.0 | 100.0 |

제4절 결론

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 주제분석과 감성분석 기술로 분류하고 네트워크분석과 데이터마이닝의 연관성분석, 의사결정나무분석 기법을 적용하여 분석함으로써 우리나라 국민의 통일인식에 대한 동향을 분석하고 통일인식의 연관규칙과 예측 모형을 개발하고자 하였다.

본 연구의 결과를 요약하면 다음과 같다. 첫째, 2016년 1월부터 3월까지 분석한 결과 북한과 관련된 이슈 발생 시에 온라인상에 통일 관련 문서가 급증하는 양상을 보이고 있으며, 연도별 통일 찬성의 감정 변화는 2011년 대비 평균 2.23배 증가하였으며, 찬성 감정의 표현 단어는 평화, 필요, 중요, 노력, 관심 등의 순으로 집중된 것으로 나타났다. 통일 반대의 감정 변화는 2011년 대비 평균 3.25배 증가하였으며, 반대 감정의 표현 단어는 문제, 위협, 갈등, 포기, 분열 등의 순으로 집중된 것으로 나타났다.

둘째, 안보와 관련된 문서는 핵무기, 정상회담, 통일대박, 천안함 등의 순으로 나타났으며 이슈와 관련한 문서는 핵무기, 정상회담, 통일대박 등의 순으로 나타났다. 그리고 통일 방식과 관련한 문서는 평화통일, 자유통일, 흡수통일 순으로 나타났다.

셋째, 통일의 필요성에 대한 국민 인식은 통일대박 감정을 포함할 경우 찬성은 2011년 55.0%, 2012년 62.2%, 2013년 57.8%, 2014년 77.1%, 2015년 56.2%로 나타났으며 2016년은 찬성이 59.8%, 반대가 27.6%로 나타났다.

넷째, 통일인식에 영향을 미치는 안보·이슈 요인에 대한 다중로지스틱 회귀분석 결과 통일대박, 남북공동선언, 민영화, 이산가족 상봉, 휴전선,

선거, 간첩 순으로 통일의 찬성에 정적(+) 영향을 주는 것으로 나타났으나 정상회담, 핵무기, 천안함 순으로 통일의 찬성에 부적(-) 영향을 주는 것으로 나타났다.

마지막으로 안보·이슈 요인이 통일인식에 미치는 영향은 통일대박의 영향력이 가장 큰 것으로 나타났는데 온라인 문서 중 통일대박이 있을 경우 통일에 대한 찬성 인식이 이전의 59.8%에서 76.0%로 증가하였고, 특히 통일대박이 있고 핵무기가 없는 경우 통일에 찬성하는 확률이 가장 높은 것으로 나타났다. 주변 4개국이 통일인식에 미치는 영향은 중국이 가장 높은 것으로 나타났으며 온라인 문서상에서 중국, 미국, 러시아가 없는 조합이 통일에 찬성할 확률이 가장 높은 반면 중국이 없고 미국이 있는 조합이 통일에 반대할 확률이 가장 높은 것으로 나타났다.

본 연구에서 우리나라의 통일 관련 인식에 대한 예측을 근거로 다음과 같은 정책적 함의를 도출할 수 있다.

첫째, 최근 북한과 관련된 이슈 발생 시에 온라인상에서 커뮤니케이션이 급증하는 것으로 나타났다. 2016년 상반기 북한의 핵실험과 개성공단 폐쇄로 인하여 남북 간의 긴장이 고조되면서 국민들의 관심이 급증한 것으로 보인다. 북한의 4차 핵실험 이후 전국 만 19세 이상의 남녀 1005명을 대상으로 '4차 북한 핵실험의 한반도 평화 위협 정도'에 대해 묻은 결과 위협적이라고 응답한 사람이 61%로 나타났다. 3차 핵실험 이후 '위협적이다'라고 응답한 사람 76%에 비해 15% 포인트 낮아졌으나 여전히 북한 핵실험 등의 이슈는 위협적이라고 인식하는 것으로 볼 수 있다.

둘째, 통일 관련 국민 인식은 소셜 빅데이터의 분석 결과와 정기적인 여론조사 결과가 비슷한 추이를 보이는 것으로 나타났다. 이는 본 연구에서 제시한 통일 관련 감정 키워드의 감성분석 방법으로 통일의 필요성에 대한 인식을 찬성, 보통, 반대로 분류하는 것의 타당성이 어느 정도 확보

된 것으로 볼 수 있다.

셋째, 통일대박이 언급된 온라인 문서의 경우 통일에 대한 찬성 확률이 높은 것은 통일대박이 통일에 대한 긍정적인 담론을 확산시키는 것으로 나타났다. 그뿐만 아니라 선거와 통일대박이 동시에 언급된 문서의 경우 찬성 확률이 높은 것은 선거를 통해 통일대박에 대한 긍정적인 담론이 확산되기 때문으로 본다. 핵무기와 천안함이 동시에 언급된 온라인 문서의 경우 통일에 대한 반대 확률이 높은 것은 북한의 위협으로 인해 안보 불안이 증가하면서 부정적인 담론이 확산되기 때문으로 볼 수 있다.

마지막으로, 통일 과정에 가장 영향력이 높은 것은 ‘통일대박’이 있고 ‘핵무기’가 없는 경우 통일에 대한 인식이 찬성으로 나타났으며, ‘통일대박’이 없고 ‘핵무기’가 있으며 ‘선거’가 없는 조합의 문서들이 온라인상에 언급될 때 통일에 반대하는 확률이 높은 것으로 나타나 ‘통일대박론’이 통일의 필요성에 대한 인식 변화에 어느 정도 영향을 주었다고 추정할 수 있다.

본 연구는 개개인의 특성을 가지고 분석한 것이 아니고 그 구성원이 속한 전체 집단의 자료를 대상으로 분석한 것이기 때문에 이를 개인에게 적용하였을 경우 생태학적 오류(ecological fallacy)가 발생할 수 있다(송태민 등, 2014).

또한 본 연구에서 감성분석 결과 정의된 통일인식은 온라인 문서 내에서 발생된 감정 단어의 빈도로 정의되었기 때문에 기존의 조사 등을 통한 통일인식의 조작적 정의와 다를 수 있으며, 2011~2015년 기간 1/4분기(15개월간)의 제한된 소셜 빅데이터를 분석함으로써 전체적인 통일 관련 인식의 예측에 한계가 있을 수 있다. 그럼에도 불구하고 본 연구는 소셜 빅데이터에서 통일 관련 주요 이슈에 대한 실제적인 내용을 빠르게 효과적으로 파악함으로써 기존 통일 관련 인식의 정보수집체계의 한계를 보완

할 수 있는 새로운 분석 방법을 제시하였다는 점에서 정책적·분석방법론적 의의를 가진다고 할 수 있다(송주영, 송태민, 2014).

끝으로 통일에 대한 찬반, 통일 방법 등에 대한 국민의 통일의식 조사와 더불어 소셜미디어에서 수집된 빅데이터를 활용, 분석할 경우 통일인식의 예측은 더욱 신뢰성이 있을 것으로 본다.



참고문헌 <<

- 김병조. (2015). 한국인의 통일인식 2007~2015. 세대별 격차와 세대내 분화, 서울대학교 통일평화연구원 통일과 평화 7(2), pp.3-41.
- 박희창. (2010). 연관규칙 마이닝에서의 평가기준 표준화 방안. 한국데이터정보과학회지, 21(5), pp.891-899.
- 송태민, 송주영, 진달래. (2014). 소셜 빅데이터를 활용한 인터넷 중독 위험예측 모형. 보건사회연구. 34(3). pp.106-134.
- 송태민, 송주영. (2014). 빅데이터 분석방법론. 서울: 한나래 출판사.
- 아산정책연구원. (2015). 한국인의 대북·통일 인식 변화, 아산정책연구원.
- 연세대학교 산학협력단. (2014). 평화통일에 대한 국민공감대 형성연구.
- 이원태, 차미영, 양해륜. (2011). 소셜미디어 유력자의 네트워크 특성: 한국의 트위터 공동체를 중심으로, 언론정보연구 48(2). pp.44-79.
- 최영준, 황태희, 최우선, 주형민. (2016). 박근혜 정부의 통일정책의 정책수용성에 대한 실증연구: 2013년과 2015년 국민인식조사 비교. 국제관계연구 21(1), pp.5-32.
- 최윤철. (2016). 통일 국민과 이주자의 사회통합. 통일 인문학, 66. pp.49-83.
- 최재원, 김성철. (2016). 통일인식 제고를 위한 통일 유관기관들의 ICT 활용방안, 언론정보연구, 53(1), pp.149-176.
- 홍종윤. (2016). 미디어의 평화통일 지향성 지수 개발 연구. 언론정보연구, 53(1), pp.5-56.



제 3 장

2016 소셜 빅데이터를 활용한 담배 위험 예측

제1절 서론

제2절 연구 방법

제3절 연구 결과

제4절 결론

참고문헌



3

2016 소셜 빅데이터를 활용한 << 담배 위험 예측

제1절 서론

우리나라 19세 이상 성인 남성 흡연율은 1998년 66.3%에서 2005년 51.6%, 2014년 43.1%였으며 2015년 가추계치(preliminary estimates)일지라도 39.3%로 최초로 30%대로 진입했다(보건복지부, 2015; 기획재정부, 보건복지부 보도자료, 2016. 5. 10.). 2012년 15세 이상 남성 흡연율은 경제협력개발기구(OECD) 평균 24.9%보다 높은 37.6%로 세계에서 가장 높은 위치를 차지하고 있으며(OECD, 2014), 세계보건기구(WHO)에서 공표한 우리나라 15세 이상 남성 흡연율(2013년, 연령표준화, Cigarette)은 51.1%로 서태평양 지역 국가 중 두 번째로 높았다(WHO, 2015). 이와 같이 우리나라 남성 흡연율이 OECD 회원국 중 최고 수준에 달하는 상황에서 현 정부는 2015년 1월 1일부터 담뱃값을 2000원 인상하는 등 범정부 차원의 금연 종합대책을 발표하였고(보건복지부 보도자료, 2014. 9. 11.), 2016년 5월 담뱃갑 경고 그림 확정 및 시행 등 비가격 정책 추진 방안을 발표하였다(기획재정부, 보건복지부 보도자료, 2016. 5. 10.). 전 세계적으로 흡연으로 인해 매년 600만 명이 사망하고 있으며(WHO, 2008), 전체 암 사망의 30.5%, 호흡기질환 사망의 19.8%, 심혈관질환 사망의 11.4%가 흡연으로 인해 사망한 것으로 예측되었다(Zheng et al., 2014). 우리나라는 1985년 2만 4338명, 2003년 4만 6207명, 2012년 5만 8155명이 흡연으로 인한 사망자 수로 보고되었고(정금지 등, 2013), 흡연으로 인해 발생하는 2013년 기준 사회경제

적 비용[직접의료비+직접비의료비(교통비, 간병비)+간접비(조기사망비, 생산성 손실비)+재산 피해 비용]의 규모는 남성의 경우 5조 9734억 원이고, 여성의 경우 1조 1523억 원으로 나타났다(이선미 등, 2015).

담배 연기는 사람에게 치명적인 화학물질 7000개 이상을 함유하고 있으며, 이로 인해 폐암을 비롯한 각종 암과 심혈관질환, 호흡기질환, 만성 질환 등 다양한 질병과 관련 있는 것으로 알려져 있다(Carter et al., 2015; CDC, 2010; Thun et al., 2013). 우리나라는 1995년 국민건강 증진법이 제정됨에 따라 본격적으로 담배 판매, 광고, 금연구역 확대 등을 추진하였고, 청소년보호법·학교보건법 등에서도 청소년 흡연과 관련하여 제도적으로 규제하고 있다. 또한 2005년 WHO 담배규제기본협약(FCTC) 비준 이후 다양한 흡연 예방 및 담배규제 정책을 시행하고 있다(강은정, 이재희, 2011). 담배규제 정책들은 선진국과 개발도상국의 차이는 있을지라도 실제 사례를 통해 효과가 입증되었다. 미국은 지속적으로 담뱃값이 인상됨에 따라 담배 소비량이 줄어들었고(Campaign for Tobacco-Free Kids, 2013), 터키도 2008년에 비해 2012년 담뱃값이 42.1% 증가했을 때 흡연율이 14.6% 감소하였다(CDC, 2014). 우리나라는 2004년 12월 2000원에서 500원 인상된 후 10년 동안 추가적인 인상이 이루어지지 않아 흡연율의 상승과 하락을 반복하여 담뱃값 인상에 대한 금연 효과는 크지 않은 것으로 나타났다(보건복지부 보도자료, 2014. 9. 11.).

비가격 정책 중 하나인 담뱃갑 경고 그림은 2000년 12월 캐나다에서 제일 먼저 시작되었고, 흡연자의 63%는 담뱃갑 경고 그림을 통해 적어도 한 번 이상의 금연 효과를 경험했으며(Hammond et al., 2004) 담뱃갑 경고 그림과 경고 문구로 인한 금연 행동 효과를 평가하는 연구에 따르면 경고 그림군에서 경고 문구군에 비해 금연 시도가 6% 포인트 더 높게(40% vs 34%, OR=1.29) 나타나(Brewer et al., 2016) 세계 여러 나라

에서도 법안으로 정하여 시행하고 있다. 우리나라도 2016년 12월 23일부터는 담뱃갑 앞뒷면 면적의 30% 이상에 해당하는 크기의 경고 그림과 20% 이상에 해당하는 크기의 경고 문구를 의무적으로 표기해야 한다. 이외에도 담배 광고 및 판촉에 대한 규제 강화, 전자담배 사용 규제 강화, 소포장 금지 및 가향 첨가 규제, 금연지원서비스 및 금연 캠페인 강화 등이 있다(기획재정부, 보건복지부 보도자료, 2016. 5. 10.).

최근 2015년 1월 1일 담뱃값 인상으로 국가 금연지원서비스를 2014년 113억 원에서 2015년 1475억 원으로 확대하였으며, 추가 확보된 재원을 금연 성공률이 가장 높은 약물·상담 치료에 지원하고, 건강보험 병의원 금연치료서비스 시행과 학교, 군부대, 사업장 등에 대한 금연 지원을 대폭 확대했다. 또한 금연광고와 금연 캠페인을 연중 실시하고 보건소 금연클리닉, 금연상담전화, 온라인 상담 등 1:1 맞춤형 금연상담서비스도 대폭 강화하였으며, 공공건물 등의 금연구역 확대로 간접흡연의 폐해도 감소시켰다(기획재정부, 보건복지부 보도자료, 2016. 5. 10.). 2016년 금연 정책의 추진 방향은 흡연자 금연지원서비스 강화 및 비흡연자 간접흡연 예방이며, 궁극적으로는 제3차 국민건강증진종합계획(Health Plan 2020)에 따라 성인 남성 흡연율을 29%, 청소년 남성 흡연율을 12%로 감소시키는 데 목적이 있다(보건복지부, 한국건강증진개발원, 2016).

본 연구는 우리나라 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 바탕으로 우리나라 국민의 담배에 대한 위험 예측 모형과 연관규칙을 파악한다.

제2절 연구 방법

1. 연구 대상

본 연구는 국내의 SNS, 온라인 뉴스 사이트 등 인터넷을 통해 수집된 소셜 빅데이터를 대상으로 하였다. 본 분석에서는 200개의 온라인 뉴스 사이트, 10개의 게시판, 1개의 SNS(트위터), 4개의 블로그 등 총 217개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서(버즈)를 소셜 빅데이터로 정의하였다. 담배 관련 토픽(topic)¹⁷⁾은 2016년 1~3월 동안 해당 채널에서 요일, 주말, 휴일을 고려하지 않고 매시간 단위로 수집하였으며, 수집된 총 18만 1395건의 텍스트(Text) 문서를 본 연구 분석에 포함시켰다. 담배 토픽은 모든 관련 문서를 수집하기 위해 ‘담배’를 사용하였으며, 토픽과 같은 의미로 사용되는 토픽 유사어로는 ‘흡연, 담뱃값, 담배 피, 담배 추천, 담배가격, 혼녀생정담배, 중딩담배, 고딩담배, 중고딩담배, 청소년 담배’ 용어를 사용하였다. 본 연구를 위한 소셜 빅데이터의 수집¹⁸⁾은 크롤러(Crawler)를 사용하였고, 이후 주제분석을 통해 분류된 명사형 어휘를 유목화(categorization)하여 분석요인으로 설정하였다.

2. 연구 도구

담배와 관련하여 수집된 문서는 주제분석¹⁹⁾의 과정을 거쳐 다음과 같

-
- 17) 토픽은 소셜 분석 및 모니터링의 ‘대상이 되는 주제어’를 의미하며, 문서 내에 관련 토픽이 포함된 문서를 수집함.
 - 18) 본 연구를 위한 소셜 빅데이터의 수집 및 토픽 분류는 ‘(주)SK텔레콤 스마트인사이트’에서 수행함.
 - 19) 주제분석에 사용되는 사전은 ‘21세기 세종계획’과 같은 범용사전도 있지만 대부분 분석의 목적에 맞게 사용자가 설계한 사전이다. 본 연구의 담배 관련 주제분석은 ‘(주)SK텔레콤 스마트인사이트’에서 관련 문서 수집 후 원시자료(raw data)에서 나타난 상위 2000

이 정형화 데이터로 코드화하여 사용하였다.

가. 담배 관련 감정

본 연구의 담배 감정 키워드는 문서 수집 이후 주제분석을 통하여 총 66개(걱정, 고민, 고생, 고통, 갈끔, 다짐, 대단, 두려움, 만족, 믿음, 부담, 불가능, 불리, 불만, 불안, 불편함, 사랑, 스트레스, 실패, 어려움, 여유, 염려, 욕구, 위험, 유혹, 응원, 의지, 의지력, 자신감, 재미, 조심, 즐거움, 짜증, 창피, 최고, 최선, 충격, 치유, 편안, 포기, 피곤, 필요, 행복, 호기심, 파이팅, 활력, 후회, 희망, 힐링, 힘들다, 성공, 도움, 문제, 추천, 관심, 도전, 결심, 잘못, 혐오, 심각, 논란, 불편, 고발, 이해, 지적, 끔찍) 키워드로 분류하였다. 본 연구에서는 66개의 담배 감정 키워드(변수)가 가지는 담배 감정 정도를 판단하기 위해 요인분석을 통하여 12개의 요인(44개 변수)으로 축약한 후 감성분석을 실시하였다. 일반적으로 감성분석은 긍정과 부정의 감성어 사전으로 분석해야 하나 본 연구에서는 요인분석의 결과로 분류된 주제어의 의미를 파악하여 감성분석을 실시하였다. 요인분석에서 결정된 12개 요인에 대한 주제어의 의미를 파악하여 ‘일반군, 잠재군, 위험군’으로 감성분석을 실시하였다. 따라서 본 연구에서 일반군은 23개 변수(스트레스, 위험, 문제, 조심, 성공, 실패, 결심, 의지, 욕구, 논란, 지적, 부담, 불만, 염려, 걱정, 짜증, 창피, 불안, 끔찍, 충격, 불편, 파이팅, 응원), 위험군은 16개 변수(믿음, 사랑, 희망, 행복, 최선, 추천, 갈끔, 만족, 고민, 최고, 즐거움, 여유, 대단, 피곤, 힐링, 치유)로 분류하였다. 그리고 일반군과 위험군의 감정을 동일한 횟수로 표현한 문서는 잠재군으로 분류하였다. 일반군은 담배를 혐오적으로 생각하는 감정이

개의 키워드를 대상으로 유목화하여 사용자 사전을 구축하였음.

고, 위험군은 담배를 애호적으로 생각하는 감정이며, 잠재군은 담배를 보통으로 생각하는 감정을 나타낸다.

나. 담배와 관련된 정책

담배와 관련된 정책의 정의는 주제분석 과정을 거쳐 ‘담뱃값 인상, FCTC (담배규제기본협약 등), 금연관련법(국민건강증진법, 학교보건법 등), 흡연규제(금연구역, 벌금 부과 등), 금연광고(공익광고, 금연 캠페인 등), 금연사업(금연상담전화, 금연클리닉 등)’ 6개 정책으로 정책이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

다. 담배와 관련된 질환

담배와 관련된 질환의 정의는 주제분석을 거쳐 ‘가래, 간암, 감기, 동맥경화, 고혈압, 구토, 뇌혈관질환, 당뇨병, 대장암, 두통, 마비, 만성질환, 발기부전, 불면증, 사망, 식도암, 심혈관질환, 염증, 우울증, 위암, 유방암, 폐암, 치매, 후두암, 구강암’의 25개로 질환이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

라. 담배에 대한 금연 도구

담배에 대한 금연 도구의 정의는 주제분석 과정을 거쳐 ‘금연껌(금연껌, 니코틴로렌즈, 니코틴껌, 니코틴엘로젠즈, 사탕, 트로키), 금연약(금연약, 약물, 니코엔, 니코스텝, 챔픽스, 니코피온, 니코그린, 니코레스, 부프로피온, 흡연욕구저해제, 챔픽스정, 바레니클린, 웰부트린), 전자담배

(전자담배, 스모키전자담배, 애니스틱, 라스트스틱), 금연패치(니코레트, 니코틴패치, 패치, 금연패치, 니코틴보조제, 금연보조제, 보조제, 금연침), 보조제(물담배, 파이프담배, 리엔파이프, 롤링토바코, 금연파이프, 금연초, 건향초)의 5개로 금연 도구가 있는 경우는 '1', 없는 경우는 '0'으로 코드화하였다.

마. 담배에 대한 도움치료

담배에 대한 도움치료의 정의는 주제분석 과정을 거쳐 '금연클리닉, 금연상담전화, 병원, 금연교실'의 4개로 해당 치료가 있는 경우는 '1', 없는 경우는 '0'으로 코드화하였다.

바. 담배와 관련된 피해

담배와 관련된 피해의 정의는 주제분석을 거쳐 '간접흡연, 알코올, 중독, 기억력, 담배꽂초, 도박·마약, 이혼, 정신건강, 폭력'의 9개 피해로 해당 피해가 있는 경우는 '1', 없는 경우는 '0'으로 코드화하였다.

사. 담배에 대한 유해물질

담배에 대한 유해물질의 정의는 주제분석 과정을 거쳐 '니코틴, 발암물질, 유해물질, 일산화탄소, 타르, 화학물질, 노폐물'의 7개 유해물질로 해당 유해물질이 있는 경우는 '1', 없는 경우는 '0'으로 코드화하였다.

아. 담배에 대한 장소

담배에 대한 장소의 정의는 주제분석 과정을 거쳐 ‘PC방, 가정, 금연 건물, 아파트, 공공장소, 흡연구역, 직장, 술집, 식당, 학교’의 10개 장소로 해당 장소가 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화하였다.

자. 담배에 대한 관련 기관

담배에 대한 관련 기관의 정의는 주제분석 과정을 거쳐 ‘청와대, 국회, 보건복지부, 여성가족부, 기획재정부, 지방자치단체, 공공기관, 세계보건기구, 금연단체(한국금연운동협의회, 한국건강관리협회, 한국보건의료연구원 등), 담배회사’의 10개 기관으로 해당 기관이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

3. 분석 방법

본 연구에서 우리나라 담배의 위험을 설명하는 가장 효율적인 예측 모형을 구축하기 위해 특별한 통계적 가정이 필요하지 않은 데이터마이닝의 연관분석(association analysis)과 의사결정나무(decision tree) 방법을 사용하였다. 소셜 빅데이터 분석에서 연관분석은 하나의 온라인 문서(transaction)에 포함된 둘 이상의 단어들에 대한 상호 관련성을 발견하는 것으로, 동시에 발생한 어떤 단어들의 집합에 대해 조건과 연관규칙을 찾는 분석 방법이다. 전체 문서에서 연관규칙의 평가 측도는 지지도(support), 신뢰도(confidence), 향상도(lift)로 나타낼 수 있다. 지지도는 자주 발생하지 않는 규칙을 제거하는 데 이용되며 신뢰도는 단어들의

연관성 정도를 파악하는 데 이용할 수 있다. 향상도는 연관규칙($X \rightarrow Y$)에서 단어 X 가 없을 때보다 있을 때 단어 Y 가 발생할 비율을 나타낸다. 연관분석 과정은 연구자가 지정한 최소 지지도를 만족시키는 빈발항목집합(frequent itemset)을 생성한 후 이들에 대해 최저 신뢰도 기준을 마련하고 향상도가 1 이상인 것을 규칙으로 채택한다(박희창, 2010). 본 연구의 연관분석은 선형적 규칙(apriori principle) 알고리즘을 사용하였으며, 담배 감정에 사용된 연관분석의 측도는 지지도 0.001, 신뢰도 0.01을 기준으로 시뮬레이션하였다. 본 연구의 의사결정나무 형성을 위한 분석 알고리즘은 CHAID(Chi-squared Automatic Interaction Detection)를 사용하였다. 정지규칙(stopping rule)으로 관찰치가 충분하여 상위 노드(부모마디)의 최소케이스 수는 100으로, 하위 노드(자식마디)의 최소 케이스 수는 50으로 설정하였고 나무 깊이는 3 수준으로 정하였다. 본 연구의 기술분석, 다중응답분석, 의사결정나무분석은 SPSS v. 21.0을 사용하였고, 연관분석은 R version 3.3.1을 사용하였다.

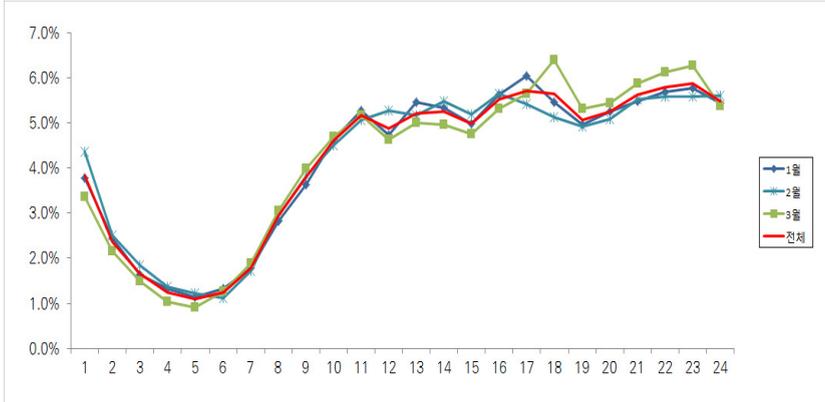
제3절 연구 결과

1. 담배 관련 온라인 문서 현황

담배와 관련된 버즈는 월별로 비슷하게 오전 7시부터 증가하여 11시 이후 감소하며, 다시 낮 12시 이후 증가하여 오후 2시 이후 감소하고, 3시 이후 증가하여 5시 이후 감소하고, 7시 이후 증가하여 밤 12시 이후 급감하는 추세를 보이는 것으로 나타났다. 요일별로 월요일, 목요일, 화요일, 수요일, 일요일, 금요일, 토요일 순으로 나타났다([그림 3-1], [그림 3-2]).

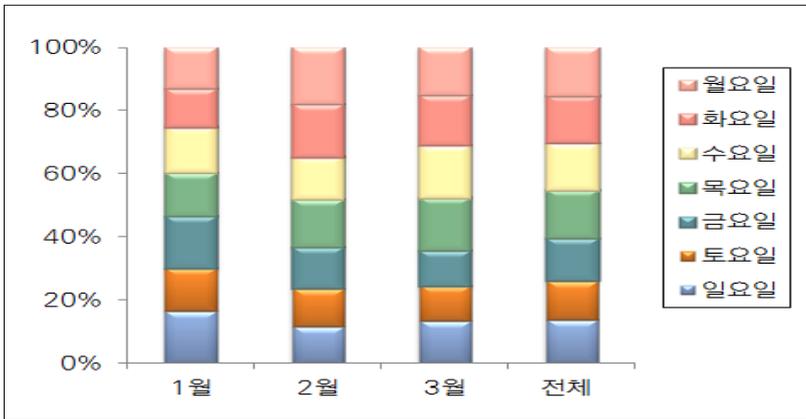
[그림 3-1] 월별, 시간별 담배 관련 온라인 문서 현황

(단위: %)



[그림 3-2] 담배 관련 요일별 온라인 문서 현황

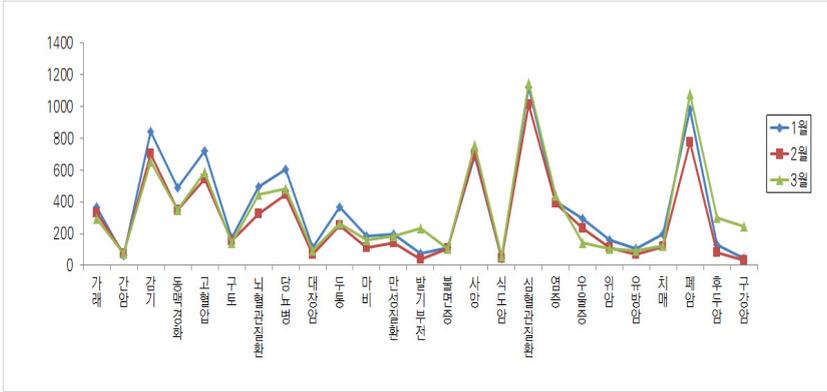
(단위: %)



[그림 3-3]과 같이 담배와 관련한 질환의 문서는 심혈관질환(3278건), 폐암(2835건), 감기(2201건), 사망(2144건), 고혈압(1849건), 당뇨병(1537건) 등의 순으로 집중된 것으로 나타났다.

[그림 3-3] 월별 담배 관련 질환별 온라인 문서 현황

(단위: 건)



<표 3-1> 담배 관련 온라인 문서 현황

(단위: 건, %)

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) |
|--------|---------------|---------------|--------|----------|----------------|
| 감정 | 일반 | 17,237(54.9) | 채널 | 블로그 | 22,794(12.6) |
| | 잠재 | 3,075(9.8) | | 카페 | 53,052(29.2) |
| | 위험 | 11,070(35.3) | | SNS | 78,369(43.2) |
| | 소계 | 31,382(100.0) | | 게시판 | 17,958(9.9) |
| 정책* | 담뱃값 인상 | 10,018(53.9) | | 뉴스 | 9,222(5.1) |
| | FCTC | 128(0.7) | | 소계 | 181,395(100.0) |
| | 금연관리법 | 1,908(10.3) | 금연 도구* | 금연검 | 850(8.1) |
| | 흡연규제 | 3,146(16.9) | | 금연약 | 772(7.3) |
| | 금연광고 | 1,112(6.0) | | 전자담배 | 7,343(69.8) |
| | 금연사업 | 2,280(12.3) | | 금연패치 | 971(9.2) |
| 소계 | 18,592(100.0) | 금연보조제 | | 578(5.5) | |
| 도움 치료* | 금연클리닉 | 2,163(25.0) | | 장소* | 소계 |
| | 금연상담전화 | 40(0.5) | PC방 | | 299(1.0) |
| | 병원 | 6,280(72.5) | 가정 | | 2,270(7.5) |
| | 금연교실 | 174(2.0) | 소계 | | 289(1.0) |
| 폐해* | 소계 | 8,657(100.0) | 아파트 | | 3,117(10.3) |
| | 간접흡연 | 15,894(48.9) | 공공장소 | | 4,911(16.2) |
| | 알코올 | 2,883(8.9) | 흡연구역 | | 2,156(7.1) |
| | 중독 | 2,017(6.2) | 직장 | | 3,389(11.2) |
| | 기억력 | 2,719(8.4) | 술집 | | 2,982(9.8) |

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) |
|-------|-------|---------------|--------|--------------|---------------|
| | 담배꽂초 | 4,949(15.2) | | 식당 | 5,078(16.7) |
| | 도박·마약 | 1,488(4.6) | | 학교 | 5,835(19.2) |
| | 이혼 | 719(2.2) | | 소계 | 30,326(100.0) |
| | 정신건강 | 699(2.2) | 관련 기관* | 청와대 | 4,252(47.6) |
| | 폭력 | 1,120(3.4) | | 국회 | 567(6.4) |
| | 소계 | 32,488(100.0) | | 보건복지부 | 979(11.0) |
| 유해 물질 | 니코틴 | 3,396(46.6) | | 여성가족부 | 113(1.3) |
| | 발암물질 | 727(10.0) | | 기획재정부 | 517(5.8) |
| | 유해물질 | 539(7.4) | | 지방자치단체 | 1,040(11.6) |
| | 일산화탄소 | 413(5.7) | 공공기관 | 507(5.7) | |
| | 타르 | 1,368(18.8) | 세계보건기구 | 480(5.4) | |
| | 화학물질 | 350(4.8) | 금연단체 | 164(1.8) | |
| | 노폐물 | 501(6.9) | 담배회사 | 310(3.5) | |
| | 소계 | 7,294(100.0) | 소계 | 8,929(100.0) | |

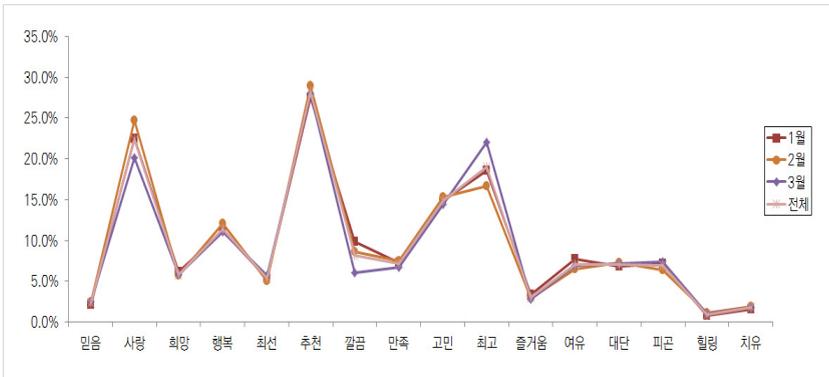
주: *다중 응답.

〈표 3-1〉과 같이 담배를 애호적으로 생각하는 위험군을 나타내는 버즈는 35.3%, 보통의 감정(잡재군)을 나타내는 버즈는 9.8%, 혐오적으로 생각하는 일반군을 나타내는 버즈는 54.9%로 나타났다. 담배와 관련한 정책은 담뱃값 인상(53.9%), 흡연규제(16.9%), 금연사업(12.3%) 등의 순으로 나타났다. 담배와 관련한 도움·치료는 병원(72.5%), 금연클리닉(25.0%), 금연교실(2.0%), 금연상담전화(0.5%) 순으로 나타났다. 담배와 관련한 폐해로는 간접흡연(48.9%), 담배꽂초(15.2%), 알코올(8.9%), 기억력(8.4%) 등의 순으로 나타났다. 담배와 관련한 유해물질로는 니코틴(46.6%), 타르(18.8%), 발암물질(10.0%), 유해물질(7.4%) 등의 순으로 나타났다. 담배와 관련한 채널로는 SNS(43.2%), 카페(29.2%), 블로그(12.6%) 등의 순으로 나타났다. 담배와 관련한 금연 도구로는 전자담배(69.8%), 금연패치(9.2%), 금연보조제(5.5%) 등의 순으로 나타났다. 담배와 관련한 장소로는 학교(19.2%), 식당(16.7%), 공공장소(16.2%) 등의 순으로 나타났다. 담배와 관련한 기관으로는 청와대(47.6%), 지방자

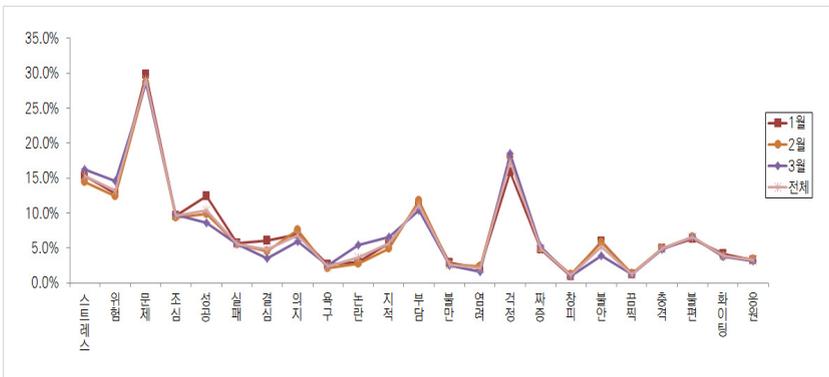
치단체(11.6%), 보건복지부(11.0%), 국회(6.4%) 등의 순으로 나타났다.

[그림 3-4], [그림 3-5]와 같이 담배를 애호적으로 생각하는 위험 감정의 토픽은 추천(28.1%), 사랑(22.4%), 최고(19.1%), 고민(14.9%), 행복(11.4%) 등의 순으로 집중된 것으로 나타났으며, 담배를 혐오적으로 생각하는 일반 감정의 토픽은 문제(29.1%), 걱정(17.4%), 스트레스(15.3%), 위험(13.2%), 부담(11.1%) 등의 순으로 집중된 것으로 나타났다.

[그림 3-4] 월별 담배 위험군(담배를 애호적으로 생각하는 감정 집단) 감정 변화 (단위: %)



[그림 3-5] 월별 담배 일반군(담배를 혐오적으로 생각하는 감정 집단) 감정 변화 (단위: %)



〈표 3-2〉와 같이 담배와 관련한 연도별 긍정적 감정(위험)을 나타내는 버즈는 청소년의 경우 2011년 30.9%, 2012년 29.2%, 2013년 27.9%, 2014년 26.8%, 2015년 25.3%, 2016년 30.0%로 나타났으며 성인의 경우 2011년 39.9%, 2012년 36.4%, 2013년 34.0%, 2014년 36.2%, 2015년 29.1%, 2016년 37.1%로 나타났다. 2015년 담뱃값 인상은 청소년보다 성인에게 더 많은 영향을 미친 것으로 나타났고, 2016년은 2015년보다 위험을 나타내는 버즈가 증가한 것으로 나타났다.

〈표 3-2〉 연도별 담배 관련 감정 변화

(단위: 건, %)

| 연도 | 청소년 | | | | 성인 | | | | 전체 | | | |
|------|------------------|-----------------|------------------|--------|------------------|-----------------|------------------|---------|------------------|-----------------|-------------------|---------|
| | 위험 | 잠재 | 일반 | 계 | 위험 | 잠재 | 일반 | 계 | 위험 | 잠재 | 일반 | 계 |
| 2011 | 2,283 (30.9) | 1,017 (13.8) | 4,085 (55.3) | 7,385 | 6,873 (39.9) | 1,747 (10.2) | 8,589 (49.9) | 17,209 | 9,156 (37.2) | 2,764 (11.2) | 12,674 (51.5) | 24,594 |
| 2012 | 2,767 (29.2) | 1,174 (12.4) | 5,542 (58.4) | 9,483 | 10,537 (36.4) | 1,938 (6.7) | 16,508 (57.0) | 28,983 | 13,304 (34.6) | 3,112 (8.1) | 22,050 (57.3) | 38,466 |
| 2013 | 2,478 (27.9) | 1,178 (13.3) | 5,228 (58.8) | 8,884 | 11,037 (34.0) | 1,956 (6.0) | 19,432 (59.9) | 32,425 | 13,515 (32.7) | 3,134 (7.6) | 24,660 (59.7) | 41,309 |
| 2014 | 2,145 (26.8) | 1,050 (13.1) | 4,799 (60.0) | 7,994 | 9,106 (36.2) | 1,906 (7.6) | 14,125 (56.2) | 25,137 | 11,251 (34.0) | 2,956 (8.9) | 18,924 (57.1) | 33,131 |
| 2015 | 2,160 (25.3) | 1,068 (12.5) | 5,324 (62.3) | 8,552 | 12,274 (29.1) | 3,172 (7.5) | 26,769 (63.4) | 42,215 | 14,434 (28.4) | 4,240 (8.4) | 32,093 (63.2) | 50,767 |
| 2016 | 2,433 (30.0) | 1,157 (14.2) | 4,533 (55.8) | 8,123 | 8,637 (37.1) | 1,918 (8.2) | 12,704 (54.6) | 23,259 | 11,070 (35.3) | 3,075 (9.8) | 17,237 (54.9) | 31,382 |
| 계 | 14,266 (28.3) | 6,644 (13.2) | 29,511 (58.5) | 50,421 | 58,464 (34.5) | 12,637 (7.5) | 98,127 (58.0) | 169,228 | 72,730 (33.1) | 19,281 (8.8) | 127,638 (58.1) | 219,649 |

자료: 2011~2015년 결과는 '송태민, 진달래(2015). 2015년 소셜 빅데이터 기반 보건복지 이슈 동향 분석. 한국보건사회연구원. p.77' 재인용.

2. 담배 위험 관련 연관성 분석

〈표 3-3〉과 이 정책 요인에 대한 담배 위험 연관성 예측에서 가장 신뢰도가 높은 연관규칙은 {담뱃값 인상, 흡연구제}⇒{일반}이며 세 변인의 연관성은 지지도 0.001, 신뢰도는 0.590, 향상도는 6.21로 나타났다. 이는 온라인 문서에서 담뱃값 인상, 흡연구제가 언급되면 담배를 부정적(일반)

으로 생각할 확률이 59.0%이며 담뱃값 인상, 흡연규제가 언급되지 않은 문서보다 담배에 대한 감정이 일반일 확률이 6.21배 높아지는 것을 나타낸다. 특히 {담뱃값 인상}⇒{위험} 두 변인의 연관성은 지지도 0.002, 신뢰도는 0.029, 향상도는 0.469로 나타나 담뱃값 인상은 담배에 대한 긍정적(위험) 감정을 감소시키는 것으로 나타났다. 반면 {금연사업}⇒{잠재} 두 변인의 연관성은 지지도 0.001, 신뢰도는 0.103, 향상도는 6.054로 나타나 담뱃값 인상은 담배에 대한 보통(잠재) 감정을 증가시키는 것으로 나타났다.

〈표 3-3〉 요인별 담배 위험 예측

| 규칙 | 지지도 | 신뢰도 | 향상도 |
|---------------------|----------|----------|----------|
| 정책 | | | |
| {담뱃값 인상, 흡연규제}⇒{일반} | 0.001246 | 0.590078 | 6.209738 |
| {금연관련법, 흡연규제}⇒{일반} | 0.001191 | 0.487585 | 5.131138 |
| {금연사업}⇒{일반} | 0.005298 | 0.421491 | 4.435598 |
| {금연관련법}⇒{일반} | 0.003572 | 0.339623 | 3.574047 |
| {금연광고}⇒{일반} | 0.001869 | 0.304856 | 3.208179 |
| {흡연규제}⇒{일반} | 0.005143 | 0.296567 | 3.120948 |
| {금연사업}⇒{위험} | 0.002078 | 0.165351 | 2.709469 |
| {담뱃값 인상}⇒{일반} | 0.007183 | 0.130066 | 1.368759 |
| {금연사업}⇒{잠재} | 0.001290 | 0.102632 | 6.054262 |
| {흡연규제}⇒{위험} | 0.001152 | 0.066434 | 1.088592 |
| {담뱃값 인상}⇒{위험} | 0.001582 | 0.028648 | 0.469438 |
| 유해물질 | | | |
| {노페롤}⇒{일반} | 0.001582 | 0.572854 | 6.028480 |
| {일산화탄소}⇒{일반} | 0.001296 | 0.569007 | 5.987995 |
| {화학물질}⇒{일반} | 0.001053 | 0.545714 | 5.742870 |
| {발암물질}⇒{일반} | 0.002051 | 0.511692 | 5.384832 |
| {유해물질}⇒{일반} | 0.001279 | 0.430427 | 4.529631 |
| {니코틴}⇒{일반} | 0.005502 | 0.293875 | 3.092620 |
| {타르}⇒{일반} | 0.001356 | 0.179825 | 1.892399 |
| {니코틴, N타르}⇒{일반} | 0.001136 | 0.175319 | 1.844986 |
| {니코틴}⇒{위험} | 0.002056 | 0.109835 | 1.799778 |
| {니코틴}⇒{잠재} | 0.001163 | 0.062132 | 3.665177 |
| 질환 | | | |
| {고혈압, 당뇨병, 사망}⇒{일반} | 0.001455 | 0.939502 | 9.886925 |

82 2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석

| 규칙 | 지지도 | 신뢰도 | 향상도 |
|----------------------------|----------|----------|----------|
| {고혈압, 뇌혈관질환, 당뇨병, 사망}⇒{일반} | 0.001036 | 0.935323 | 9.842953 |
| {뇌혈관질환, 당뇨병, 사망}⇒{일반} | 0.001103 | 0.925926 | 9.744058 |
| {고혈압, 당뇨병, 사망, 심혈관질환}⇒{일반} | 0.001020 | 0.925000 | 9.734314 |
| {고혈압, 뇌혈관질환, 당뇨병, 두통}⇒{일반} | 0.001169 | 0.917749 | 9.658007 |
| {고혈압, 뇌혈관질환, 사망}⇒{일반} | 0.001152 | 0.916667 | 9.646618 |
| {뇌혈관질환, 당뇨병, 두통}⇒{일반} | 0.001196 | 0.915612 | 9.635517 |
| {당뇨병, 사망, 심혈관질환}⇒{일반} | 0.001196 | 0.896694 | 9.436436 |
| {당뇨병, 사망}⇒{일반} | 0.001737 | 0.889831 | 9.364205 |
| {고혈압, 뇌혈관질환, 두통}⇒{일반} | 0.001235 | 0.881890 | 9.280640 |
| {고혈압, 구토, 당뇨병}⇒{일반} | 0.001003 | 0.866667 | 9.120439 |
| {뇌혈관질환, 두통}⇒{일반} | 0.001345 | 0.859155 | 9.041388 |
| {고혈압, 구토}⇒{일반} | 0.001097 | 0.846809 | 8.911460 |
| {구토, 당뇨병}⇒{일반} | 0.001092 | 0.838983 | 8.829108 |
| {고혈압, 사망}⇒{일반} | 0.001814 | 0.837150 | 8.809819 |
| {고혈압, 사망, 심혈관질환}⇒{일반} | 0.001307 | 0.814433 | 8.570753 |
| {동맥경화, 사망, 심혈관질환}⇒{일반} | 0.001047 | 0.791667 | 8.331170 |
| {동맥경화, 고혈압, 사망}⇒{일반} | 0.001003 | 0.787879 | 8.291308 |
| {구토, 두통}⇒{일반} | 0.001218 | 0.786477 | 8.276555 |
| 금연 도구 | | | |
| {금연약}⇒{일반} | 0.002398 | 0.563472 | 5.929739 |
| {금연패치}⇒{일반} | 0.002547 | 0.475798 | 5.007101 |
| {금연껌}⇒{일반} | 0.001632 | 0.348235 | 3.664683 |
| {전자담배}⇒{위험} | 0.005518 | 0.136320 | 2.233769 |
| {전자담배}⇒{일반} | 0.004956 | 0.122430 | 1.288397 |
| {전자담배}⇒{잠재} | 0.001703 | 0.042081 | 2.482362 |
| 치료 | | | |
| {금연클리닉, 병원}⇒{일반} | 0.002271 | 0.615845 | 6.480891 |
| {병원}⇒{일반} | 0.014868 | 0.429459 | 4.519443 |
| {금연클리닉}⇒{일반} | 0.005033 | 0.422099 | 4.441993 |
| {금연클리닉}⇒{위험} | 0.002029 | 0.170134 | 2.787847 |
| {병원}⇒{위험} | 0.004868 | 0.140605 | 2.303980 |
| {병원}⇒{잠재} | 0.003831 | 0.110669 | 6.528379 |
| {금연클리닉}⇒{잠재} | 0.001284 | 0.107721 | 6.354474 |
| 폐해 | | | |
| {알코올, 기억력}⇒{일반} | 0.002624 | 0.607918 | 6.397478 |
| {기억력, 담배꽂초}⇒{일반} | 0.001483 | 0.580994 | 6.114134 |
| {기억력}⇒{일반} | 0.008203 | 0.547260 | 5.759136 |
| {간접흡연, 중독}⇒{일반} | 0.001472 | 0.531873 | 5.597204 |
| {간접흡연, 기억력}⇒{일반} | 0.002067 | 0.525210 | 5.527092 |
| {간접흡연, 담배꽂초}⇒{일반} | 0.003815 | 0.437698 | 4.606148 |
| {정신건강}⇒{일반} | 0.001626 | 0.422031 | 4.441283 |
| {간접흡연, 알코올}⇒{일반} | 0.001020 | 0.404814 | 4.260094 |
| {중독}⇒{일반} | 0.004427 | 0.398116 | 4.189607 |

| 규칙 | 지지도 | 신뢰도 | 향상도 |
|-------------------|----------|----------|----------|
| {알코올}⇒{일반} | 0.006218 | 0.391259 | 4.117448 |
| {이혼}⇒{일반} | 0.001417 | 0.357441 | 3.761559 |
| {담배꽂초}⇒{일반} | 0.008953 | 0.328147 | 3.453283 |
| {폭력}⇒{일반} | 0.001907 | 0.308929 | 3.251035 |
| {도박마약}⇒{일반} | 0.002012 | 0.245296 | 2.581390 |
| {간접흡연}⇒{일반} | 0.017840 | 0.203599 | 2.142589 |
| {기억력}⇒{위험} | 0.002415 | 0.161089 | 2.639627 |
| {간접흡연, 담배꽂초}⇒{위험} | 0.001191 | 0.136622 | 2.238719 |
| {기억력}⇒{잠재} | 0.002018 | 0.134608 | 7.940577 |
| {알코올}⇒{잠재} | 0.002040 | 0.128339 | 7.570722 |
| {담배꽂초}⇒{위험} | 0.003457 | 0.126692 | 2.076002 |
| {중독}⇒{위험} | 0.001351 | 0.121468 | 1.990389 |
| {간접흡연, 담배꽂초}⇒{잠재} | 0.001058 | 0.121442 | 7.163901 |
| {알코올}⇒{위험} | 0.001703 | 0.107180 | 1.756271 |
| {중독}⇒{잠재} | 0.001147 | 0.103123 | 6.083278 |
| {간접흡연}⇒{위험} | 0.008539 | 0.097458 | 1.596967 |
| {담배꽂초}⇒{잠재} | 0.002089 | 0.076581 | 4.517539 |
| {간접흡연}⇒{잠재} | 0.004449 | 0.050774 | 2.995163 |

3. 담배에 영향을 미치는 정책 및 도구 요인

〈표 3-4〉와 같이 금연과 관련한 정책 중 FCTC를 제외한 모든 정책 요인은 담배의 위험에 부적인 영향을 미치는 것으로 나타나 담뱃값 인상, 금연관련법, 흡연규제, 금연광고가 온라인상에서 많이 언급될수록 담배에 대한 애호적인 감정(위험)이 감소하는 것으로 나타났다. 금연 도구 요인의 영향과 관련해서는 담배 위험군에 금연약, 금연패치는 부적인 영향을 미치는 것으로 나타나 금연약·금연패치와 관련한 금연 도구가 온라인상에서 많이 언급될수록 담배에 대한 애호적인 감정(위험)은 감소하는 것으로 나타났다. 그러나 전자담배는 정적인 영향을 미치는 것으로 나타나 전자담배와 관련한 금연 도구가 많이 언급될수록 담배에 대한 긍정적 감정(위험)은 증가하는 것으로 나타났다.

〈표 3-4〉 담배의 위험군과 잠재군에 영향을 미치는 정책 및 도구 요인¹⁾

| 변수 | 위험 | | | | 잠재 | | | | |
|----|--------|--------|-----------------|-------|----------------|--------|-----------------|-------|------|
| | b | S.E. † | OR [§] | P | b [†] | S.E. † | OR [§] | P | |
| 정책 | 담뱃값 인상 | -.988 | .067 | .372 | .000 | -.397 | .089 | .672 | .000 |
| | FCTC | .496 | .326 | 1.642 | .128 | -.676 | .602 | .509 | .262 |
| | 금연관련법 | -1.076 | .109 | .341 | .000 | -.230 | .122 | .794 | .059 |
| | 흡연규제 | -.774 | .080 | .461 | .000 | -.184 | .101 | .832 | .067 |
| | 금연광고 | -.720 | .129 | .487 | .000 | -.215 | .159 | .807 | .177 |
| | 금연사업 | -.422 | .063 | .656 | .000 | .370 | .076 | 1.447 | .000 |
| 도구 | 금연검 | -.129 | .111 | .879 | .248 | .700 | .123 | 2.013 | .000 |
| | 금연약 | -1.155 | .124 | .315 | .000 | -.078 | .130 | .925 | .550 |
| | 전자담배 | .638 | .048 | 1.892 | .000 | .707 | .070 | 2.027 | .000 |
| | 금연패치 | -1.154 | .117 | .315 | .000 | -.169 | .124 | .844 | .173 |
| | 금연보조제 | -.091 | .170 | .913 | .592 | -.135 | .257 | .874 | .600 |

주: 1) 기본범주-일반, † Standardized coefficients, ‡ Standard error, § Adjusted odds ratio.

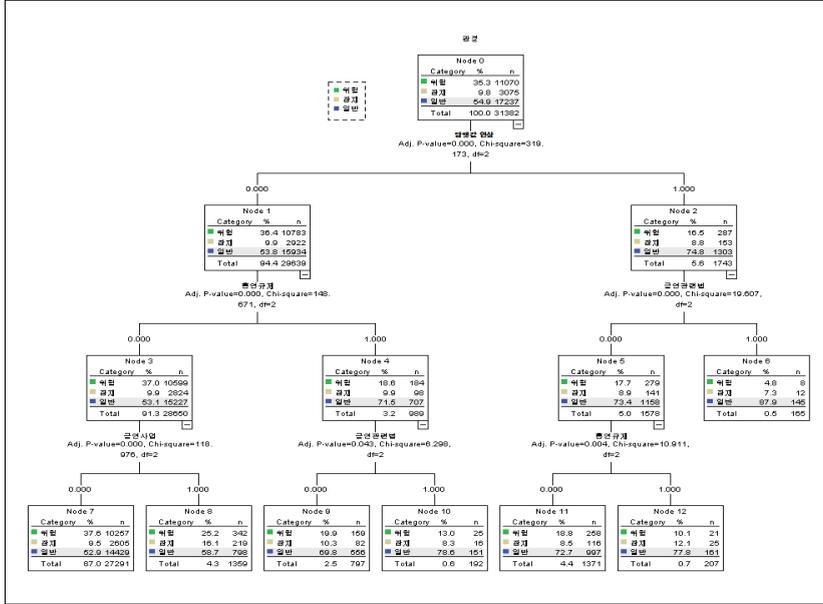
4. 담배 관련 위험 예측 모형

본 연구에서는 담배 관련 위험을 예측하기 위하여 담배와 관련한 정책 요인에 대해 데이터마이닝 분석을 실시하였다. 담배 관련 정책 요인이 담배의 위험 예측 모형에 미치는 영향은 [그림 3-6]과 같다.

나무 구조의 최상위에 있는 네모는 루트노드로서, 예측변수(독립변수)가 투입되지 않은 종속변수(위험, 잠재, 일반)의 빈도를 나타낸다. 루트노드에서 담배의 위험은 35.3%(1만 1070건), 잠재는 9.8%(3075건), 일반은 54.9%(1만 7237건)로 나타났다. 루트노드 하단의 가장 상위에 위치하는 요인은 담배의 위험 예측에 가장 영향력이 높은(관련성이 깊은) 정책 요인으로 ‘담뱃값 인상 요인’의 영향력이 가장 큰 것으로 나타났다. ‘담뱃값 인상 요인’이 있을 경우 담배의 위험은 이전의 35.3%에서 16.5%로 크게 감소하였고 잠재는 이전의 9.8%에서 8.8%로 1.0% 포인트 감소

하였다. 반면에 일반은 이전의 54.9%에서 74.8%로 증가하였다. ‘담뱃값 인상 요인’이 있고 ‘금연관련법 요인’이 있는 경우 담배의 위험은 이전의 16.5%에서 4.8%로, 잠재는 이전의 8.8%에서 7.3%로 감소한 반면 일반은 이전의 74.8%에서 87.9%로 증가하였다. <표 3-5>의 담배와 관련한 정책 요인의 위험 예측 모형에 대한 이익도표와 같이 담배의 위험에 가장 영향력이 높은 경우는 ‘담뱃값 인상 요인’이 없고 ‘흡연규제 요인’이 없으며 ‘금연사업 요인’이 없는 조합으로 나타났다. 즉 7번 노드의 지수(index)가 106.5%로, 뿌리마디와 비교했을 때 7번 노드의 조건을 가진 집단이 담배에 대한 위험이 높을 확률이 1.07배로 나타났다. 담배의 잠재에 가장 영향력이 높은 경우는 ‘담뱃값 인상 요인’이 없고 ‘흡연규제 요인’이 없으며 ‘금연사업 요인’이 있는 조합으로 나타났다. 즉 8번 노드의 지수가 164.5%로, 뿌리마디와 비교했을 때 8번 노드의 조건을 가진 집단이 담배에 대한 보통의 감정(잠재)이 높을 확률이 1.65배로 나타났다. 담배의 일반에 가장 영향력이 높은 경우는 ‘담뱃값 인상 요인’이 있고 ‘금연관련법 요인’이 있는 조합으로 나타났다. 즉 6번 노드의 지수(index)가 160.0%로 뿌리마디와 비교했을 때 6번 노드의 조건을 가진 집단이 담배에 대한 부정의 감정(일반)이 높을 확률이 1.60배로 나타났다.

[그림 3-6] 정책 요인의 예측 모형



<표 3-5> 정책 요인의 예측 모형에 대한 이익도표

| 구분 | 노드 | 이익지수 | | | | 누적지수 | | | |
|----|----|-------|-------|-------|-------|-------|-------|-------|-------|
| | | 노드(n) | 노드(%) | 이익(%) | 지수(%) | 노드(n) | 노드(%) | 이익(%) | 지수(%) |
| 위험 | 7 | 27291 | 87.0 | 92.7 | 106.5 | 27291 | 87.0 | 92.7 | 106.5 |
| | 8 | 1359 | 4.3 | 3.1 | 71.3 | 28650 | 91.3 | 95.7 | 104.9 |
| | 9 | 797 | 2.5 | 1.4 | 56.6 | 29447 | 93.8 | 97.2 | 103.6 |
| | 11 | 1371 | 4.4 | 2.3 | 53.3 | 30818 | 98.2 | 99.5 | 101.3 |
| | 10 | 192 | 0.6 | 0.2 | 36.9 | 31010 | 98.8 | 99.7 | 100.9 |
| | 12 | 207 | 0.7 | 0.2 | 28.8 | 31217 | 99.5 | 99.9 | 100.5 |
| 잠재 | 6 | 165 | 0.5 | 0.1 | 13.7 | 31382 | 100.0 | 100.0 | 100.0 |
| | 8 | 1359 | 4.3 | 7.1 | 164.5 | 1359 | 4.3 | 7.1 | 164.5 |
| | 12 | 207 | 0.7 | 0.8 | 123.3 | 1566 | 5.0 | 7.9 | 159.0 |
| | 9 | 797 | 2.5 | 2.7 | 105.0 | 2363 | 7.5 | 10.6 | 140.8 |
| | 7 | 27291 | 87.0 | 84.7 | 97.4 | 29654 | 94.5 | 95.3 | 100.9 |
| | 11 | 1371 | 4.4 | 3.8 | 86.3 | 31025 | 98.9 | 99.1 | 100.2 |
| 일반 | 10 | 192 | 0.6 | 0.5 | 85.0 | 31217 | 99.5 | 99.6 | 100.1 |
| | 6 | 165 | 0.5 | 0.4 | 74.2 | 31382 | 100.0 | 100.0 | 100.0 |
| | 10 | 192 | 0.6 | 0.9 | 143.2 | 357 | 1.1 | 1.7 | 151.0 |

| 구분 | 노드 | 이익지수 | | | | 누적지수 | | | |
|----|----|-------|-------|-------|-------|-------|-------|-------|-------|
| | | 노드(n) | 노드(%) | 이익(%) | 지수(%) | 노드(n) | 노드(%) | 이익(%) | 지수(%) |
| | 12 | 207 | 0.7 | 0.9 | 141.6 | 564 | 1.8 | 2.7 | 147.5 |
| | 11 | 1371 | 4.4 | 5.8 | 132.4 | 1935 | 6.2 | 8.4 | 136.8 |
| | 9 | 797 | 2.5 | 3.2 | 127.0 | 2732 | 8.7 | 11.7 | 133.9 |
| | 8 | 1359 | 4.3 | 4.6 | 106.9 | 4091 | 13.0 | 16.3 | 125.0 |
| | 7 | 27291 | 87.0 | 83.7 | 96.3 | 31382 | 100.0 | 100.0 | 100.0 |

제4절 결론

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, SNS, 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 주제분석과 감성분석 기술로 분류하고 데이터마이닝의 연관성분석과 의사결정나무분석 방법을 적용하여 분석함으로써 우리나라 국민의 담배에 대한 위험요인을 예측하고자 하였다. 본 연구의 주요 분석 결과는 다음과 같다.

첫째, 담배 관련 버즈는 월별로 오전 7시부터 증가하여 오전 11시 이후 감소하고 오후 7시 이후 증가하여 밤 12시 이후 감소하는 것으로 나타났다. 요일별로 월요일, 목요일, 화요일 등의 순으로 나타났다.

둘째, 2015년 담뱃값 인상은 청소년보다 성인에게 더 많은 영향을 미친 것으로 나타났고, 2016년은 2015년보다 담배의 위험이 증가한 것으로 나타났다. 담뱃값 인상의 금연 정책이 시행된 2015년 담배를 애호적으로 생각하는 감정(위험군)을 나타내는 버즈는 2014년에 비해 5.6% 포인트 감소한 28.4%로 나타났다. 하지만 2016년 위험군 버즈량은 2014년보다 1.3% 포인트 증가한 35.3%로 나타났다. 연도별 1/4분기 담배 판매량 변화를 살펴보면 2014년 122억 개비에서 2015년 71억 개비로 2014년 대비 41.4% 감소했지만 2016년은 105억 개비로 2015년에 비해 47.9% 증가한 것으로 나타났다(KT&G, 2016. 9. 5.). 마찬가지로 최

병호와 이근재(2015)의 연구에서 담배 가격 인상 충격에 대한 1인당 담배 소비량의 단기 변동 과정을 살펴보면 가격 인상 충격에 따라 담배 소비량은 즉각적으로 상당히 하락하지만 대체로 5개월 정도 지나면 가격 충격이 소비량에 미치는 효과가 사라지는 것으로 나타났다. 이는 담배가 중독성이 있고, 주기적으로 피울 수 있으며, 손쉽게 살 수 있기 때문에 담뱃값 인상의 금연 정책 효과의 지속성이 오래가지 않은 것으로 보인다.

셋째, 버즈에서 담뱃값 인상, 흡연규제 등이 동시에 언급되면 일반군이 될 확률이 증가하며, 담뱃값 인상만 언급되어도 위험군을 감소시키는 것으로 나타났다.

넷째, 전자담배의 언급은 일반군보다 위험군의 확률이 높은 것으로 나타났다. 2014년 담뱃값 인상 발표 이후 전자담배 판매량이 급증하였고, 2015년 1월에서 4월 전자담배기기 수입 금액은 전년 대비 24배, 전자담배 액상은 15배 이상 급증하였다(한국소비자원, 2015). 전자담배는 흡입 정도에 따라 니코틴 흡입량이 달라지기 때문에 흡연자 스스로 유입량을 예측하기가 어렵고, 안전성에 대한 의혹도 제기되며, WHO는 전자담배의 안전성과 금연 효과가 과학적으로 증명된 것은 아니기 때문에 단순 금연보조제로 사용해서는 안 된다고 권고하고 있다(최성은, 2015; WHO, 2014).

보건복지부는 가격 인상 정책 이후로 담뱃갑 경고 그림과 경고 문구 삽입, 금연치료서비스, 1:1 맞춤형 금연상담서비스, 담배 광고 및 판촉에 대한 규제 강화, 전자담배 사용 규제 강화, 소포장 금지 및 가향 첨가 규제, 금연지원서비스 및 금연 캠페인 강화 등의 비가격 정책을 시행해 왔다(기획재정부, 보건복지부 보도자료, 2016. 5. 10.). 또한 지난해 말 흡연의 폐해를 몸으로 표현하는 ‘발레 광고’와 흡연은 스스로 질병을 구입한다는 콘셉트의 ‘폐암 하나 주세요’, 실제 흡연의 후유증으로 고통받는 환자와

가족의 모습을 보인 광고는 TV뿐만 아니라 SNS 등에서 큰 이슈가 되었다. 그러나 본 연구에서는 연도별 1/4분기의 버즈를 분석했기 때문에 각 연도 전체를 대표하기에는 제한점이 따른다.

끝으로 금연 정책의 효과에 대한 대국민 조사와 더불어 소셜미디어에서 수집된 빅데이터의 활용과 분석을 병행할 경우 정부의 금연 정책에 대한 예측 및 평가의 신뢰성이 더욱 제고될 것으로 예상된다. 또한 국민들이 금연에 적극적으로 동참할 수 있도록 소셜 빅데이터 분석을 통해 담배를 애호적으로 생각하는 위험군을 감소시킬 수 있는 SNS 홍보가 강화되어야 할 것이다.



참고문헌 <<

- 강은정, 이재희. (2011). 성인 흡연자의 금연의향 담배가격에 영향을 미치는 요인. 보건교육 건강증진학회지, 28(2), pp.125-137.
- 기획재정부, 보건복지부 보도자료. (2016. 5. 10.). 비가격 금연정책 추진방안. 보건복지부 보도자료.
- 박희창. (2010). 연관규칙 마이닝에서의 평가기준 표준화 방안. 한국데이터정보과학회지, 21(5), pp.891-899.
- 보건복지부, 질병관리본부. (2015). 국민건강통계 2014: 국민건강영양조사 VI. 보건복지부, 질병관리본부.
- 보건복지부 보도자료. (2014. 9. 11.). 범정부, 「금연 종합대책」 발표. 보건복지부 보도자료.
- 보건복지부, 한국건강증진개발원. (2016). 2016년 지역사회 통합건강증진사업 안내: 지역사회 중심 금연지원 서비스. 보건복지부, 한국건강증진개발원.
- 이선미, 윤영덕, 백종환, 현경래, 강하림. (2015). 주요건강위험요인의 사회경제적 영향과 규제정책의 효과 평가. 국민건강보험 건강보험정책연구원.
- 정금지, 윤영덕, 백수진, 지선하, 김일순. (2013). 한국인 성인 남녀의 흡연관련 사망에 관한 연구. 한국보건정보통계학회지, 38(2), pp.36-48.
- 최병호, 이근재. (2015). 우리나라 담배수요함수의 추정과 담뱃세 정책에 관한 함의. 경제학회지, 63(4), pp. 53-80.
- 최성은. (2015). 전자담배 관리방안 및 전자담배 제세부담금 개편방안연구. 한국조세재정연구원.
- 한국소비자원. (2015). 전자담배 안전실태조사. 한국소비자원.
- KT&G. (2016. 9. 5.). 경영실적 2014년 1분기-2016년 1분기. KT&G IR자료실.
- Brewer, N. T., Hall, M. G., Noar, S. M., Parada, H., Stein-Seroussi, A., Bach, L. E., & Hanley, S., et al. (2016). Effect of Pictorial Cigarette pack warnings on changes in Smoking Behavior. *JAMA Intern Med*, 176(7), pp.905-912.

- Campaign for Tobacco-Free Kids. (2013). *Increasing the federal tobacco tax reduces tobacco use*, Washington DC.
- Carter, B. D., Abnet, C. C., Feskanich, D., Freedman, N. D., Hartge, P., Lewis, C. E., Ockene, J. K., et al. (2015). Smoking and mortality: Beyond established causes. *New England Journal of Medicine*, 372, pp.631-640.
- Centers for Disease Control and Prevention. (2014). Cigarette prices and smoking prevalence after a tobacco tax increase-Turkey, 2008 and 2012. *MMWR Morbidity and Mortality Weekly Report*, 63, pp.457-461.
- Hammond, D., Fong, G. T., McDonald, P. W., Brown, K. S. and Cameron, R. (2004). Graphic Canadian cigarette warning labels and adverse outcomes. *American Journal of Public Health*, 94, pp.1442-1445.
- Organization for Economic Cooperation and Development. (2014). *Health data 2014*. Paris, OECD.
- Thun, M. J., Carter, B. D., Feskanich, D., Freedman, N. D., Prentice, R., Lopez, A. D., & Hartge, P. et al. (2013). 50-year trends in smoking-related mortality in the United States. *New England Journal of Medicine*, 368, pp.351-364.
- World Health Organization. (2008). *Report on the global tobacco epidemic - The MPOWER package*. WHO, Geneva.
- World Health Organization. (2014). *Electronic nicotine delivery systems. WHO Framework Convention on Tobacco Control, Geneva*.
- World Health Organization. (2015). *WHO REPORT ON THE GLOBAL TOBACCO EPIDEMIC, 2015: Raising taxes. on tobacco*. WHO, Geneva.
- Zheng, W., McLerran, D. F., Rolland, B. A., Fu, Z., Boffetta, P., He, J.,

& Gupta, P. C., et al. (2014). Burden of total and cause-specific mortality related to tobacco smoking among adults aged ≥ 45 years in Asia: A pooled analysis of 21 cohorts. *Public Library of Science Medicine*, 11, e1001631.



제 4 장

지카바이러스 감염증 소셜
빅데이터 분류를 위한
온톨로지 개발

제1절 서론

제2절 지카바이러스 소셜 빅데이터 온라인 문서 현황

제3절 연구 결과

제4절 결론

참고문헌



4

지카바이러스 감염증 소셜 << 빅데이터 분류를 위한 온톨로지 개발²⁰⁾

제1절 서론

2015년 4월 브라질에서 시작된 지카바이러스의 감염증이 급속히 확산되어 2016년 5월 12일 현재 58개국에서 감염이 보고되었다(WHO, 2016.5.12.). 중남미 지역에서 지카바이러스 감염증에 연계된 것으로 추정되는 선천성 결함과 자가 면역질환의 급속한 증가로 인해 세계보건기구(WHO)에서는 2016년 2월 1일 ‘국제공중보건 위기’를 선언하였다(질병관리본부, 2016.2.2.; WHO, 2005).

국제 공중보건 비상사태는 국제보건규정(IHR)에 따라 질병이 다른 나라의 공중보건을 위협할 만큼 국제적으로 확산 중이며 심각성과 예외성, 예측 불가능성 때문에 국제사회의 즉각적인 공동 대응이 필요한 비상 상황을 뜻한다(WHO, 2016.8.3.). 세계보건기구는 2009년 신종플루, 2014년 소아마비 바이러스와 그해 8월에 발생한 에볼라 바이러스 확산, 이렇게 세 차례에 대해 비상사태를 선언하였다(The New York Times, 2016.2.1).

지카바이러스 증가로 인해 국내외에서는 지카바이러스에 대한 루머 또한 확산되고 있는 실정이다. 우리나라의 경우 3월 지카바이러스 감염 확진자가 발생함에 따라 소셜미디어를 통해 괴담이 급속도로 확산되었다.

20) 본 논문은 국내외 학술지에 게재하기 위하여 한국보건사회연구원(송태민, 진달래), 펜실베이니아주립대학교(송주영 교수), 인디애나주립대학교(서동철 교수)에서 공동으로 작성한 것임을 밝힌다.

소셜미디어에 떠도는 괴담으로는 “잠복기가 2년이고 감염되면 모두 소두 증에 걸린다”, “태아에게 전파되려면 태반을 통해 감염되어야 하는데 지카바이러스는 보통 혈액에서 1주일가량 존재한다”, “소변이나 키스(침)를 통해 감염된다” 등이 있다(동아일보, 2016.3.22.). 계속되는 루머 확산으로 인해 WHO, CDC에서는 지카바이러스 예방 대책 권고안과 함께 루머 확산 방지를 위한 정보를 공개하였다.

최근 지카바이러스 발생 후 소셜 분석 회사인 줌프(Zoomph)의 분석 결과에 따르면 지카바이러스 확산으로 인해 여행 계획에 차질이 생길 것을 우려하는 사람들의 반응과 함께 #WHO(World Health Organization), #Salud(스페인어로 ‘건강’을 의미함) 등의 해시태그를 연계하여 트윗을 하는 것으로 나타났다(The Straitimes, 2016.2.2.). 이뿐만 아니라 국내에서 발생한 2014년 4월 ‘세월호 사고’는 소셜네트워크서비스(SNS) 이용의 상반된 모습을 보여 주는 의미 있는 사례이다. 여객선 침몰 사고가 발생한 직후부터 온라인 희생자 추모를 비롯해 트위터, 페이스북 등의 SNS 네트워크를 활용하여 무사 귀환을 기원하는 범국민적 동참 분위기를 조성하였다. 반면 다른 한쪽에서는 각종 유언비어를 끊임없이 확대 재생산하면서 피해 가족들을 더욱 아프게 하는 것은 물론 수색과 구조 작업에 혼선을 가져왔다(이윤희, 2014).

2015년 발생한 국내 메르스(MERS) 사태에서도 근거 없는 루머가 확대 재생산되면서 예방 대응에 혼선을 빚었다. 메르스 발생 시 대부분의 한국인은 메르스에 대한 정보를 SNS를 통해 얻거나 의존하는 경향이 있었으며(Woohyun Yoo, Doo-Hun Choi, Keeho Park, 2016) 지나친 걱정과 염려로 인해 시장, 극장, 레스토랑 등 사람이 많은 공공장소에서 인파가 현저하게 줄어드는 현상까지 나타났다(Woohyun Yoo, Doo-Hun Choi, Keeho Park, 2016). 이처럼 루머는 다양한 채널을 통해 근거 없

는 소문이 확산되어 영향을 끼치는 것으로 정의한다(장덕진, 김기훈, 2011). 확산되는 루머가 자극적일수록 이용자들이 코멘트를 통해 서로 정보 교환을 할 뿐 아니라 국민들의 불안감을 야기하고, 정부 정책에 대한 불신을 초래하는 등 개인적·사회적으로 큰 손실을 끼칠 수 있다(Hong, Ju-Hyun, 2014).

초기 SNS의 등장을 단순히 새로운 비즈니스 영역 또는 소통과 정보 공유 도구로 평가하던 전문가들도 정치·사회·경제·문화 전 사회의 변화를 이끄는 SNS의 위력에 관심을 보이며 다양한 해석을 시도하고 있다. 이처럼 트위터나 페이스북 등의 SNS는 다양한 의견 교환 및 공유를 할 수 있는 곳인 반면에 루머가 생성되거나 확산되는 주요 플랫폼 중의 하나로 자리매김하고 있는 실정이다(Ling-Ling Xia, Guo-Ping Jiang, et al., 2015). 이에 따라 악의적인 루머에 대한 신속하고 적절한 대응은 매우 중요하며(Woohyun Yoo, Doo-Hun Choi, Keeho Park, 2016), 이들의 사회문화적 영향력에 대한 체계적인 연구의 필요성이 대두되고 있다(이원태, 차미영, 양해륜, 2011).

이러한 문제에도 불구하고 SNS는 전염병 발생 시 적시에 신뢰할 수 있는 정보를 제공하고 피해 확산을 감소시키는 데 큰 역할을 할 수가 있다(WHO, 2016.8.). SNS는 공중보건 문제 및 질병 위험에 대한 정확한 정보 및 정책을 홍보하는 커뮤니케이션 도구로 활용할 수 있다(WHO 2016.8.; Isaac Chun-Hai Fung, Zion Tsz Ho Tse, King-Wa Fu, 2015).

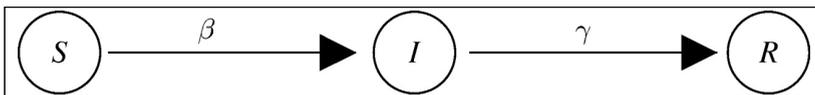
소셜 빅데이터 수집을 통해 데이터마이닝, 시각화, 통계분석, 기계학습 등의 분석으로 정확한 정보를 예측할 수 있으며(Ibrar Yaqoob, Victor Chan, et al., 2016), 질병 동향에 대한 감지, 각종 통신별 인식과 반응 등의 비상 상황을 모니터링하고 측정할 수 있다(Isaac Chun-Hai Fung, Zion Tsz Ho Tse, King-Wa Fu, 2015).

1. 감염병 질환 확산 모델

전염병은 인구 집단 내에서 일정한 전파 속도로 퍼져 나가며 전염병에 영향을 끼치는 요소들을 수학적인 모델로 만들 수 있다. 전염병의 수학적인 모델로는 특성에 따라 SIS, SIR, SEIR 등이 있으며, 이 모형들은 질병의 진행 과정에 따라 구분된다. SIS(susceptible infected susceptible) 모형은 평형 상태를 이루는 전염 확산 과정을 설명하는 데 적합한 모형이다. SIS 모형에서는 전염가능자가 감염이 되었을 때 치유 단계가 되는 것이 아니라 또다시 감염된 사람과 접촉하거나 환경에 노출될 때 감염이 가능한 상태가 되어 전염 확산에 기여하는 것이다(Juacong Chen, Huiling Zhang, et al., 2012).

감염성 바이러스 질환을 설명하는 모형으로 전염병 확산 정도를 예측하는 SIR(Susceptible-infected-recovered) 모형은 3개의 클래스로 분류될 수 있다. 전염가능자(Susceptible), 감염자(Infected), 치유자(removed) 사이에서 전염병이 어떻게 확산되는지 보여 주는 모델이다 (Sebastian Funk, Marcel Salathe, et al., 2010). SIS 모형과 SIR 모형의 다른 점은 감염 후 면역력을 갖게 되는 회복(치유) 상태가 되는 것이다.

[그림 4-1] SIR 모형



자료: Sebastian Funk, Marcel Salathe, Vincent A. A. Janesen. (2010). Modeling the influence of human behaviour on the spread of infectious disease: a review. *Journal of the Royal Society interface*, 7, pp.1247-1256.

모든 전염가능자(S), 감염자(I) 구성원 사이에서 β 만큼의 비율(rate)로 감염이 발생하며 감염자(I) 구성원은 γ 만큼의 비율로 회복(R) 상태로 변하는 것을 뜻한다. 미분 방정식으로 각각을 설명할 수 있으며(Mathematical association of America) 사람 간 접촉의 균질성, 전염병 치료에 아무런 조치를 취하지 않은 가정이 포함된다.

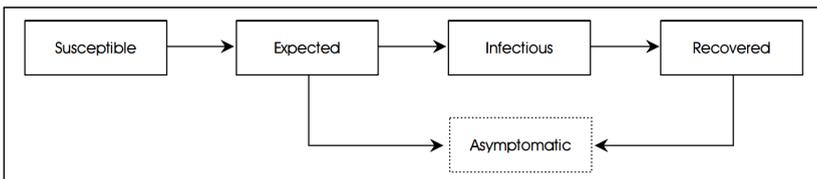
$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

SEIR(susceptible-expected-infectious-recovered) 모형은 인구 집단을 통해 전파되는 질병의 전달 과정 각 단계에서 개입되는 질병 매개체의 동적인 상태를 묘사한 모형이다(신호성, 김동진, 2008). 전염가능 단계에서 사람들이 질병에 노출된 후 일정한 시간 경과 후 감염력을 갖게 되어 전파시키고 회복되는 모형을 말한다. 질병 회복 단계에서는 면역력이 생기거나 사망에 이를 경우도 본 모형에 포함되어 있다(김성순, 이상원, 최보율, 2008).

[그림 4-2] SEIR 모형



자료: 김성순, 이상원, 최보율. (2008). 신종인플루엔자 유행관리를 위한 수학적 모형과 전략에 대한 고찰. 한국역학회지, 30(2), pp.156-167.

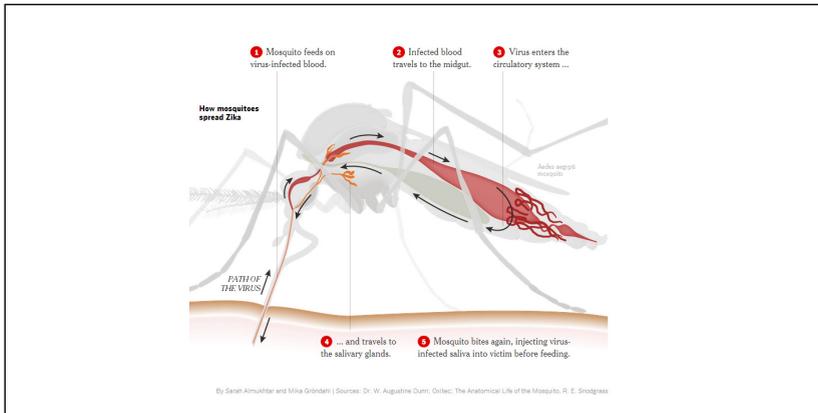
2. 지카바이러스 감염 경로 및 확산 현황

지카바이러스는 1947년 우간다 지카(Zika) 숲의 붉은털원숭이에서 발견된 바이러스이며, 1952년 우간다와 탄자니아에서 인간에게 최초로 감염된 것으로 보고되었다(Rodolphe Hamel, Florian Liègeois, et al., 2016; Centers for Disease Control and Prevention 2016). 지카바이러스는 아프리카, 동남아, 태평양 섬 지역에서 2015년 이전에 발생하였으며, 2015년 5월 이후 브라질에서 첫 보고가 된 후 확산되어 중남미 지역에서 지속적으로 환자가 발생하고 있다(질병관리본부, 2016).

지카바이러스는 Aedes 모기에 의해 감염이 되어 발생하고, 지카바이러스의 질환을 가진 사람들은 대개 발열, 피부 발진, 결막염, 근육과 관절 통증, 권태감 및 두통을 동반하는 증상을 가지고 있으며(Camila Zanluca, Claudia Nunes Duarte dos Santos, 2016), 이러한 현상은 2일에서 7일간 지속된다. 지카바이러스에 감염되면 발열, 두통, 근관절통을 동반하는 증상이 있을 뿐만 아니라 뎅기열, 발진 등의 증상도 함께 발생한다(WHO, 2016.2.; 질병관리본부, 2016). 또한 지카바이러스 감염 증가로 인해 소두증 신생아 출산 증가와 길랭바레증후군 증가 경향이 보고되고 있으나 원인에 대해서는 현재 조사가 진행 중이다. 모기에 의한 바이러스 이동 경로는 ① 흡입한 혈액을 바이러스에 감염시켜 ② 감염된 혈액이 중장(midgut)으로 이동한 다음 ③ 순환 시스템에 의해 바이러스가 순환하여 ④ 침샘으로 이동하고, 마지막으로 ⑤ 침샘을 통해 감염시키는 구조로 되어 있다(그림 4-3] 참조). 그러나 이는 사람 간의 일상적인 접촉으로는 감염되지 않는다. 다만 감염된 사람의 혈액을 수혈받거나 감염된 사람과 성적 접촉을 할 때 감염될 가능성이 미미하게 있다고 보고되고 있다(그림 4-4] 참조).

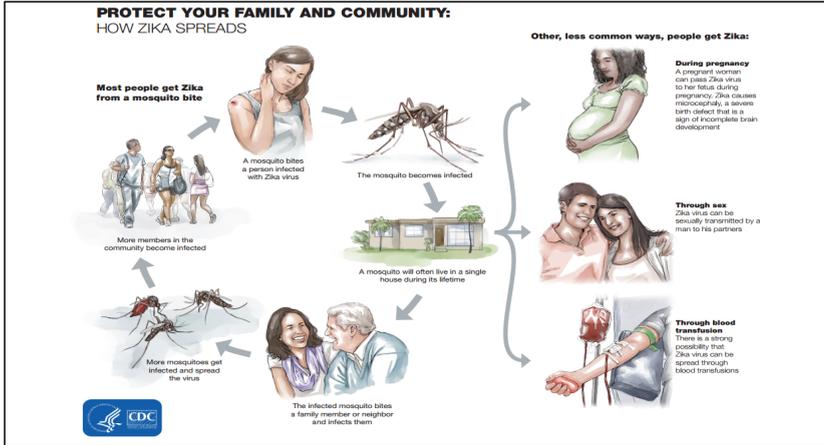
질병관리본부에서는 가임 여성의 경우 최소 2개월 동안 임신을 연기하고, 남성의 경우 배우자 등이 임신 중인 경우 임신 기간 동안 금욕 또는 콘돔 사용, 임신 중이 아닐 경우 최소 2개월 동안 금욕 또는 콘돔 사용, 확진 환자의 경우 회복 후 최소 6개월 동안 금욕 또는 콘돔 사용을 할 것을 발표했다(질병관리본부, 2016.2.2.).

[그림 4-3] 모기에 의한 바이러스 이동 경로



자료: New York Times, Short answers to Hard Questions About Zika Virus.
Updated March 18, 2016.
<http://www.nytimes.com/interactive/2016/01/16/health/10000004149566.mobile.html>에서 2016. 1. 16. 인출.

[그림 4-4] 지카바이러스 감염 경로



자료: Centers for Disease Control and Prevention. How Zika Spreads.

<http://www.cdc.gov/zika/transmission/>에서 2016. 5. 9. 인출.

이 외에도 국민 행동 수칙을 발표하여 예방 대책을 내놓았지만, 국내 지카바이러스 감염 확진자가 발생함에 따라 의심 신고 건수도 증가하고 있는 것으로 나타났다(연합뉴스, 2016.5.5.). 현재까지 국내 지카바이러스 감염 확진자는 5명으로 신고되었다. 첫 번째 확진자부터 네 번째 확진자까지 모두 모기에 의해 감염된 것으로 확인되었으며, 다섯 번째 확진자도 모기에 물린 것으로 추정하고 있다(표 4-1) 참조).

<표 4-1> 국내 지카바이러스 확진자 현황

| 순번 | 성별 | 연령 | 감염국 | 감염 매개 | 체류 기간 |
|----|-----|----|-----|-------|-------------|
| 1 | 남자 | 43 | 브라질 | 모기 | 2.17.~3.9. |
| 2 | 남자 | 20 | 필리핀 | 모기 | 4.10.~4.14. |
| 3 | 남자* | 21 | 필리핀 | 모기 | 4.10.~4.14. |
| 4 | 여자 | 25 | 베트남 | 모기 | 4.10.~5.1. |
| 5 | 남자 | 39 | 필리핀 | 모기 | 4.27.~5.4. |

주: *2의 친형.

자료: 질병관리본부 보도자료.

우리나라의 경우 지카바이러스가 동남아를 통해 유입되고 있음이 확인돼 각별한 주의가 필요할 뿐 아니라 지카바이러스 초기 발병 시 질병 관리본부에서는 국내의 계절이 겨울철이기 때문에 모기를 통한 바이러스 유입 가능성이 낮을 것으로 전망했다. 바이러스 매개체가 될 수 있는 ‘흰줄숲모기’가 국내에서 6월부터 증가하기 시작해 8~9월에 정점을 이루는 것으로 알려져 지카바이러스에 대한 정부 차원의 대응책 마련이 필요하다. 2016년 5월 12일 현재 지카바이러스는 58개국에서 감염이 보고되었다. 58개국 중 45개국이 모기에 의한 감염국으로 최초 보고되었으며, 13개국은 2007년부터 2014년 사이 지속적으로 지카바이러스 감염을 보고하였다. 이 중 Cook Islands, French Polynesia, ISLA DE PASCUA - Chile and YAP, Federated States of Micronesia은 지카바이러스 유행 종식을 보고하였다[표 4-2 참조(Rodolphe Hamel, Florian Litgeois, et al., 2016; WHO, 2016.05.12.)].

〈표 4-2〉 발생 구분별 지카바이러스 감염 국가

| 구분 | WHO 지역 | 발생국(지역) | |
|---------------------|-------------------|--|----|
| 최초 발생 -모기에 의한 감염 | 아프리카(AFRO) | Cabo Verde | 1 |
| | 미주 (AMRO/PAHO) | Aruba, Barbados, Belize, Brazil, Bolivia (Plurinational State of), BONAIRE Netherlands, Colombia, Costa Rica, Cuba, Curacao, Dominica, Dominican Republic, Ecuador, El Salvador, French Guiana, Grenada, Guadeloupe, Guatemala, Guyana, Haiti, Honduras, Jamaica, Martinique, Mexico, Nicaragua, Panama, Paraguay, Peru*, Puerto Rico, Saint Barthelemy, Saint Lucia, Saint Martin, Saint Vincent and the Grenadines, Sint Maarten, Suriname, Trinidad & Tobago. | 38 |

| 구분 | WHO 지역 | 발생국(지역) | |
|------------------------------|------------|--|----|
| | | United States Virgin Islands, Venezuela (Bolivarian Republic of) | |
| | 동남아(SEARO) | Maldives | 1 |
| | 서태평양(WPRO) | American Samoa, Fiji, Marshall Islands, Samoa, Tonga | 5 |
| 계 | | | 44 |
| 과거 발생 현재 전파 중이거나 종식된 곳 | 아프리카(AFRO) | Gabon | 1 |
| | 동남아(SEARO) | Bangladesh, Indonesia, Thailand | 3 |
| | 서태평양(WPRO) | Cambodia, Cook Islands, French Polynesia, Lao People's Democratic Republic, Malaysia, Micronesia (Federated States of), New Caledonia, Papua New Guinea, Philippines, Solomon Islands, Vanuatu, Viet Nam | 12 |
| | 미주(PAHO) | SLA DE PASCUA - Chile | 1 |

자료: World Health Organization, SITUATION REPORT, ZIKA VIRUS MICROCEPHALY GUILLAIN-BARRÉ SYNDROME(2016), pp.3.

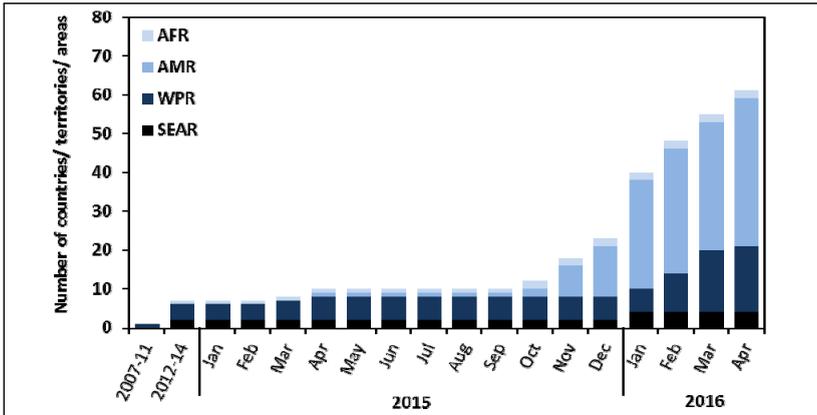
이들 국가 중 성관계를 통한 지카바이러스 감염 국가는 <표 4-3>과 같이 보고되었다.

<표 4-3> 성관계를 통한 지카바이러스 감염 국가

| 구분 | WHO 지역 | 발생국(지역) | |
|-----------|----------------|--|---|
| 성관계 통한 전파 | 미주 (AMRO/PAHO) | Argentina, Canada, Chile, Peru, United States of America | 5 |
| | 유럽(EURO) | France, Italy, Portugal | 3 |
| | 서태평양 (WPRO) | New Zealand | 1 |
| 계 | | | 9 |

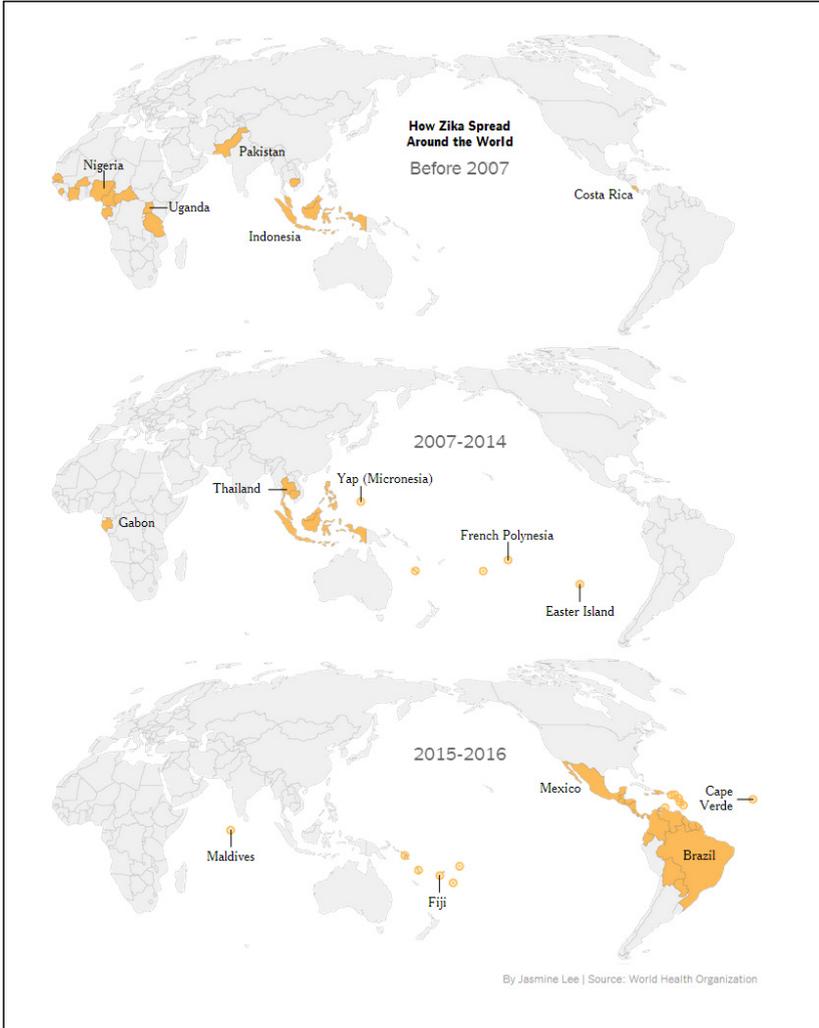
자료: World Health Organization, SITUATION REPORT, ZIKA VIRUS MICROCEPHALY GUILLAIN-BARRÉ SYNDROME(2016), pp.3.

[그림 4-5] WHO 지역 구분별 지카바이러스 감염 누적 국가 수
(기간 2007.~2016.4.27.)



자료: World Health Organization, SITUATION REPORT, ZIKA VIRUS MICROCEPHALY
GUILLAIN-BARRÉ SYNDROME(2016), pp.2.

[그림 4-6] 지카바이러스 확산 현황



자료: New York Times, Short answers to Hard Questions About Zika Virus.

Updated March 18, 2016,

<http://www.nytimes.com/interactive/2016/01/16/health/10000004149566.mobile.html>에서

2016. 1. 16. 인출.

제2절 지카바이러스 소셜 빅데이터 온라인 문서 현황

1. 분석 방법

가. 연구 대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜네트워크서비스(SNS), 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 대상으로 하였다. 본 분석에서는 99개의 온라인 뉴스 사이트, 3개의 블로그(네이버·다음·티스토리), 2개의 카페(네이버·다음), 1개의 SNS(트위터), 10개의 게시판(네이버지식인·네이트지식·네이트톡·네이트판 등)의 총 115개 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서(버즈)를 소셜 빅데이터로 정의하였다. 지카바이러스 관련 토픽은 2016. 1. 1.~2016. 4. 15. 해당 채널에서 요일별, 주말, 휴일을 고려하지 않고 매시간 단위로 수집하였으며, 수집된 총 6만 5673건의 텍스트(Text) 문서를 본 연구의 분석에 포함시켰다. 본 연구를 위한 소셜 빅데이터의 수집²¹⁾은 크롤러(Crawler)를 사용하였고, 토픽의 분류는 주제분석(text mining) 기법을 사용하였다. 지카바이러스 토픽은 모든 관련 문서를 수집하기 위해 ‘지카바이러스’, ‘Zikavirus’ 그리고 ‘지카바이러스 감염’ 등을 사용하였다.

21) 본 연구를 위한 소셜 빅데이터의 수집 및 토픽 분류는 ‘(주)SK텔레콤 스마트인사이트’에서 수행함.

나. 연구 도구

지카바이러스와 관련하여 수집된 버즈는 주제분석(text mining)과 요인분석(factor analysis)의 과정을 거쳐 <표 4-4>와 같이 정형화 데이터로 코드화하여 사용하였다.

<표 4-4> 지카바이러스 소셜 빅데이터 수집을 위한 분류

| 대분류 | 중분류 | 키워드 |
|----------------|------------------|---|
| 개인 수준의 인구학적 특성 | 대상 (일반 시민) (305) | 남편, 농부, 노약자, 가족들, 고딩, 가족, 개인, 국민, 고삐리, 국민들, 노인, 대딩, 둘째, 본인, 부모, 부인, 비지니스맨, 사람, 사람들, 성인, 소상공인, 시민, 시민들, 신랑, 아기, 아내, 아들, 아빠, 아이, 아이들, 애들, 어린이, 어린이들, 어머니, 언니, 엄마, 오빠, 와이프, 외국인, 인간, 일반인, 임산부, 자영업자, 주민들, 주부, 중딩, 중삐리, 직장인, 질환자, 첫째, 친구, 친구들, 커플, 패밀리, 패밀리, 학생, 학생들, 환자, 환자들, 웨미리, 할머니, 할아버지, 아버지, 기러기아빠, 기러기엄마, 부부, 계부, 계모, 장모, 동생, 애인, 연인, 대학생, 회사원, 가해자, 피해자, 성추행범, 성폭행범, 도둑, 어른, 노인들, 집단, 공무원, 간호조무사, 간호사, 의사, 수급자, 근로자, 엘리트, 단역배우, 서민, 비정규직, 여성환자, 남성환자, 여대생, 직딩, 편부, 편모, 계부모, 조부모, 조부, 조모, 형, 누나, 남동생, 여동생, 부모간, 부모자녀, 형제, 청년, 부, 모, 학교교사, 상담교사, 교내학생, 초등학생, 중등학생, 교사, 일반대중, 일반청소년, 사회전체, 자살시도청소년, 생존가족, 생존친구, 나, 선배, 후배, 선생, 한국인, 새터민, 미국교포, 미국동포, 중국동포, 중국교포, 친부, 친모, 형제자매, 외조부모, 삼촌, 외삼촌, 고모, 고모부, 이모, 이모부, 자매, 일반의, 정신과전문의, 심리전문상담사, 정신보건전문가, 보건의료사, 가족치료사, 음악치료사, 미술치료사, 아로마치료사, 운동치료사, 남매, 감염자, 격리대상자, 자가격리자, 격리자, 파병, 장병, 파병자, 확진자, 접촉자, 밀접접촉, 관광객, 의료진들, 의료진, 교수, 교수님, 전문가, 발병자, 만성질환자, 신고자, 사망자, 의심환자, 사병, 의심병사, 중국인, 미국인, 중동인, 아랍인, 의심자, 유포자, 감염유포자, 직원, 보균자, 초중학생, 우리애, 병사, 확진환자, 미군병사, 감염환자, 경찰, 승객, |

| 대분류 | 중분류 | 키워드 |
|--------------------------------|----------------------------|---|
| | | <p>관계자, 종사자, 학자들, 학자, 슈퍼전파 SFTS환자, 중증열성혈소판감소증후군환자, 무슬림, 첫감염자, 목사님, 목사, 당철자, 탑승객, 엘리어답터, 아르바이트생, 쌍둥이, 여직원, 업주, 민간전문가, 해외건설근로자, 건설근로자, 쪽바리, 박사, 연예인, 감염전문가들, 직장인들, 입국객, 낚시꾼, 급식요원, 여행객들, 여행객, 우리국민, 증상자, 출국, 간병자, 여자환자, 남자환자, 판료들, 판료, 직장동료, 이탈자, 무단이탈, 슈퍼보균자, 사무차장, 대변인, 중증환자, 능동감시대상자, 능동감시자, 상인들, 상인, 예약자, 최초감염자, 최초확진자, 퇴원자, 소비자, 판매자, 투자자, 신생아, 임신부, 태아, 유전자, L씨, 산모, 여성들, 임신부들, 책임자, 배우자, 입국자, 임신부들, 가입여성, 주민, 대국민, 대상자별, 아기들, 여행, 방문객, 태국인, 성인환자, 소비자들, 선수들, 내국인, 첫환자, 영국인, 맘머느리, 직원들, 조선족, 유아, 빈곤층, 관광객들, 젊은이들, 운전기사들, 영유아, 안전요원, 여행자들, 외국인들, 양성환자, 여자친구, 방문객들, 군인, 일반인들, 지카 바이러스환자, 신생아들, 승객들, 부모님, 여중생, 남성분, 고등학생, 감염환자, 만성간염환자, 발병환자</p> |
| | <p>성별/ 나이 (31)</p> | <p>gender, 남자, 여자, 남성, 여성, 남녀, 10대, 20대, 30대, 40대, 50대, 60대, 70대, 80대, 60대남성, 70대환자, 10대여성, 20대여성, 30대여성, 40대여성, 50대여성, 60대여성, 70대여성, 80대여성, 10대남성, 20대남성, 30대남성, 40대남성, 50대남성, 70대남성, 80대남성</p> |
| <p>개인의 주관적 인지 및 감정</p> | <p>감정(긍정) (194)</p> | <p>다행, 해결, 행복, 든든, 완벽, 안정, 안전, 깨끗, 기대, 감동, 격려, 극복, 긍정, 기대감, 기쁨, 도움, 미소, 믿음, 따뜻, 선호, 성공, 소망, 소중, 희망, 힐링, 훈훈, 흥미, 활력, 활발, 배려, 보답, 활약, 사랑, 선처, 스마트, 승리, 화이팅, 환영, 최고, 최선, 치유, 친절, 재미, 웃음, 즐거움, 찬양, 양보, 상승세, 여유, 열정, 염원, 자랑, 자신감, 자유, 우수, 응원, 이해, 인기, 지지, 진심, 진정, 추천, 축하, 존경, 존중, 편리, 편안, 평화, 평화적, 풍요, 가능성, 간단, 간편, 강조, 강추, 개선, 결심, 고급, 공감, 과감, 귀여움, 긍정적, 깔끔, 께찍, 날씬, 노력, 농담, 느긋, 느낌, 늠름, 다양, 다짐, 단순, 달콤, 당당, 대단, 대박, 도전, 독특, 동심, 득템, 릴렉싱, 만끽, 만족, 매력, 매력적, 몰입, 무난, 반복, 변화, 보호, 비교, 부듯, 산뜻, 상쾌, 선택, 세련, 센스, 소소, 소중한, 소통, 순수, 시원, 신기, 신뢰, 신비, 신선, 신중, 실속,</p> |

| 대분류 | 중분류 | 키워드 |
|-----|-------------------------|---|
| | | <p>솔솔, 아름다움, 안심, 알뜰, 애국심, 애도, 약속, 양호, 역대급, 열광, 완화, 용서, 위대, 위로, 위안, 유명, 유용, 유의, 의지, 익숙, 인내, 인정, 자제, 장난, 적극, 적극적, 적절, 정직, 즐기기, 지름신, 집중, 짱짱, 최애, 추억, 흥분, 흥성, 쾌적, 탁월, 포근, 푸짐, 풍부, 풍성, 합리, 호기심, 화려, 화제, 환상, 회복, 효과, 훌륭, 휴식, 모범, 정확, 신속, 상승, 추진, 협력, 권장, 각별, 협조, 도입, 구조, 안전성, 금상첨화, 긍정적</p> |
| | <p>감정(부정) (268)</p> | <p>답답, 거짓말, 비상, 판단, 불안, 스트레스, 냉소, 공포, 혼란, 엄벌, 공포증, 위협, 우려, 문제, 긴급, 부담, 약화, 난리, 비판, 무책임, 갈등, 감소, 갑갑, 강제, 거부, 거짓, 검토, 경고, 경박, 경악, 고민, 고생, 고통, 곤혹, 골치, 과도, 과시, 귀차니즘, 규제, 규탄, 극심, 급락, 기만, 긴장, 까기, 꼼수, 끄찍, 나쁨, 난형난제, 논란, 눈물, 눈치, 다툼, 단점, 답답, 두려움, 두렵다, 둔화, 똥판지, 리스크, 마음고생, 막장, 모욕, 무관심, 무능, 무리, 무시, 미안, 민감, 바보, 반대, 반발, 반성, 방탕, 방해, 배신감, 변명, 복수, 복잡, 부적절, 부정, 부정적, 부족, 부진, 분노, 분열, 불가능, 불과, 불구, 불리, 불만, 불법, 불순, 불신, 불쌍, 불안감, 불청객, 불쾌감, 불편, 불편함, 불행, 불확실, 비난, 비방, 비아냥, 비통, 비하, 비하한, 빠친다, 사기, 상실, 선정적, 섭섭, 성질, 소름, 손실, 수상, 수치심, 슬픔, 시비, 신경질, 실망, 실수, 실패, 심각성, 쓰레기, 아쉬움, 아쉬워, 아쉽다, 아유, 악마, 악영향, 안타까움, 안타깝다, 압박, 야유, 약세, 어려움, 어물, 어물쩍, 억울, 엄격, 엄두, 엄중, 염려, 예민, 오해, 왜곡, 외면, 욕설, 우울, 우울증, 울적, 울화, 위기, 위반, 위험성, 위협, 유죄, 유치, 은밀, 은폐, 의심, 의혹, 자극, 자극적, 잔인, 잔혹, 잘못, 재앙, 저렴, 저주, 적나라, 조롱, 죄송, 죽음, 지랄, 지루, 지적, 지친, 집착, 짜증, 짜증나, 쪽박, 착각, 창피, 천박, 철회, 최악, 충격, 치명, 치열, 침해, 칼갈, 탄핵, 탐욕, 통감, 통제, 속수무책, 퇴진, 퇴치, 퇴폐, 파괴, 파렴치, 폐지, 포기, 피곤, 피로, 피해, 하락, 하한가, 한숨, 한심, 향의, 해명, 해지, 허약, 현혹, 혐오, 혐오감, 협박, 호소, 화, 환장, 황당, 후폭풍, 후회, 흐림, 흥분, 힘들다, 고발, 단절, 무제한, 사회악, 상처, 수상한, 야한, 욕구, 위협한, 유혹, 자극적인, 중독성, 천박한, 충격적인, 퇴폐적, 파렴치한, 취소, 혼탁, 난동, 후덜덜, 비상사태, 제거, 무단, 감시, 설치, 불가피, 제한, 공격, 사상최악, 침투, 공포감, 자살, 개고생, 단속, 경각심, 고리타분, 큰일, 낙폭</p> |

| 대분류 | 중분류 | 키워드 |
|-------------|----------------------|---|
| 감염 관련 정보 | 바이러스 (질병) (91) | 사스, 중증급성호흡기증후군, 신종플루, 바이러스, severeacuterespiratorysyndrome, 조류독감, 구제역, SARScoronavirus, avianinfluenza, 조류인플루엔자, H1N1, 신종인플루엔자, 에볼라, 에볼라바이러스, ebolahemorrhagicfever, 에볼라출혈열, ebola, 라사열, Lassafever, 베타코로나바이러스, 질병, 생탄저균, 신종바이러스, 변이형, 중증열성혈소판감소증후군, 병균, 베타코로나, 항바이러스, 항체, 급성신부전증, 급성신부전, 죽상경화증, 사구체염, 급성호흡곤란증후군, 다기관부전증, 복막염, 췌장염, 패혈증, 부정맥, 심부전, 간질, 발작, 혼수, 인수공동전염병, 신종감염병, 감염병, 전염병, 법정전염병, 기저질환, 고열증, 질병정보, 중증급성호흡기질환, 증후군, 독감, 지카 바이러스, 바이스, 지카, ZIKA, VIRUS, 법정감염병, VIRUSZIKA, 뎅기, 황열, 뎅기열, 치쿤구니아, 황열병, 플라비바이러스, 길랑바레증후군, 길바레, 유행병, 에이즈, 지카바이러스, 소두증바이러스, 특소플라스증, 결핵, 인플루엔자, 자카바이러스, 가카바이러스, 바이러스 감염, 바이러스들, 바아르스, 바레증후군, 뎅기열바이러스, 동물인플루엔자, H7N9, H5N6, 호흡기질환, 제두이러스, Guillainbarresyndorme, 에이즈바이러스, 길랑바레증후군, |
| | 감염 경로 (98) | 밀접접촉, 농작물재배, 접촉, 가금류, 미확정, 1감염, 2차감염, 3차감염, 박쥐, 염소, 감염로, 공기감염, 호흡기 감염, 공기, 대상, 공기전염, 재채기, 원인, 감염원, 밀접, 동물, 작은소참진드기, 진드기, 공기호흡, 농장방문, 농장, 비말감염, 동물들, 감염원, 야생진드기, 인체감염, 지역사회, 사회감염, 4차감염, 지역사회감염, 가족감염, 전파경, 공기전파, 경구전파, 모기, 여행, 해외여행, 흰줄숲모기, 성관계, 숲모기, 이트숲모기, 매개체, 귀국, 출장, AEDES, 성접촉, 유충, 원숭이, 입국, AEGYPTI, 전쟁, 정액, 매개모기, 키스, 서식지, 모기들, 작은빨간집모기, 토종모기, AEDES종, CULEX, 열대모기, 감염성질환, 타액, 성적접촉, 뇌상, 성관계한, 성행위, 질병매개, 월동모기, 해충, 전염경로, GM모기, 모기충, 성관계만, 질병매개모기, 뇌염모기, 모기문제, 암모기, 일본뇌염모기, 모기활동, 지카숲, AedesAegypti, 기생충, 해외여행력, 현혈, AedesAlbopictus, 수혈, 감염자접촉, 모기전파, 성관계, 체액감염, 아테스기, 임신 |
| | 증상 (151) | 감기, 기침, 폐감염, 발열, 고열, 치사율, 취사율, 전염성, 전염, 잠복기, 잠기간, 면역력저하, 호흡기증상, 위 |

| 대분류 | 중분류 | 키워드 |
|-----|------------------------------|---|
| | | <p>장장애, 합법증, 경련, 혼수상태, 혈액, 식부진, 구토, 설사, 오심, 혈전증, 외상, 콩팥내종양, 콩팥종양, 전립선비대, 결석, 종양, 탈수, 과다출혈, 화상, 심낭액저류, 감각이상, 수지진전, 섬망, 숨가쁨, 호흡기이상, 호흡이상, 무력감, 두통, 근육통, 인후통, 복통, 폐렴, 가래, 오한, 숨가쁨, 탄저균, 감염, 발생, 확산, 증세, 호흡곤란, 양성, 음성, 호흡기, 38도, 전염력, 사망, 증상, 감염율, 급성, 만성, 질환, 의심증상, 목숨, 전파력, 생명, 혈소판감소, 염증반응, 출반응, 병감, 미열증세, 미열, 중증질환, 신부전증세, 폐렴증세, 출혈, 신장기, 발병, 목감기, 최대잡복기, 흉통, 호흡곤란증세, 감염력, 의심증세, 치사40, 메스꺼움, 분비물, 소화기증상, 소화기관, 불면, 14일, 2주, 소두증, 감증, 발진, 관절통, 결막염, 충혈, 기형, 합병증, 마비, 건강상태, 장애, 정신체, 일본뇌염, 길랭, 전신마비, 신경마비, 뇌성마비, 발달장애, 소아마비, 기아, 반점구신성, 안구충혈, 당뇨, 뇌염, 피부발진, 염증, 몸살기운, 감염증, 신경장애, 정신질환, 중추신경마비, 심장마비, 피부노화, 태아소두증, 소뇌, 뇌졸중, 뇌수막염, 소두증신생아, 소아소두증, 태아기형, 피부염, 피부붉은진, 눈통증, 눈충혈, 결절증, 손목관절통, 정신장애, 소안구증, 관절구축, 말신경계이상, 파다반사증, 권태감, 운동신경마비, 감각신경마비, 전신마비, 소두증</p> |
| | <p>대처 (치료) (207)</p> | <p>확진, 격리, 치료약, 백신, 대처법, 치료, 지원, 긴급복지지원제도, 생계비, 교육비, 전기세, 주거비, 연료비, 진단, 채혈, 항체검사, 자가격리, 격리대, 인공호흡기, 입원, 감염증세, 양성판정, 음성판정, 유전자검사, 치료제, 초증상, ZMapp, 지맵, TKM에볼라, 파비피라비르, 브린시도피어, 1차치료, 2차치료, 추가, 판정, 확인, 조치, 대응, 대처, 검사, 대책, 폐쇄, 발표, 초, 진료, 관리, 초기발견, 발견, 초기대응, 해결방법, 자택격리, 집중치료, 감염부, 신고, 종합대응방안, 수사, 초강경대책, 자연신고, 차단, 방역체계, 역학사, 총력, 방지, 철저, 실험, 개발, 강화, 의료계열, 검사결과, 격리조치, 방, 혈액검사, 산소공급, 병원진찰, 전파가능성, 약물, 산소포화도, 긴급비상회, 국가지정입원치료, 대응책, 후속조치, 종합대책, 긴급현안, 가택격리, 대응상황, 대처상황, 양성반응, 음성반응, 유사증상, 신원검증, 특별검역, 정상범, 엑스레이, 발열감지, 검역절차, 검역체계, 격리종료, 감염가능성, 대응단계, 의심사례, 수술, 적외선, 발열검사, 위기대응, 격리검사, 환자접촉일, 위급상황, 치료백신, 발열감시, 혈액투석, 마련, 긴급생계지</p> |

| 대분류 | 중분류 | 키워드 |
|-----|---------------------|--|
| | | <p>원, 음압, 구호품, 자가진단표, 휴교, 피해저금리신용대출, 금융지원, 폴리펩티드인HR2P, 경영안전지원자금, 대출, 긴급운영자금, 징수유예, 납기연기, 자금지원, 국방역망, 확산방지, 살균, 방역, 퇴원, 재난관리기금, 예비비, 에크모, 브린시포비르, 수혈, 당부, 연구, 분석, 회의, 치료법, 공중보건, 방제, 방안, 기록, 상담, 박멸, 마거릿, 검역, 예방접종, 진상규명, 완치, 연구소, 소년, 처방, 진통제, 해열제, 가임기, 초음파, 격리치료, 진단 키트, 보호조치, 진단검사, 방역소독, 임상시험, 연구개발, 정밀검사, 예방백신, 피검사, 지카백신, 영양상태, 보건위생,약품, 백신개발, 에볼라백신, 접종, 감염회, 면역체계, 뇌염백신, 신종플루백신, 진료가이드라인, 간염치료제, 간염치료백신, 정액검사, 치료방법, 지카 바이러스검출, 치료제백신, 의뢰진상담, 황열예방접종, 콜레라예방접종, 말라리아예방약, 예방약, 기타예방접종, 항생, 면역항암제, 피임약, 사후피임약, 병원치료, 지카 바이러스백신, 바이러스백신, 에볼라치료제, 독감백신, 소변검사, 간염치료백신, 만성간염균자치료제, 약물치료, 아세트아미노펜, 이부프로펜</p> |
| | <p>예방 (182)</p> | <p>손씻기, 마스크, 예방, 예방법, 입막음, 비누, 티슈, 손수건, 손세정제, 손소젼, 손날, 소독, 청결, 면역력강화, 면역력, 유산균, 쇼핑, 나들이, 건강관리, 소풍, 위생, 개인위생, 감염예방수칙, 예방수칙, 음식, 감염예방, 양파, 알코, 알콜, 소독제, N95, 눈보호장비, 증상체크, 체온, 환기, 오렌지, 마늘, 고구마, KF80, 특등급, 1등급, 공산품마스크, 보건용마스크, KF94, 방역마스크, 황사용마스크, 애니가드, 에탄올, 글리세린, 정제수, 방호복, 계란, 달걀, 모히또, 슈퍼푸드, 흑마늘진액, 흑마늘, 상추, 깻잎, 미나리, 샐러드, 채소, 야채, 매실, 매실차, 아사이베리, 그라비올라, 안토시아닌, 피토케미, 파이토케미컬, 단백질, 비타민E, 비타민D, 칼슘, 휴지, 고등어, 브로콜리, 철포라판, 유허화합물, 비타민C, 베타카로틴, 항산화물질, 오메가3, 토마, 무즙, 에스터C, 스피루리나, 핸드워시, 프로폴리스, 위생용품, 초유, 프로바이오틱스, 향균, 클로렐라, 과즙, 건강즙, 건강기능식품, 건기식, 홍삼정, 호암타블렛, 뿌리삼, 홍이장균, 프로바이오틱, 듀오락, 쉐바이오틱스, 듀오락케, 듀오락얌, 통마늘진액, 도라지배즙, 천호식품, 알리신, 산시울, 곤돌, 살제, 모기장, 모기기피제, 기피제, 방충망, 행동수칙, 긴바지, 긴급회의, 지, 긴팔, 좋은정보, 소매, 수분섭취, 대책마련, 가이드라인, 공동대응, 방역조, 긴소매, 예방하기, 피임, 면역, 퇴치제, 차관회의, 대책회</p> |

| 대분류 | 중분류 | 키워드 |
|--------------|-------------|---|
| | | 의, 모기방제, 경계, 금지령, 모기조심, 모기퇴치제, 예방책, 보도자료, 살충방제, 모기약, 대책, 대응조치, 모기퇴치, 신속대응, 회피요령, 방제증명서, 모기방역, 시민동수칙, 예방조치, 가습기살균제, 감염병예방법, 모기분포조사, 콘돔광고시, 키스금지령, 콘돔사용, 집중방역, 예방대책, 예방활동, 모기살충제, 긴옷, 밝은색옷, 모기퇴치제품, 냉방, 예방접종기관안내, 마스크착용, 황사마스크, 지카 바이러스예방용품, 천연모기퇴치제, 담검사, 카바이러스예방, 메트린, 충분한휴식, 휴식, 살충제살포, 향수자제, 긴팔옷착용 |
| 사회적 관련 요소 | 병원 (247) | 상급종합병원, 종합병원, 국민안심병원, 가톨릭서울성모병원, 경희대학교병, 경희대병원, 고려대학교구로병원, 고려대학교안암의과대학부속병원, 고대학교안암병원, 고대안암병원, 연세대강남세브란스병원, 강남세브란스병, 연세대세브란스병원, 이대목동병원, 중앙대학교병원, 중앙대병원, 한양대학병원, 한양대병원, 강북삼성병원, 서울대학교병원, 서울대병원, 서울아산병, 강동성심병원, 명지성모병원, 부민병원, 삼육서울병원, 순천향대학교서울병, 순천향대서울병원, 을지병원, 인제대학교상계백병원, 인제대상계백병, 인제상계백병원, 인제대학교서울백병원, 인제서울백병원, 중앙보훈병원, 가톨릭대학교여의도성모병원, 가톨릭대여의도성모병원, 가톨릭여의도성모병원, 베스티안병원, 보라매병원, 서울성심병원, 서울적십자병원, 성바오로병원, 한림대강남성심병원, 한림대학교강남성심병원, 영등포병원, 한강수병원, 강북힘찬병원, 더조은병원, 목동힘찬병원, 고신대학교복음병원, 고신대복음병원, 인제대학교부산백병원, 동아대학교병원, 부산대학교병원, 광혜병원, 삼육부산병원, 온종합병원, 인제대학교해운대백병원, 좋은삼선병원, 해동병원, 인당의료재단부민병원, 김원목기념봉생병원, 대동병원, 부산성모병원, 영도병원, 왈레스기념침례병원, 부산힘찬병원, 장립한서병원, 인하대학교병원, 길의료재단길병원, IS한림병원, 검단탑병원, 부평세림병원, 인천광역시의료원, 나은병원, 온누리병원, 의료법인담우의료재단현대유비스병원, 현대유비스병원, 부평힘찬병원, 계명대동산병원, 계명대학교동산병원, 대구가톨릭대칠곡가톨릭병원, 대구의료원, 칠곡경북대학교병원, 칠곡경북병원, 칠곡경북대병원, 드림병원, 한영한마음아동병원, 울산대병원, 울산대학교병원, 전남대병원, 전남대학교병원, 조선대병원, 조선대학교병원, 서광병원, 충남대병원, 충남대학교병원, 대전선병원, 대전한국병원, 유성선병원, 고려대안산병원, 고대 |

| 대분류 | 중분류 | 키워드 |
|-----|-----|--|
| | | 안산병원, 고려대학교안산병원, 순천향대학교부천병원, 순천향대부천병원, 아주대병원, 아주대학교병원, 가톨릭성빈센트병원, 가톨릭의정부성모병원, 경기도의료원 안성병원, 국민건강보험공단일산병원, 남양주한양병원, 뉴고려병원, 동국대학교일산병원, 동국대일산병원, 명지병원, 분당제생병원, 세종병원, 시화병원, 신천연합병원, 안양샘병원, 오산한국병원, 원광대의과대학산본병원, 원광대의산본병원, 원광대학교의과대학산본병원, 인제대학교일산백병원, 지샘병원, 차의과대학교분당차병원, 분당차병원, 현대병원, 가톨릭부천성모병원, 강남병원, 다보스병원, 동의성단원병원, 센트럴병원, 우리의료재단김포우리병원, 김포우리병원, 의료법인동수원병원, 한양대구리병원, 한양대학교구리병원, 안성성모병원, 아가페의료재단시티병원, 이천파티마병원, 장산의료재단이춘택병원, 추병원, 연세대원주세브란스기독병원, 강릉동인병원, 속초보광병원, 충북대학교병원, 건국대학교충주병원, 제천시울병원, 한마음의료재단하나병원, 건명의료재단진천성모병원, 옥천성모병원, 의료법인인화재단한국병원, 청주의료원, 청주효성병원, 충주의료원, 괴산성모병원, 순천향대학교천안병원, 단국대학교의과대학부속병원, 백제병원, 천안충무병원, 당진종합병원, 원광대학교병원, 전북대학교병원, 동군산병원, 부안성모병원, 고창병원, 예수병원유지재단전주예수병원, 전주열린병원, 목포기독병원, 목포중앙병원, 성가롤로병원, 세안종합병원, 순천한국병원, 순천중앙병원, 해남병원, 전남중앙병원, 차의과학대학교구미차병원, 김천제일병원, 동국대학교의과대학경주병원, 안동병원, 포항세명기독병원, 창원파티마병원, 대우의료재단대우병원, 진주고려병원, 창원한마음병원, 제주한마음병원, 의료법인승연의료재단삼천포서울병원, 창원협찬병원, 제주대학교병원, 제주한라병원, 서귀포의료원, 중앙의료재단에스중앙병원, 중앙의료재단이중앙병원, 365서울열린의원, 삼성서울병원, 평택성모병원, 아산서울의원, 대청병원, 하나로의원, 윤창옥내과의원, 성모가정의학과의원, 평택굿모닝병원, 평택푸른의원, 365연합의원, 박애병원, 연세허브가정의학과, 한림대동탄성심병원, 메디홀스의원, 가톨릭대부천성모병원, 신영내과의원, 단국대의대부속병원, 대천삼육오연합의원, 폐쇄병원, 병원폐쇄, 병원, 메디힐병원, 서울시립보라매병원, 건양대학교병원, 건양대병원, 창원sk병원, 집중관리병원, 진영의원, 현대의원, 원주세브란스기독병원, 힘찬병원, 가족보건의원, 이운우내과, 서울내과, |

| 대분류 | 중분류 | 키워드 |
|-----|------------------------|--|
| | | <p>이피부과의원, 어린이병원, 서북병원, 은평병원, 서울의료원, 동부병원, 북부병원, 서남병원, 국립중앙의료원, 강남성심병원, 국립의료원, 건국대학교병원, 건대원, 산부인과, 안암병원</p> |
| | <p>산업/경제 (260)</p> | <p>남대문시장, 시장, 코엑스, 박람회, 행사, 명동, 관광지, 백화점, 마트, 슈퍼마켓, 재래시장, 모란시장, 대중교통, 운송수입금, 관광산업, 서비스업, 숙박업, 요식업, 음식업, 소비심리, 매출감소, 매출상승, 전통시장, 경제손실, 항공사, 항공수요, 증가, 대한항공, 아시아나, 홈쇼핑, 대형마트, 온라인주문, 온라인몰, 택배, 배달, 배송, 면세점, 요우커, 급감, 화장품, 의류, 생활용품, 지역경제, 웰크론, 케이엠, 고객수, 경영난, 면담거부, 계약취소, 계약연기, 내부행사취소, 시가총액, 매도, 코스피, 코스닥, 주식, 주가, 외식업계, 외식, 개점휴업, 휴업, 성수기, 서양식, 양식, 일식, 중식, 기피심리, 취소율, 감소율, 서민경제, 경기보완책, 불황, 경제적손실, 통신요금면제, 데이터무제한, 이통사, 이동통신사, 요금감면, 가든5, 카드사용액, 야외활동, 배송주문, 수요, 공급, 의약품, 애슬리, 자연별곡, 할매순대구, 이랜드, 계절밥상, 빙스, CJ푸드빌, 롯데리아, 신세계푸드.</p> <p>인터넷쇼핑, 인터넷상거래, 여행업체, 소비유행, 리조트, 단체관광, 놀이공원, 유원지, 식품류, 건강식품, 구강청결제, 이마트, 롯데마트, 지마켓, 옥션, 홈플러스, 현대홈쇼핑, 롯데백화점, 신세계백화점, 통합온라인몰, SSG닷컴, 할인, 상품기획전, 공기청정기, 가든파이프, 판촉전, 주문, 판매, 신선식품, 가공식품, 오픈마켓, G마켓, 11번가, 쿠팡, 소셜커머스, 티몬, 위메프, 간편가정식, 생수, 주방용품, 화장지, 세제, 생필품, 커피, 음료, 공산품, 반찬류, 즉석밥, 카레, 홍삼, 인삼, 미네랄, 즉석식품, 아울렛, 여행사, 밀리오레, 두타, 동대문시장, 동대문상가, 이케아, 세일, 생활필수품, 국외여행, 테마주, 불공정거래, 내수회복세, 부산시티투어, 추가경정예산, 추경, 경제통인, 국가부채, GDP, 국내총생산, 관광수입, 크루즈, 로얄캐리비안크루즈, 킴오브더시즈호, 코스타빅토리아호, 중화태산호, 보하이크루즈, 코스타크루즈, 세레나호, 보이저오브더시즈호, 마리너오브더시즈호, 내수경기, 호텔, 유통, 항공, 유커, 건설, 건설업계, 숙박, 아모레퍼시픽, LG생활건강, 영화관, 영화, 건설공사, 현대건설, 현대백화점, 할인매장, 방역업계, 매출시장, 극장가, 극장, 경제, 온라인, 상한가, 태교여행, 충신, PCR, 신혼여행, 인수, 바이오니아, 국제사회, 국제보건규정, 삼성전자, 국제유가, 유유계약, 국</p> |

| 대분류 | 중분류 | 키워드 |
|-----|---------------------------------|--|
| | | <p>민행동수칙, 알파고, 우진비앤지, 더불어민주당, 정당, 중남미여행, 매출액, 순이익, 국세청, 대기업, 현대상선, 뉴욕주식시장, 무소속, 현대증권, 흑자전환, 최대주주, 이태원, 국내여행, 민주노총, 제주항공, 야권단일화, 금융시장, 부동산중개, 공인중개사법, 중개사무소, 중개행위, 공천문제, 두산그룹, 서울도시철도공사, 대한상공회의소, 콘돌회사, 삼성엔지니어링, 대우증권, 시드니국제공항, 대한민국정부, 바른생각콘돌, 삼성SDS, 국영제약회사, 바이오기술회사, 국영기업, 경기침체, 경제제재, 한국투자증권, 비영리단체, 올림픽, 하계올림픽, 경제위기, 제주국제공항</p> |
| | <p>기관 (장소/직책) (321)</p> | <p>보건복지부, 질병관리본부, 국가방역체계, 국민안전처, 세계보건기구, WHO, 종합병원응급실, 전국보건의료산업노동조합, 보건복지위원회, 새정치민주연합, 새누리당, 정부, 합동대책반, 보건당국, 국립보건원, 보안국장, 레지나입, 입법회, 홍콩주재한국총영사관, 홍콩총영사관, 아시아나항공, 위생방역센터, 홍콩공항, 인천국제공항, 인천공항, 홍콩국제공항, 방역상황반, 정부서울청사, 김포국제공항, 김포공항, UAE수의진단센터, 바이오노트, 바이오기업, 한국생명공학연구원, 미국국립보건원, NIH, 미국질병통제예방센터, CDC, 국경없는의사회, 두바이국제공항, 공항, 국방부, 감염내과, 긴급당정회의, 중국보건당국, 미국보건당국, 홍콩보건당국, 중동보건당국, 우리나라보건당국, 한국보건당국, 회사, 대책본부, 박근혜정부, 병실, 복지부, 아산병원, 수용시설, 박근혜정, 병동, ICU, 중환자실, 만원지하철, 응급실, 학원, 지하철, 질병통지센터, 지정격리병원, 격리병원, 대학교, 고등학교, 청와대, 주한미군, 초등학교, 국가지정격리병상, 도심, 시골, 화성국제공항, 한라체육관, 한국농수산물유통공사, 한국보건산업진흥원, 충주휴게소, 직거래장터, 중부내륙고속도로, 자택, 주택, 왕십리광장, 영양사협회, 역사학회, 식품공장, 보건산업진흥원, 호흡기내과, 집중치료소, 집중치료실, 질병본부, 제주검역소, 소화기내과, 제주공항, 제주대학교, 국립보건연구원, 김해공항, 진원생명과학, 방역대책본부, 공공장소, 검역소, 검사반, 인천국제공항입국장, 서울정부청사, 고대구로병원, 파병부대, 정부당국, 사우나, 국가지정격리치료병원, 정부세종청사, 동물원, 해외여행질병정보센터, 식품의약품안전처, 국립인천공항검역소, 국가격리병원, 대법원, 정부조직, 유럽질병통제청, 공공기관, 검찰, 감염관리실, 외교부, 이스타항공, 도하공항, 경남기업, 정읍보건소, 서태평양지역사무소, 지방자치</p> |

| 대분류 | 중분류 | 키워드 |
|-----|-----|--|
| | | <p>단체, 새정치연합, 보건소, 새정연, 직장, 대통령, 감염병리과장, 복지장관, 보건장관, 보건복지위원장, 감염병관리과장, 수석대표, 국무총리, 장관, 총리, 제약회사, 서울시장, 성남시장, 경기도지사, 합동평가단, 유엔전문기구, 누리고1727호, 누리고, 객차, 열차, 기차, 버스, KTX123호, KTX, 터미널, 광주관천터미널, 지역응급의료센터, 지역응급의료기관, 응급실신고기관, 권역응급의료센터, 선별진료소, 격리병동, 어린이집, 학교, 도시교통본부, 푸단대, 제주신용보증재단, 수원시장, 심리위기지원단, 콘트롤타워, 컨트롤타워, 심리지원, 마음돌봄, 상황실, 약국, 킹파드왕립병원, 창조경제혁신센터, 정책위의장, 한국거래소, 서울상인연합회, 충주시장, 의료혁신투쟁위원회, 남구보건소, 대구시의원, 대구공무원, 금융감독원, 금감원, 금융위원회, 금융위, KDI, 한국개발연구원, 원내대표, 재난상황실, 한국은행, 한국경제연구원, 총재, 원장, 인천항, 의원, 보좌진, 신라호텔, 힐집, 전경련, 대한상의, 무역협회, 중기중앙회, 전국경제인연합회, 한국관광공사, 차관, 기획재정부차관, 기획재정부, 의료기관, CDC, 당국, 보건, 보건부, 유니테스, 방역당국, 질병관리본부장, 명문제약, 긴급위원회, 문산도, 의사단체, 보건환경연구원, 국민의당, 더민주당, 더민주, 개누리, 기관, UN, 대한산부인과학회, 지자체, 녹십자엠에스, 국제보건, 보건국, 연구팀, 질병통제센터, 호세프, 국토교통부, 위원장, 후생노동성, 오리엔트바이오, 대책반, 운영위원회, 유엔, 의료재단, 오스바우두, 신속대응반, 대한감염학회, 국가위생계획생육위원회, 농림축산식품부, 식약처, 동화약품, 질병통제예방센터, 긴급상황센터, 문화체육관광부, 군대, 미국보건당국, 미국의사협회저널, 공군, 의과대학, 공정위, 연방보건, 미주지역본부, 국립보건연구소, 미래부, 방통위, 공천관리위원회, 자유권규약위원회, 재심위원들, 질병예방통제센터, 한국공인중개사협회, 국제올림픽위원회, 보건소장, 정책위원회, 미래창조과학부, 국립감염증연구소, 긴급당정협의, 환경부, 감염병관리팀, 의료계, 보건전문가들, 방역대책, 국립검역소, 범미보건기구, 공중보건국, 위생계획생육위원회, 질병관리센터, 유엔본부, 현대경제연구원, 보건위원, 융합생명공학부, 동성제약, 건강기능식품벤처제조업, 한국온라인신문협회, 위기대응총괄과, 질병예방센터, 위생당국, 국회의장, 특별위원회, 긴급대책위원회, 건강보험심사평가원, 경제공동위원회, 국가위생계획위원회, 해외여행질병정보센터, 미국질병통제센터, 질병연구소, 보건노동복지, 비상대책</p> |

| 대분류 | 중분류 | 키워드 |
|-----------|------------|--|
| | | 위원, 한국의학연구소, 경주시보건소, 감염병관리본부, 경기도재난안전본부, 알레르기전염병연구소, 문화사회정책부, 행정자치부 |
| 국가적 관련 요소 | 경보 수준 (4) | 관심, 주의, 경계, 심각 |
| | 국가 (199) | 중국, 중동, 홍콩, 대만, 대한민국, 한국, 바레인, 우리나라, 이라크, 이란, 사우디아라비아, 터키, 아시아, 두바이, Dubai, 북한, UAE, 중동국가, HongKong, Taiwan, 레바논, Lebanon, 리비아, Libya, 모로코, Morocco, Bahrain, SaudiArabia, 수단, Sudan, 시리아, Syria, 아랍에미리트연합, 알제리, Algeria, 예멘, Yemen, 오만, Oman, 요르단, Jordan, Iraq, 이집트, Egypt, 지부티, Djibouti, 카타르, Qatar, 쿠웨이트, Kuwait, 튀니지, Tunisia, 팔레스타인, Palestine, Iran, 이스라엘, Israel, Turkey, China, 서아프리카, 기니, Guinea, 포레카리아, 두브레, 킨디아, 프리아, 보파, 코나크리, 라이베리아, Liberia, 시에라리온, SierraLeone, 나이지리아, Nigeria, 콩고, RepublicoftheCongo, 미국, UnitedStatesofAmerica, 스페인, Spain, 말리, Mali, 영국, UnitedKingdom, 프랑스, France, 인도, India, 독일, Germany, 노르웨이, Norway, 케냐, Kenya, WestAfrica, 나라, 중동지역, 해외, 세계, 전계, 하와이, 아랍에미레이트, 아부다비, 국외, 중동봉, 뉴욕, 아라비아 반도, 중동여행, 감염국가, 최대감염국가, 최저감염국가, 이슬람, 중남미, 국가, 아프리카, 남미, 국제, 유럽, 일본, 동남아, 우간다, 인도네시아, 발생국, 과말라, 미주, 대륙, 스위스, 각국, 동남아시아, 제네바, 아르헨티나, 외국, 리데자네이루, 국제적, KOREA, 국가들, 캐나다, 말라리아, 국제적인, LA, 칠, 미주대륙, 슬로베니아, 전세계적, 탄자니아, 벨기에, 22개국, 말레이시아, 텍사스, 카리브해, 상파울루, 브뤼셀, 델러스, 이탈리아, 23개국, 북동, 러시아, 남미지역, 폴리네시아, 호주, 아메리카, 세아라주, 싱가포르, 21개, 카탈루냐, 플로리다, 파리, 뉴욕시, 워싱턴, 남아메리카, 북미, 동북지역, 오나와, 우루과이, 52개국, 25개국, 포르투갈, 26개국, 싱가포르, 상파울로, 베징, 캘리포니아, 캄보디아, 동아시아, 뉴질랜드, 파키스탄, 에노스아이레스, 안국가, 7개국, 중앙남미아프리카, 태평양, 수마트라섬, 남태평양, 미크로네시아, 몰디브, 세부, 섬지역, 깰버스턴, 타히티섬 |
| | 발생 국가 (51) | 세인트루시아, 볼리비아, 태국, 가이아나, 바베이도스, 프랑스령세인트마틴섬, 프랑스령과들루프, 도미니카공화국, 미국령버진아일랜드, 니카라과, 네덜란드령퀴라 |

| 대분류 | 중분류 | 키워드 |
|---------------------|--------------------|---|
| | | <p>소, 자메이카, 코스타리카, 통가, 사모아, 미국령사모아, 네덜란드령보네르, 네덜란드령아루바, 네덜란드령퀴라소, 미셜제도, 파라과이, 트리니다드, 토바고, 세인트빈센트그레나딘, 네덜란드령신트마르텐, 과테밀라, 마르티니크, 멕시코, 베네수엘라, 브라질, 수리남, 아이티, 에콰도르, 엘살바도르, 온두라스, 콜롬비아, 트리니다드 토바고, 파나마, 푸에르토리코, 프랑스령기아나, 프랑스령 생마르탱, 마셜제도, 피지, 베트남, 카보베르데, 도미니카, 쿠바, 프랑스령 뉴칼레도니아, 코스라이 제도, 파푸아뉴기니, 필리핀</p> |
| <p>시간 요인</p> | <p>계절 (13)</p> | <p>봄, 여름, 가을, 겨울, 겨울철, 봄철, 여름철, 가을철, 여름에, 황사, 겨울, 강추위, 초봄</p> |
| <p>기타 (141)</p> | | <p>필요, 세기말, 문화, 설명, 분리, 2012년, 2013년, 2014년, 2015년, 김치, 유연, 소규모, 추정, 비슷, 반응, 지정, 중심, 노출, 사과, 가능, 경미, 능력, 상태, 요구, 이유, 이동, 문의, 소리, 제외, 태도, 경험, 사실, 최초, 참고, 자료, 기본, 행동, 처벌, 심판, 아픔, 걱정, 통증, 부탁, 파악, 구명, 소식, 방문, 연락, 아침, 주말, 언론, 발표기사, 기준, 사례, 이야기, 전환, 사안, 건강, 최소, 낭비, 조심, 중대, 책임, 억지, 조사, 과징금, 중요, 각오, 정체, 조용, 주목, 중독, 필수, 전파, 유발, 유입, 진행, 유행, 선포, 사태, 영향, 최대, 동반, 실시, 예상, 연관성, 착용, 두뇌, 고려, 보유, 선천성, 시행, 지속적, 급증, 대표, 증거, 환경, 존재, EL, 급등, 관련주, 확률, 요청, 공유, 참사, 확보, 배포, 선언, 낙태, 지구촌, 발령, 대유행, 생존, 열니노, 출산, 관할, PAHO, 외출, 붉은털원숭이, 시민안전, 붉은원숭이, 해외여행질병정보사이트, 소두증원인, 급성신경계, 신경세포, 신경질환, 즐기세포, 유전적인, 뇌신경, 신체, 유전자변형, 임신초기, 임신, 임신중, 머리둘레, 두개골, 임신기간, 무증상, 2016올림픽, 리오, 리오데자네이로</p> |

제3절 연구 결과

지카바이러스와 관련하여 긍정적인 감정(안심)을 나타내는 온라인 문서는 36.1%, 보통의 감정을 나타내는 문서는 12.2%, 부정적인 감정(불안)을 나타내는 문서는 51.6%로 나타났다. 최초로 작성된 문서는 57.5%, 확산된 문서는 42.5%로 나타났다. 수집된 채널은 SNS(트위터), 카페, 뉴스, 블로그, 게시판 순으로 나타났다. 대상은 의료계, 일반인, 증상자 등의 순으로 언급되었으며 언급 국가는 아시아, 북미/호주, 남미 등의 순으로 나타났다. 감염은 모기, 성관계, 감염 경로, 접촉 등의 순으로 나타났다. 증상은 전염, 소두증, 열 등의 순으로 언급된 것으로 나타났다. 대처 방법은 초기 대응, 정부 대응, 감염 검사 등의 순으로 나타났다. 예방 방법은 예방수칙, 모기 퇴치, 피임 등의 순으로 나타났다. 산업은 올림픽, 주식, 관광산업의 순으로 나타났다.

〈표 4-5〉 지카바이러스 소셜 빅데이터 수집을 위한 분류

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) | |
|------|---------|--------------|-------|-------|--------------|--------------|
| 감정 | 긍정 | 8,968(36.1) | 감염 | 감염 경로 | 3,042(10.6) | |
| | 보통 | 3,037(12.2) | | 접촉 | 2,551(8.9) | |
| | 부정 | 12,820(51.6) | | 동물 | 1,309(4.6) | |
| | 계 | 24,825 | | 공기 | 931(3.2) | |
| 순계정 | 최초 문서 | 37,758(57.5) | | 진드기 | 499(1.7) | |
| | 확산 문서 | 27,915(42.5) | | 모기 | 15,063(52.5) | |
| | 계 | 65,673 | | 성관계 | 5,323(18.5) | |
| 채널 | 블로그 | 7,189(10.9) | | 계 | 28,718 | |
| | 카페 | 18,988(28.9) | | 증상 | 전염 | 20,778(33.5) |
| | SNS | 24,163(36.8) | | | 의심증상 | 3,152(5.1) |
| | 게시판 | 5,939(9.0) | 열 | | 5,677(9.2) | |
| | 뉴스 | 9,394(14.3) | 호흡기증상 | | 1,173(1.9) | |
| | 계 | 65,673 | 소화기증상 | | 511(0.8) | |
| 바이러스 | 메르스 | 4,127(6.7) | 사망 | | 2,160(3.5) | |
| | 사스 | 217(0.4) | 마비증상 | | 1,955(3.2) | |
| | 조류인플루엔자 | 116(0.2) | 소두증 | | 12,405(20.0) | |

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) |
|--------|---------|--------------|-------------|-------------|-------------|
| | 신종플루 | 701(1.1) | | 피부발진 | 5,050(8.1) |
| | 에볼라 | 1,520(2.5) | | 통증 | 4,410(7.1) |
| | 지카바이러스 | 38,545(62.8) | | 출혈 | 2,083(3.4) |
| | 덴기열 | 3,539(5.8) | | 합병증 | 976(1.6) |
| | 기타 바이러스 | 12,635(20.6) | | 기타 증상 | 1,651(2.7) |
| | 계 | 61,400 | | 계 | 61,981 |
| 대상 | 일반인 | 12,288(11.3) | 대처 | 초기 대응 | 8,514(25.3) |
| | 남성 | 1,216(1.1) | | 격리 | 3,323(9.9) |
| | 여성 | 4,529(4.1) | | 치료 | 3,988(11.8) |
| | 노인 | 353(0.3) | | 감염 검사 | 6,255(18.6) |
| | 아이/학생 | 4,425(4.1) | | 예방백신 | 4,880(14.5) |
| | 가족 | 1,310(1.2) | | 정부 대응 | 6,730(20.0) |
| | 싱글 | 832(0.8) | 계 | 33,690 | |
| | 외국인 | 602(0.6) | 예방 | 예방수칙 | 7,616(42.4) |
| | 여행객 | 1,814(1.7) | | 위생 | 996(5.5) |
| | 증상자 | 15,167(13.9) | | 마스크 | 305(1.7) |
| | 의료계 | 65,673(60.2) | | 면역 강화 | 1,162(6.5) |
| | 직장인/군인 | 946(0.9) | | 외출 | 500(2.8) |
| | 계 | 109,155 | | 피임 | 1,930(10.8) |
| | 국가 | 아시아 | 8,958(27.7) | 모기 퇴치 | 5,444(30.4) |
| | | 중동 | 772(2.4) | 계 | 17,953 |
| | | 아프리카 | 3,438(10.6) | 경기 | 564(5.2) |
| | | 유럽 | 4,634(14.3) | 전통시장 | 1,065(9.7) |
| | | 북미/호주 | 6,577(20.3) | 대형마트 | 371(3.4) |
| 남미 | | 4,779(14.8) | 관광산업 | 1,601(14.6) | |
| 동남아/인도 | | 3,230(10.0) | 숙박업 | 384(3.5) | |
| 계 | | 32,388 | 요식업 | 181(1.7) | |
| 기관 | 정부 | 13,052(49.2) | 산업 | 주식 | 1,864(17.0) |
| | 국제기구 | 6,098(23.0) | | 건설업 | 236(2.2) |
| | 병원 | 3,079(11.6) | | 영화산업 | 202(1.8) |
| | 민간기관 | 1,605(6.1) | | 생필품 | 860(7.9) |
| | 정당 | 707(2.7) | | 온라인쇼핑몰 | 146(1.3) |
| | 공항 | 1,511(5.7) | | 교통 | 849(7.8) |
| | 학교 | 451(1.7) | | 올림픽 | 1,957(17.9) |
| | 계 | 26,503 | | 바이오산업 | 660(6.0) |

제4절 결론

지난 3년간 에볼라 바이러스와 메르스로 인해 국내 유입이 우려되는 해외 유행 감염병에 대한 국민의 불안이 증폭되었다. 최근 브라질에서 지카바이러스가 발생하면서 그에 대한 공포의식은 더욱 증가하였으며, 국민들이 쉽게 의사소통할 수 있는 SNS를 통하여 유언비어가 확산되기도 하였다. 그럼에도 불구하고 SNS는 정보를 적시에 제공하고 정책을 전달하는 커뮤니케이션 수단으로서 질병 위험에 대한 우려를 감소시키기도 한다. 본 연구는 지카바이러스 정보 확산과 위험을 예측하는 데 필요한 온톨로지 개발을 통해 공중보건 비상 상황에 대한 대응 방안을 마련하는데 기여할 수 있을 것이다.

1. 지카바이러스 주제 분류

가. 온톨로지 구축

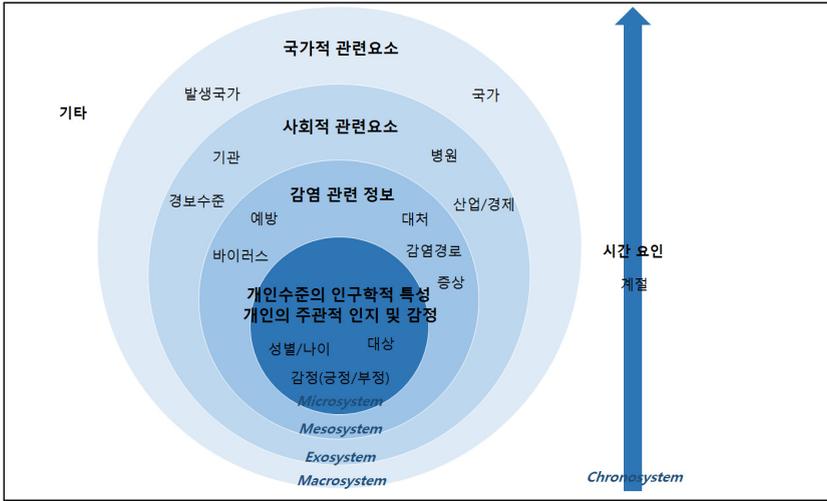
지카바이러스 소셜 빅데이터 수집을 위하여 Ontology Development 101에 근거하여 온톨로지를 개발하였으며 지카바이러스 관련 분류체계와 용어체계를 구축하였다. 첫째, 온톨로지 개발 범위는 개인 수준의 인구학적 특성, 개인의 주관적 인지 및 감정, 지카바이러스 감염 관련 정보, 사회 및 국가 수준의 관련 요소, 시간 요인으로 선정하였다. 둘째, 기존의 온톨로지를 확인하였는바, 본 연구에서는 메르스 소셜 빅데이터 분석 시 사용된 검색체계를 활용하여 온톨로지를 개발하였다. 셋째, 생태학적 이론을 바탕으로 지카바이러스 관련 분류체계를 대분류, 중분류 순으로 하였다. 넷째, 지카바이러스 관련 용어를 소셜미디어 검색을 통해 추출하였

다. 이후 단계인 EAV 모델링 개발 및 Protege를 이용한 온톨로지 표현, 온톨로지 content coverage 평가는 향후에 진행하여야 한다.

나. 온톨로지 분류

지카바이러스와 관련한 소셜 빅데이터 수집에 앞서 용어 추출에 대한 다 수준적 접근을 위해 생태학적 체계 이론을 바탕으로 주제 분류를 하였다. 생태학적 관점에서의 미시체계(Microsystem), 중간체계(Mesosystem), 외체계(Exosystem), 거시체계(Macrosystem), 시간체계(Chronosystem)를 기준으로 지카바이러스와 관련한 온톨로지를 개인 수준의 인구학적 특성, 개인의 주관적 인지 및 감정, 감염 관련 정보, 사회적 관련 요소, 국가적 관련 요소, 시간 요인 등 6개의 대분류 및 기타로 구분하였다. 하위 영역은 17개의 중분류로 구성하여 2763개의 용어를 추출하였다.

[그림 4-7] 지카바이러스 온톨로지 분류



[그림 4-8] 생태학적 관점의 지카바이러스 온톨로지 분류체계

| 생태학적 관점 | 대분류 | 중분류 |
|--------------|-----------------------------------|-----------|
| Microsystem | 개인 수준의 인구학적 특성 개인의 주관적 인지 및 감정 | 대상(일반 시민) |
| | | 성별/나이 |
| Mesosystem | 감염 관련 정보 | 긍정(감정) |
| | | 부정(감정) |
| | | 바이러스(질병) |
| | | 감염 경로 |
| Exosystem | 사회적 관련 요소 | 증상 |
| | | 대처(치료) |
| | | 예방 |
| | | 병원 |
| Chronosystem | 국가적 관련 요소 | 산업/경제 |
| | | 국가 |
| | 시간 요인 | 기타 |
| | | 계절 |

미시체계에 속하는 개인 수준의 인구학적 특성은 대상 305개, 성별/나이 31개, 개인의 주관적 인지 및 긍정(감정) 194개, 부정(감정) 268개를 포함하여 총 798개의 용어가 배치되었다. 중간체계에 속하는 감염 관련 정보는 증상 151개, 바이러스 91개, 감염 경로 98개, 대처(치료) 207개, 예방 182개로 구성되었다. 중분류 중 대처(치료)와 예방의 경우는 격리, 백신, 예방수칙 등과 같이 조직 수준에서도 개입이 가능한 개념을 포함하며, 감염 경로 또한 단순 매개감염뿐만 아니라 SIS 모형 및 SEIR 모형과 같은 감염병 질환 확산 모델에 의하여 사람들 간의 감염병 전파가 설명되므로 감염 관련 정보는 중간체계의 개념이라고 볼 수 있다. 중간체계는 총 729개의 용어로 추출되었다.

외체계에 속하는 사회적 관련 요소는 병원 247개, 산업/경제 260개, 기관(장소/직책) 321개, 경보 수준 4개로 총 832개, 거시체계에 속하는 국가적 관련 요소는 지카바이러스 발생 국가 51개와 국가 199개로 총 250개를 포함하였다. 시간체계에 속하는 시간 요인은 계절을 포함하며 18개의 용어를 추출하였다. 그 밖에 생태학적 체계 이론의 범주에 포함시키기 부적절하다고 간주되는 용어 141개는 기타로 분류하였다.

본 연구에서는 생태학적 체계 이론을 통해 지카바이러스에 관한 요소를 개인 수준의 인구학적 특성, 개인의 주관적 인지 및 감정, 감염 관련 정보, 사회적 관련 요소, 국가적 관련 요소, 시간 요인 및 기타로 나누어 살펴보았다. 이를 통해 개발된 지카바이러스 관련 온톨로지와 분류체계는 SNS에서의 감염병 감정 위험을 예측하여 추후 감염병 위기관리 방안을 마련하는 데 활용될 수 있을 것이다.

2. 지카바이러스 정부 대응 방안

본 지카바이러스 발생에 대한 분석을 기반으로 정부는 향후 신종 전염병 질환이 발생할 때 신속하고 체계적으로 대응해야 한다. 구체적으로 정부가 수립할 수 있는 대응 방안을 제시하면 다음과 같다.

첫째, SNS상의 무분별한 허위 정보 유포를 방지하고 사용자가 올바른 정보를 제공받고 공유할 수 있는 체계를 구축해야 한다. SNS는 파급력이 뛰어난 매체 중 하나이며 이는 정보의 질과 무관하게 적용된다. 바이러스와 같은 질병은 해당 질병에 대한 부정적인 정보 유포와 낙인화의 충격을 초래할 수도 있으며, 이로 인해 사용자들은 잘못 조성된 정보로 인한 불안 속에서 위협에 대한 해결 방안을 찾지 못할 수 있다. 따라서 정부에서는 SNS 사용자들에게 공중보건 문제에 대한 올바른 정보를 제공할 수 있는 방안을 마련함과 동시에 이 정보들이 확산될 수 있고 잘못된 정보의 전파를 방지할 수 있는 체계를 구축해야 한다.

둘째, 지카바이러스와 같은 고위험 신종 전염병이 출현할 경우 SNS 빅데이터 분석을 통한 정부 차원의 위기관리체계 구축을 통해 치밀하고 신속한 대응이 이루어져야 한다. 고위험 신종 전염병에 대한 불안한 감정이 SNS상에서 확산되면 사회 전반에 그 악영향이 미칠 수 있다. 따라서 빅데이터 분석을 통해 SNS 사용자들이 보이는 불안한 감정을 감지 및 예측할 수 있는 질병관리 모니터링체계를 구축해 지나친 우려에 대해선 초기에 대처할 수 있어야 한다.

셋째, SNS는 위기 커뮤니케이션으로서 중요한 역할을 할 수 있지만 메르스나 에볼라바이러스, 지카바이러스 등의 고위험 신종 전염병에 대한 정보가 다뤄질 경우 미디어상에서 해당 정보가 반복적으로 조명됨으로써 대중의 공포심을 자극하게 될 수도 있다. 이는 정부 혹은 관련 조직에 대

한 불신 등의 형태로 전이될 수 있으며 다른 요소들과 겹쳐 위험이 확산될 수 있으므로 위험에 대한 올바른 정보와 대책을 사용자에게 제공하고 공유하도록 하는 체계가 구축되어야 할 것이다.

3. 감염병 위기 정보 확산 리스크 커뮤니케이션 구축 방안

감염병 발생 시 수집된 소셜 빅데이터를 활용하여 감염병 위기 대응센터를 통해 위기 단계에 따른 대응 및 시나리오별 대응 조치 방안을 마련해야 한다. 소셜네트워크는 즉각적인 정보의 전달과 사람들 간의 상호작용이 가능한 커뮤니케이션 수단이다. 이에 따라 감염병 위기 정보의 지속적인 모니터링을 통하여 정보를 관리해야 한다. 또한 소셜네트워크상에서 국민들의 공포의식을 증대시키는 요인을 파악하고 분석한 결과를 바탕으로 신종 감염병에 대한 여론 동향을 예측해 올바른 정보를 제공해야 한다.

참고문헌 <<

- 김민정, 윤영민. (2015). 복합매체 이용과 사회적 동조가 온라인상에서 루머신뢰 및 루머전파의도에 미치는 영향, 한국PR학회, 19(2), pp.96-119.
- 김성순, 이상원, 최보율. (2008). 신종인플루엔자 유행관리를 위한 수학적 모형과 전략에 대한 고찰. 한국역학회지, 30(2), pp.156-167.
- 김태형, 간형식. (2014). 루머의 특성과 확산에 관한 이론적 고찰. 글로벌 경영연구, 26(1), pp.1-18.
- 동아일보. 지카 바이러스 한국인 감염자 첫 발생, SNS 괴담 진실은?, <http://news.donga.com/BestClick/3/all/20160322/77140871/2>에서 2016. 3. 22. 인출.
- 이윤희. (2014). 국내 SNS 이용현황과 주요 이슈 분석, Internet & Security, 8월호.
- 이원태, 차미영, 양해륜. (2011). 소셜미디어 유력자의 네트워크 특성: 한국의 트위터 공동체를 중심으로. 언론정보연구, 48(2), pp.44-79.
- 연합뉴스, 지카 바이러스 의심 신고 87건, 1주일 새 4.8배로 늘어, <http://www.yonhapnews.co.kr/bulletin/2016/05/05/0200000000AKR20160505026000017.HTML>에서 2016. 5. 5. 인출.
- 신호성, 김동진. (2008). 기후변화와 전염병 질병부담. 한국보건사회연구원.
- 장덕진, 김기훈. (2011). 한국인 트위터 네트워크의 구조와 동학. 언론정보연구, 48(1), pp.59-86.
- 질병관리본부, 보도자료. 2016.02.02.
- 질병관리본부. 지카 바이러스 발생국가 현황. <http://www.cdc.go.kr/CDC/contents/CdcKrContentView.jsp?cid=66993&menuIds=HOME001-MNU2374-MNU2365-MNU2367>에서 2016. 2. 22. 인출.
- 질병관리본부. 지카 바이러스 감염증. <http://www.cdc.go.kr/CDC/contents/CdcKrContentView.jsp?cid=66993&menuIds=HOME001-MNU2374-MNU2365-MNU2366>.에

서 2016. 2. 22. 인출.

Camila Zanluca, Claudia Nunes Duarte dos Santos. (2016). *Zika virus - an overview*. Microbes and Infection, pp.1-7.

Centers for Disease Control and Prevention. *About Zika Virus Disease*.
<http://www.cdc.gov/zika/about/index.html>에서 2016. 2. 22. 인출.

Hong, Ju-Hyun. (2014). *A crisis of confidence and the media*, Newspaper and broadcast 10, pp.15-20.

Ibrar Yaqoob, Victor Chang, Abdullah Gani, Salimah Mokhtar, Ibrahim Abaker Targio Hashem, Ejaz Ahmed, Nor Badrul Anuar, Samee U. Khan. (2016). *Information Fusion in Social Big Data: Foundations, State-of-the-art, Applications, Challenges, and Future Research Directions*. International Journal of Information Management.

Isaac Chun-Hai Fung, Zion Tsz Ho Tse, King-Wa Fu. (2015). *The use of Social media in public health surveillance*, Western Pacific Surveillance and Response Journal. 6(2), pp.3-6.

Juacong Chen, Huiling Zhang, Zhi-hong Guan, Tao Li. (2012). *Epidemic spreading on networks with overlapping community structure*. Physica A: Statistical Mechanics and its Application, 39(4), pp.1848-1854.

Ling-Ling Xia, Guo-Ping Jiang, Bo Song, Yu-Rong Song. (2015). *Rumor spreading model considering hesitating mechanism in complex social networks*. Physica A 437, pp.295-303.

Mathematical association of America, *The SIR Model for Spread of Disease - The differential Equation Model*.

<http://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model>에서 2016. 5. 4. 인출.

Michael B Prentice, Lila Rabalison. (2007). *Plague*. Lancet, 369, pp.

1196-1207.

Rodolphe Hamel, Florian Lefebvre, Sineewanlaya Wichit, Julien Pompon, Fodje Diop, Loïc Talignani, Frédéric Thomas, Philippe Despres, Hans Yssel, Dorothee Missre. (2016). *Zika virus: epidemiology, clinical features and host-virus interactions*. *Microbes and Infection*, pp.1-9.

Sebastian Funk, Marcel Salathe, Vincent A. A. Janesen. (2010). *Modeling the influence of human behaviour on the spread of infectious disease: a review*. *Journal of the Royal Society interface*, 7, pp.1247-1256.

The New York Times, Zika Virus a Global Health Emergency, W.H.O. Says 1. Feb, 2016.

http://www.nytimes.com/2016/02/02/health/zika-virus-world-health-organization.html?_r=0에서 2016. 2. 22. 인출.

Woohyun Yoo, Doo-Hun Choi, Keeho Park. (2016). The effects of SNS communication: How expressing and receiving information predict MERS-preventive behaviour intentions in south Korea.

World Health Organization, Zika Virus, Fact sheet, Updated February 2016, <http://www.who.int/mediacentre/factsheets/zika/en/>에서 2016. 2. 22. 인출.

World Health Organization. Frequently asked question on IHR Emergency Committee.

World Health Organization, Situation report, Zika Virus Microcephaly Guillain-Barre Syndrome. Update 12 May 2016.

World Health Organization. What social media offers to health professionals citizens. Updated August, 2009, <http://www.who.int/bulletin/volumers/87/8/09-006712/en/>에서 2016. 4. 27. 인출.

World Health Organization. WHO statement on the first meeting of the International Health Regulations (2005) (IHR 2005) Emergency Committee on Zika virus and observed increase in neurological disorders and neonatal malformations. <http://www.who.int/mediacentre/news/statements/2016/1st-emergency-committee-zika/en/>에서 2016. 2. 22. 인출.

Zika Virus a trending topic on social media platform. <http://www.straitstimes.com/world/united-states/zika-virus-trending-topic-on-social-media-platforms>에서 2016. 2. 22. 인출.

제 5 장

어린이 예방접종 소셜 빅데이터 분류를 위한 온톨로지 개발

제1절 서론

제2절 어린이 예방접종 온톨로지 개발

제3절 연구 결과

제4절 논의

제5절 결론

참고문헌



5

어린이 예방접종 소셜 빅데이터 << 분류를 위한 온톨로지 개발²²⁾

제1절 서론

전 세계적으로 인류를 위협하는 다양한 감염병이 끊임없이 발생하고 있는데 우리나라도 예외가 아니다. 최근 들어 사스, 에볼라, 메르스 등과 같은 신종 감염병뿐만 아니라 수두, 홍역, 백일해 등과 같이 이미 퇴치되었다고 생각되었던 감염병도 지속적으로 다시 유행하고 있는 상황이다. 이러한 감염병으로부터 건강을 보호하기 위해 예방접종의 중요성이 강조되고 있다. 예방접종은 미생물의 병원성을 죽이거나 약하게 하여 만든 백신을 사람에게 투여해 능동면역이 생기는 것을 유도하여 감염질환을 예방할 수 있는 효과적인 수단으로 한 개인의 질병 예방뿐만 아니라 사회집단 전체로 감염이 전파되는 것을 막는 집단면역의 효과까지 얻을 수 있다(질병관리본부, 2013).

2013년 전국 예방접종률 조사에 따르면 우리나라의 3세 이상 아동의 예방접종률은 95% 이상으로 세계보건기구(WHO)에서 정한 감염병 퇴치 수준에 이를 정도로 높다. 그러나 이렇게 높은 예방접종률에도 불구하고 최근 홍역, 유행성이하선염 등의 예방접종 대상 감염병이 간헐적으로 유행하면서 예방접종의 효과나 필요성을 의심하는 사람이 생겨나고 있고, 예방접종 후 부작용이나 사망과 같은 백신 안전성 관련 문제가 발생하면서 온라인상에 자녀의 예방접종에 대한 부모들의 걱정이나 우려, 예방접

22) 본 연구는 해외 학술지에 게재하기 위하여 서울대학교 간호대학(박현애, 온정아)과 한국 보건사회연구원(송태민)에서 공동으로 작성한 논문임을 밝힌다.

종 경험 등의 글이 늘어나고 있다.

예방접종에 대한 걱정이나 우려와 같은 부정적인 감정, 예방접종의 효과나 유의성과 같은 긍정적인 감정은 예방접종 의도에 영향을 주며 나아가 예방접종률의 감소나 증가를 초래한다(임은실 외, 2006; 김금순 외, 2007). 비록 우리나라 예방접종률이 감염병 퇴치 수준까지 도달하였지만 온라인에 게시되는 예방접종에 대한 걱정이나 우려의 글이 늘어나면 예방접종 의도에 영향을 미칠 수 있다. 따라서 예방접종에 영향을 줄 수 있는, 예방접종에 대한 부모들의 걱정이나 우려 등을 정확히 파악할 필요가 있다.

본 연구팀이 사전 조사에서 온라인에 게시된 예방접종 관련 감정이나 경험 등을 표현한 글을 살펴본 결과 예방접종에 대한 긍정적인 글과 부정적인 글 모두를 확인할 수 있었다. 이들 게시 글을 자세히 살펴보면 예방접종 후 발생하는 부작용과 이에 대처하는 방법, 예방접종 날짜의 조정, 예방접종 여부 고민과 같은 다양한 주제에 대한 개인들의 감정, 경험, 태도와 예방접종 정보 등이 게시되고 있음을 알 수 있었다.

예방접종에 대한 감정, 태도, 경험, 지식과 관련한 국내의 연구 중 예방접종 실태와 부모들의 예방접종에 대한 인식을 조사한 연구(임정우 외, 2006)에 의하면 반드시 시행해야 할 예방접종은 BCG, MMR, DTP, 소아마비, B형간염이라고 인식하였고 비용이 비싼 예방접종이 더 안전할 것이라고 생각하는 것으로 나타났다. 또한 제때 예방접종을 하지 못한 이유는 자녀가 아파서, 접종 날짜를 잘 몰라서, 집안일 때문인 것으로 나타났다. 12세 이하 자녀를 둔 보호자를 대상으로 방문 면접 조사 방식으로 시행한 질병관리본부 주관의 2014년 국가예방접종 지원 정책 만족도 및 보호자 인식조사(고재영 외, 2015)에 의하면 대부분의 부모는 예방접종을 하지 않으면 감염병에 걸릴 가능성이 커진다고 생각하고 있으며, 예방

접종 대상 감염병을 심각하게 여기고, 예방접종이 건강 증진에 도움을 준다는 긍정적인 인식을 가지고 있는 것으로 나타났다. 앞에 소개한 연구들은 온라인 및 방문 설문조사를 통해 예방접종에 대한 인식을 파악하여 조사 시간, 비용, 노력이 많이 소요되고 사람들의 인식을 실시간 파악할 수 없다는 단점이 있다. 또한 이러한 연구 방법은 감염병 유행이나 새로운 백신 도입과 같은 특정 사건이 발생한 시점에서 사람들의 생각을 실시간으로 파악하는 데 한계가 있다.

예방접종에 대한 부모들의 감정, 태도, 경험, 지식들을 좀 더 빠르게 파악하기 위해 소셜 데이터를 활용한 연구가 국외에서 소개되고 있다. 예를 들어 미국에서 2009년 신종플루가 유행한 시기에 신종플루 예방접종과 관련한 트위터 메시지를 수집하여 예방접종에 대한 감정을 분석한 연구(Salathé et al., 2011)에서는 트위터에 표현된 예방접종에 대한 감정과 예방접종을 간에 밀접한 관계가 있는 것으로 나타났고, 소셜 데이터가 예방접종률을 예측할 수 있는 효과적인 수단임을 보여 주었다. 또한 유엔아동기금(UNICEF, 2013)에서 유럽에서의 대중의 예방접종에 대한 반대 감정을 파악하기 위해 블로그, 포럼, 트위터, 텀블러, 유튜브, 페이스북과 같은 소셜미디어에 게시된 글을 수집하여 분석한 결과, 예방접종 반대 감정은 종교적·도덕적 신념, 이상반응, 장애 발생, 화학물질/독성/불필요성, 음모이론 등에 기인하는 것으로 나타났다. 이 연구에서는 예방접종을 반대하는 감정에 기여하는 영역을 부작용에 대한 우려나 두려움과 같은 개인 영역, 정부나 제약 산업에 대한 불신과 같은 상황적 영역, 예방접종이 필요 없다는 종교적인 신념과 같은 초월적 영역으로 구분하였고, 이들 세 영역의 조합에 따른 각기 다른 예방접종 캠페인 방법을 제시하였다. 또 다른 예로 UN global pulse(2014)에서 인도네시아 대중이 소셜미디어인 트위터에 게시한 예방접종 관련 글을 분석하여 예방접종에 대한 중

교적인 입장, 질병 발생 관련 대화, 백신 이상반응으로 여겨지는 증상 혹은 건강 상태, 오가 백신인 새로운 백신 도입에 대한 글이 많이 게시되는 것을 파악하였고, 게시 글의 추이를 분석해 대중이 예방접종 관련 우려를 언제 가장 많이 하는지 그 시점과 우려의 이유를 파악할 수 있었다.

소셜미디어에 게시된 글을 이용한 이들 연구에서 문자로 구성된 비정형 데이터를 분석하기 위해 수작업 분석 방법 혹은 자연어 처리 시스템을 활용하였다. 이러한 분석 방법은 주로 게시된 글에 포함된 용어의 빈도수에 기반을 두어 분석 과정에서 용어의 의미론적 관계를 명확히 파악하지 못한다는 한계를 가진다. 이러한 한계를 극복하기 위해 동음이의어(homonym)와 이음동의어(synonym)를 설정한 유의어 사전을 개발하여 이용하고 있으나 그 용어가 사용된 문맥을 이해하지 못하면 용어의 정확한 의미를 파악하기 어렵다(유은지 외, 2012). 소셜미디어에 게시된 비정형 데이터의 의미적 모호성을 해결하고 용어의 의미를 정확하게 파악하기 위해서는 개념의 의미와 개념의 고유한 속성, 개념들 간의 관계 및 제약 조건 등을 정의하고 유의어/동의어를 망라한 용어체계가 포함된 온톨로지가 필요하다. 온톨로지는 공유된 개념화에 대한 정형화되고 명시적인 명세(Gruber, 1993)이며, 어휘의 내용을 명확하게 정의하고 어휘들로 표현된 사실들 사이의 논리적 관계를 나타내는 일종의 지식 표현(knowledge representation)이다(김학래, 2010). 따라서 소셜미디어에 게시된 글을 이용하여 자녀의 예방접종에 대한 부모의 감정, 태도, 경험, 지식을 정확히 파악하기 위해서는 온톨로지를 개발할 필요가 있다.

이에 본 연구에서는 부모가 자녀의 예방접종을 결정하는 과정에 영향을 주는 감정, 경험, 태도, 지식과 예방접종 경험 및 예방접종 후의 감정, 태도 등을 포함한 온톨로지를 개발하고자 한다. 본 연구에서 개발된 온톨로지는 부모가 자녀의 예방접종에 대해 어떤 감정, 경험, 태도, 지식을 가

지고 있는지를 파악하는 분석의 틀로 활용될 수 있을 것이다. 또한 온톨로지 개발 과정에서 비정형 소셜 데이터의 빅데이터 분석에 활용할 수 있는 주요 용어들의 유의어/동의를 기술한 용어체계를 개발하고자 한다.

제2절 어린이 예방접종 온톨로지 개발

1. 연구 목표

연구 목표는 다음과 같다.

첫째, 어린이 예방접종에 대한 부모의 감정, 경험, 태도, 지식과 관련된 용어를 수집하고, 용어로부터 관련 개념을 추출하여 추출된 개념의 데이터 모델을 개발한다.

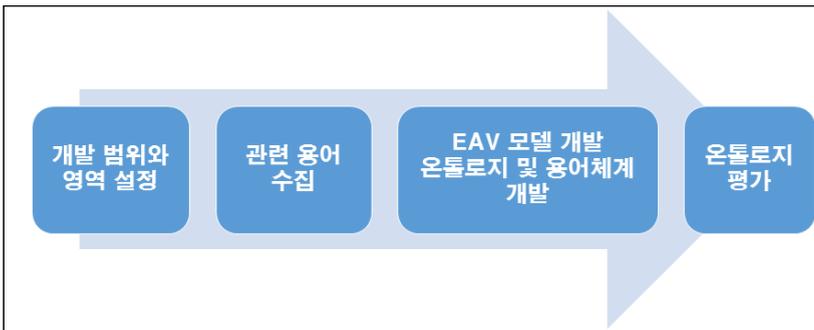
둘째, 어린이 예방접종에 대한 부모의 감정, 경험, 태도, 지식 관련 영역과 영역 간의 관계를 나타내는 온톨로지와 유의어/동의를 포함한 용어체계를 개발한다.

2. 연구 방법 및 내용

어린이 예방접종에 대한 부모의 감정, 경험, 태도, 지식 관련 온톨로지를 Ontology development 101(Noy & McGuinness, 2001)을 참조하여 구축할 예정이며 이 과정에서 온톨로지에 포함된 개념들의 유의어/동의를 포함한 용어체계를 개발할 예정이다. 연구의 진행 과정은 [그림 5-1]과 같다. 우선 개발할 어린이 예방접종 온톨로지의 범위와 영역을 설정하고, 관련 문헌을 통해 용어를 수집한다. 수집된 용어로부터 개념을 추출

하여 추출된 개념의 특성을 파악하는 EAV 모델(Entity-Attribute-Value model)을 개발하고, 관련 개념들의 관계와 유의어/동의어를 포함하는 온톨로지 및 용어체계를 개발할 예정이다. 마지막으로 개발된 온톨로지가 계획 목적에 맞게 개발되었는지 평가할 예정이다.

[그림 5-1] 어린이 예방접종 온톨로지 개발 과정



가. 어린이 예방접종 온톨로지의 개발 범위와 영역 설정

1) 예방접종의 대상과 범위

예방접종은 영유아에서부터 노인에 이르기까지 다양한 연령을 대상으로 시행되고 있으나 그중에서 면역력이 약한 어린이를 대상으로 한 예방접종은 기초 면역을 형성하는 데 중요하여 다른 연령보다 많은 항목의 예방접종이 요구된다. 또한 우리나라 정부에서는 어린이 예방접종 비용 지원 사업, 취약아동 예방접종 확인 사업과 같은 정책을 시행하며 어린이 예방접종을 독려하기 위해 애쓰고 있다. 따라서 본 연구에서는 예방접종의 범위를 0세에서 12세에 시행되는 어린이 예방접종으로 한정하였으며 여기에 포함되는 세부 예방접종으로는 결핵 예방접종, B형간염 예방접

중, 디프테리아/파상풍/백일해 예방접종, 폴리오 예방접종, 홍역/볼거리/풍진 예방접종, 일본뇌염 예방접종, 수두 예방접종, b형 헤모필루스 인플루엔자 예방접종, 폐렴구균 예방접종, A형간염 예방접종, 인플루엔자 예방접종, 로타바이러스 예방접종, 인유두종바이러스 예방접종이 있다.

어린이 예방접종은 다른 연령층의 예방접종과 달리 육아를 담당하는 부모의 예방접종에 대한 감정, 경험, 태도, 지식에 의해 예방접종 여부가 결정된다. 부모는 자녀의 건강에 관심이 많으며 건강에 영향을 주는 예방접종과 관련된 이슈에 대해서도 민감하게 반응하고, 소셜미디어를 통해 예방접종에 대한 감정, 경험, 태도 등을 표현한다. 본 연구에서 개발할 어린이 예방접종 온톨로지는 부모의 관점에서 어린이 예방접종에 대한 감정, 태도, 경험, 지식을 파악할 필요가 있다.

자녀의 예방접종에 대한 부모의 감정, 태도, 경험, 지식은 예방접종을 할 때뿐만 아니라 예방접종 여부를 결정하고 예방접종을 하고 난 후의 모든 과정에 영향을 미친다(Katz et al., 2010). 본 연구에서 개발할 온톨로지는 예방접종 이전부터 예방접종을 할 때, 예방접종을 하고 난 후에 이르는 전체 과정에서의 부모의 감정, 경험, 태도, 지식을 파악할 필요가 있다.

예방접종과 관련된 개인의 감정, 태도, 경험, 지식 등은 주변의 다양한 요인에 영향을 받는다. 부모의 개인적 경험이나 생각, 지식뿐 아니라 사회지리적인 특성, 예방접종과 관련된 의료기관, 예방접종에 대한 사회적 인식, 국가 정책, TV나 신문을 통한 보도와 같은 요인들이 예방접종에 영향을 주는 요인으로 알려져 있다(Larson et al., 2014). 본 연구에서 개발할 어린이 예방접종 온톨로지는 개인적 측면에 국한하지 않고 구조적·사회적·문화적·환경적 요인들을 모두 포함할 필요가 있다.

따라서 본 연구에서 개발할 어린이 예방접종 온톨로지는 0세에서 12세에 시행하는 예방접종을 대상으로 부모 관점에서 예방접종에 대한 감

정, 경험, 태도, 지식을 파악하며 예방접종 이전 단계에서부터 예방접종 시점, 예방접종 후에 이르는 예방접종 전 과정에서 개인적·구조적·사회적·문화적·환경적 영역을 포함할 필요가 있다.

2) Competency questions

어린이 예방접종 온톨로지의 영역과 내용을 좀 더 명확히 설정하고자 Competency questions(Grüniger & Fox, 1995)를 작성하였다. Competency questions는 온톨로지 개발 시 영역과 범위를 구체적으로 정하는 데 도움을 주며, 온톨로지 개발 후에는 원래 목적에 맞는 내용이 충분히 포함되었는지를 평가하는 데 이용된다. 연구를 시작하기에 앞서 본 연구팀이 온라인에 표현된 예방접종 관련 감정 및 경험을 조사한 내용을 참고로 Competency questions를 작성하였다. 작성된 Competency questions는 <표 5-1>과 같다.

<표 5-1> Competency questions

| 범위 | Competency questions |
|----------------|--------------------------------------|
| 개인적 측면 | 예방접종을 결정하는 데 영향을 주는 요인은 무엇인가? |
| | 예방접종에 대해 어떤 감정을 가지고 있는가? |
| | 예방접종을 하는 데 고려하는 자녀의 건강 상태는 무엇인가? |
| | 예방접종에 대한 지식 수준은 어떠한가? |
| 예방접종에 대한 정보 | 어린이가 맞아야 하는 예방접종은 무엇인가? |
| | 결핵 예방접종 후 발생 가능한 이상반응은 무엇인가? |
| | 만 1세인 자녀는 어떤 예방접종을 해야 하는가? |
| | 예방접종에 대한 정보는 어떤 매체를 통해 얻는가? |
| 구조적/사회적/문화적 요인 | 예방접종에 영향을 주는 자녀와 관련된 요인은? |
| | 예방접종에 영향을 주는 과거 예방접종 관련 경험은 무엇인가? |
| | 예방접종 비용에 대해 어떻게 생각하는가? |
| 예방접종 후 경험 | 디프테리아/파상풍/백일해 예방접종 후 경험한 이상반응은 무엇인가? |

| 범위 | Competency questions |
|--------|----------------------------|
| | 예방접종 후 발열 시 어떻게 해야 하는가? |
| | 예방접종을 하러 간 의료기관에 대해 만족하는가? |
| | 의료진은 친절한가? |
| 환경적 요인 | 정부의 예방접종 관련 정책은 무엇이 있는가? |
| | 예방접종 관련 미디어의 성향은 어떠한가? |

3) 기개발된 온톨로지 확인

본 연구에서 예방접종 관련 감정, 경험, 태도, 지식을 확인하는 온톨로지를 개발하고자 기개발된 예방접종 관련 온톨로지를 검색하였으나, 백신 자체에 대한 온톨로지(He et al., 2009) 외에 예방접종 행위나 인식 및 태도와 관련된 온톨로지는 찾을 수 없었다. 예방접종에 대한 감정, 경험, 태도, 지식과 관련된 요인들에 대한 연구 개념 틀을 검색하여 이를 온톨로지 개발에 참조하였다.

본 연구에서 고찰한 연구 개념 틀은 Sturm et al.(2005)의 ‘Factors influencing parental decision making about childhood immunization’과 Larson et al.(2014)의 ‘Model of determinants of vaccine hesitancy’와 Katz et al.(2010)의 ‘The vaccine perception, accountability and adherence model’이다.

Sturm et al.(2005)의 연구에서 사용된 개념 틀에서는 예방접종 결정에 영향을 주는 요인으로 개인적/부모 요인, 사회 환경적 요인, 제도적 요인, 의료서비스 환경 요인, 물리적 환경 요인을 제시하고 각 요인들 간의 관계를 표현하였다. 개인/부모의 건강신념, 건강에 대한 태도 등이 예방접종을 결정하며 사회 환경적 요인, 제도적 요인, 의료서비스 환경 요인이 개인/부모 요인에 영향을 미치는 것으로 나타났다. 사회 환경적 요인은 예방접종에 대한 문화적 태도와 신념, 부모의 사회적 집단규범, 백신

에 대한 미디어의 보도 등을 포함하며, 제도적 요인은 국가 정책뿐 아니라 백신에 대한 사회적 집단이나 전문가 집단의 행동을 포함하고 있다. 의료서비스 환경 요인은 예방접종의 위험성과 장점에 대한 의사소통, 의료서비스 제공자의 태도나 행동을 포함하며, 물리적 환경 요인은 감염병의 유행, 감염병에 대한 효과적인 새로운 예방법이나 치료법 개발 등을 포함하고 있다.

Larson et al.(2014)의 연구에서 사용된 개념 틀에서는 예방접종을 결정하는 데 관련된 요인을 개인적/사회적 집단 영향, 백신과 예방접종 관련 이슈, 상황적 영향으로 구분하였다. 개인적/사회적 집단 영향은 건강에 대한 신념과 태도, 예방접종에 대한 지식이나 인식, 의료서비스 경험, 예방접종에 대한 사회적 규범 등을 포함하고 백신과 예방접종 관련 이슈는 백신의 과학적인 위험성과 이득, 예방접종 일정, 백신 공급의 신뢰성, 예방접종 비용 등으로 구성되어 있다. 상황적 영향은 지리적 장애, 미디어 환경, 정책 및 정치, 지역적/문화적 요소, 사회경제적 요인 등을 포함하고 있다. 세 가지 요인 간의 관계는 제시되지 않았다.

Katz et al.(2010)의 연구에서 사용된 개념 틀에서는 인유두종바이러스 예방접종에 대한 인식, 결정, 행위 유지에 영향을 미치는 요인으로 청소년 개인 요인, 부모와 같은 돌봄 제공자 요인, 구조적/사회 환경적 요인, 행동반응요인, 환경적 요인을 제시하였고 각 요인이 예방접종 과정에서 어떤 시점에 영향을 미치는지 표현하였다. 예방접종 시행 전에는 구조적/사회 환경적 요인이 영향을 주며 사회인구학적 요인, 사회지리적 요인, 과거 예방접종 관련 행동, 예방접종 이용 가능성과 비용이 포함된다. 예방접종 시점에는 청소년 개인 요인과 돌봄 제공자 요인이 영향을 주며 청소년 개인 요인에는 청소년의 성숙 발달, 자기 효능감, 지식, 돌봄 제공자와의 관계, 정신건강, 건강신념이 포함되고, 돌봄 제공자 요인에는 건

강신념, 돌봄 제공자와 청소년의 관계, 의료서비스 이용이 포함된다. 예방접종 후에는 행동반응요인이 영향을 주며 첫 번째 예방접종 후 반응, 첫 방문 시 의료서비스의 상호작용이 포함된다. 예방접종의 모든 시점에 환경적 요인이 영향을 주며 사회적 건강신념, 정부의 보도자료, 미디어가 포함된다. 세 가지 연구 개념 틀에서 제시한 예방접종과 관련된 요인들의 개념과 요인들 간의 관계를 본 연구에서 개발할 어린이 예방접종 온톨로지에 포함할 필요가 있다. <표 5-2>와 같이 각 연구 개념 틀에서 제시한 요인들을 비슷한 개념으로 분류하였고, 이를 이용하여 온톨로지의 세부 범위를 설정하고 관련 용어를 추출하였다.

<표 5-2> 연구 개념 틀에서 제시한 관련 요인들의 내용 분류

| 분류 | 내용 | 출처 |
|-------------|---|--------------------------------------|
| 개인 | 건강신념(지각된 위험성, 지각된 유익성, 지각된 민감성, 지각된 장애 요인, 자기효능감) | Sturm(2005), Larson(2014) Katz(2010) |
| | 건강에 대한 태도 | Sturm(2005), Larson(2014) |
| | 예방접종의 이상반응 발생 우려 | Sturm(2005) |
| | 예방접종에 대한 지식이나 인식 | Larson(2014), Katz(2010) |
| | 과거 예방접종 경험 | Larson(2014), Katz(2010) |
| | 의료서비스 이용 | Katz(2010) |
| | 예방접종 대상과 돌봄 제공자와의 관계 | Katz(2010) |
| | 자녀의 정신건강, 자녀의 성숙도 | Katz(2010) |
| 제도적 | 정책 | Sturm(2005), Katz(2010) |
| | 사회적 집단이나 전문가 집단의 행동 | Sturm(2005), Larson(2014) |
| 사회적 환경적 구조적 | 문화적 태도와 신념 | Sturm(2005), Larson(2014) |
| | 사회적 집단규범 | Sturm(2005), Larson(2014) Katz(2010) |
| | 미디어 | Sturm(2005), |

| 분류 | 내용 | 출처 |
|------------|--------------------------|------------------------------|
| | | Larson(2014) Katz(2010) |
| | 타인의 영향 | Larson(2014) |
| | 사회경제적 요소 | Larson(2014) |
| | 지리적 요소 | Larson(2014), Katz(2010) |
| | 종교적 요소 | Larson(2014) |
| | 성별 요소 | Larson(2014) |
| | 제약산업 | Larson(2014) |
| | 역사적 영향 | Larson(2014) |
| | 인구사회학적 요소 | Katz(2010) |
| | 백신 이용 가능성 | Katz(2010) |
| | 감염병의 유형 | Sturm(2005) |
| | 감염병에 대한 새로운 예방법이나 치료법 개발 | Sturm(2005) |
| 의료서비스 경험 | 예방접종의 위험성과 장점에 대한 의사소통 | Sturm(2005), Larson(2014) |
| | 예방접종 서비스 제공자의 태도와 행동 | Sturm(2005) |
| | 의료서비스와 제공자에 대한 개인적 경험 | Larson(2014), Katz(2010) |
| | 의료서비스와 제공자에 대한 신뢰 | Larson(2014) |
| | 예방접종 후 반응 | Katz(2010) |
| 예방접종 관련 정보 | 백신의 효과와 위험성 | Larson(2014) |
| | 예방접종 일정 | Larson(2014) |
| | 백신 관리 방법 | Larson(2014) |
| | 백신 전달 방법 | Larson(2014) |
| | 새로운 백신, 새로운 제형 도입 | Larson(2014) |
| | 백신 공급의 신뢰성 | Larson(2014) |
| | 예방접종 비용 | Larson(2014), Katz(2010) |

나. 관련 용어 수집

예방접종 실무지침과 어린이 예방접종 관련 연구논문 및 연구보고서를 고찰하여 어린이 예방접종에 대한 부모의 감정, 경험, 태도, 지식과 관련

된 인적요인, 구조사회문화요인, 정보요인, 환경요인, 행동반응요인에 포함된 용어를 수집하고, 마지막으로 SNS 검색을 통해 위의 자료원에 포함되지 않은 용어들을 수집하였다.

1) 예방접종 실무지침

본 연구에서는 우리나라 질병관리본부에서 제공한 ‘예방접종 대상 감염병의 역학과 관리’, 미국의 CDC(Centers for Disease Control)와 영국의 NHS(National Health Service)에서 제공하는 예방접종 실무지침을 고찰하였다. 예방접종 실무지침에는 예방접종의 원리, 일반 원칙, 예방접종 대상 감염병 등에 대한 설명이 포함된다. 예방접종의 원리 영역에서 백신의 분류 기준과 관련된 용어를 수집하였고 예방접종의 일반 원칙 영역에서 예방접종 시기와 간격, 예방접종 후 이상반응, 예방접종 금기 사항 및 주의 사항, 백신의 접종 방법과 관련된 용어를 수집하였다. 예방접종 대상 감염병 영역에서 결핵, B형간염, 디프테리아, 파상풍, 백일해, 폴리오 등과 같은 감염병의 원인균, 예방법과 관련된 용어를 수집하였다.

2) 예방접종 관련 요인에 대한 연구논문과 연구보고서

Pubmed와 Google scholar에서 ‘예방접종’, ‘영유아’, ‘어린이’, ‘부모’, ‘인식’, ‘태도’, ‘지식’, ‘vaccination’, ‘immunization’, ‘child’, ‘parent’, ‘perception’, ‘attitude’, ‘knowledge’ 등의 검색어로 확인된 101개의 국내외 논문 중 예방접종에 대한 감정, 경험, 태도, 지식과 관련된 논문을 고찰하였다.

예방접종에 대한 부모와 자녀의 인식 및 개인적 요인과 관련된 논문을

고찰하여 예방접종 필요성, 예방 효과, 예방접종의 해로움, 예방접종 부작용 발생, 감염병 위험성, 예방접종 불신 등과 같은 예방접종에 대한 신념(임은실 외, 2006; 김금순 외, 2007; 차혜경 외, 2012; 오현경, 2015)에 대한 용어, 부모의 지식 수준, 예방접종에 대한 지식을 얻는 지식원(정미은, 2007; 진선미, 이석구, 2007; 최인영 외, 2007)과 같은 예방접종 관련 지식에 대한 용어, 부모의 나이, 교육 수준, 직업, 자녀의 나이, 성별, 출생 순서, 가족 형태, 소득 수준 등의 인구사회적 요인(진선미, 이석구, 2007; 최윤경 외, 2008; 이무식 외, 2012; 이석구, 전소연, 2015)에 대한 용어가 수집되었다. 자녀의 질병, 장애 유무, 자녀의 건강에 대한 인식(이석구, 전소연, 2015; 박명배 외, 2013) 등 자녀의 건강과 관련된 용어, 외래 이용 횟수, 입원 여부, 건강검진 시행 여부(박명배 외, 2013) 등의 의료서비스 이용과 관련된 용어가 수집되었다.

예방접종에 영향을 미치는 구조사회문화요인과 관련된 논문을 고찰하여 사회적 건강신념, 주변 사람의 예방접종에 대한 태도, 예방접종 추천 유무(최인영 외, 2007; Brunson et al., 2012), 거주지(박명배 외, 2013) 같은 용어가 수집되었다.

예방접종과 관련한 환경적 요인에 대한 논문을 고찰하여 필수 예방접종, 예방접종 일정, 어린이 예방접종 지원 정책, 예방접종을 조사, 예방접종 국민 의식 조사, 예방접종 인식 조사, 예방접종 지원 정책 만족도 조사(이종구, 최원석, 2008; 고재영 외, 2015)와 같은 정책 관련 용어, TV·신문·라디오를 통한 미디어 보도, 보도 내용, 보도 성향(이현우, 2010)과 같은 미디어 보도 관련 용어, 감염병의 유행, 전파 정도, 심각성(sturm et al., 2005)과 같은 감염병 유행 관련 용어를 수집하였다.

3) 소셜미디어 검색

예방접종 실무지침과 예방접종 관련 연구논문에서 수집한 용어를 보완하기 위해 본 연구에서는 소셜미디어 검색을 시행하였다. 육아 전문 카페 중 하나인 '맘스홀릭'에서 2016년 3월 1일부터 2016년 4월 1일까지 게시된 글 중 '예방접종'을 키워드로 검색하여 추출된 총 285개의 예방접종 관련 게시 글을 분석하였다. 게시 글의 주제를 예방접종 후 발생한 부작용 및 대처 방안, 예방접종 날짜를 조정하는 방법, 예방접종 여부에 대한 고민, 두 가지 이상의 백신 상품명 중 하나를 선택하는 방법, 예방접종 전에 주의할 사항, 예방접종 기관에 대한 경험 및 의견, 예방접종 비용, 하루에 투여 가능한 예방접종 개수 등으로 분류한 후 예방접종과 관련된 개념을 추출하였다. 소셜미디어에서 추출한 예방접종 관련 게시 글의 주제 분류는 <표 5-3>과 같다.

<표 5-3> 예방접종 관련 게시 글의 주제 분류

| 주제 | 게시 글 개수 | 비율 |
|----------------------------|---------|-------|
| 예방접종 후 부작용과 대처 방안 | 89 | 31.2% |
| 예방접종 날짜 조정 방법 | 62 | 21.8% |
| 예방접종 여부에 대한 고민 | 21 | 7.4% |
| 두 가지 이상의 백신 상품명 중 하나 선택 방법 | 19 | 6.7% |
| 예방접종 전 주의사항 | 17 | 6.0% |
| 예방접종 기관 | 16 | 5.6% |
| 예방접종 비용 | 12 | 4.2% |
| 하루에 접종 투여 가능한 예방접종 개수 | 12 | 4.2% |
| 예방접종 정보 요청 | 11 | 3.9% |
| 해외 예방접종 정보 요청 | 5 | 1.8% |
| 예방접종 후 주의사항 | 4 | 1.4% |
| 예방접종 일정 알람 | 3 | 1.1% |
| 예방접종 후 인증 | 2 | 0.7% |

| 주제 | 게시 글 개수 | 비율 |
|---------|---------|--------|
| 예방접종 수첩 | 1 | 0.4% |
| 합계 | 285 | 100.0% |

이들 예방접종 관련 게시 글에서 추출한 용어를 word cloud로 표현하면 [그림 5-2]와 같다. 수집한 용어를 클래스, 속성, 속성값으로 분류하고 수집된 용어를 검색 키워드로 각 용어의 유의어/동의어를 수집하였다. 예를 들어 예방접종 종류의 하나인 ‘결핵 예방접종’이라는 용어를 클래스로 정의하였고 그 동의어로 ‘BCG’, ‘비씨지’, ‘불주사’라는 용어를 수집하였다.

[그림 5-2] 예방접종 관련 게시 글에 포함된 용어의 word cloud



다. 어린이 예방접종 온톨로지 개발

1) EAV 모델(Entity-Attribute-Value model)

앞 단계에서 수집한 어린이 예방접종 관련 용어에서 의미론적 개념을 추출하고, 동일하고 특징적인 성질을 지닌 개념들을 클래스로 생성하였다. 하나의 클래스 개념을 개체(entity)로 규정하고, 특징적인 성질들을 속성(attribute)으로 표현하며, 고유의 속성값(value)을 정의하여 EAV 모델(Entity-Attribute-Value model)을 개발하였다.

2) 온톨로지 및 용어체계 개발

어린이 예방접종과 관련된 클래스 개념을 일반적인 개념에서부터 시작하여 구체적인 세부 개념으로 계층화하여 배치하고, 연구 개념 틀을 기초로 최상위 클래스 사이의 관계를 정의하여 온톨로지를 개발하였다. 클래스, 속성, 속성값에 해당하는 개념의 유의어/동의를 수집하고 배열하여 용어체계를 개발하였다.

라. 온톨로지 평가

개발된 온톨로지가 개발 목적에 맞는 내용을 충분히 포함하고 있는지 앞 단계에서 개발한 Competency questions를 이용하여 평가하였다. Protégé 5.0 프로그램 내의 DL-Query를 활용하여 Competency questions를 query로 입력했을 때, 개발된 온톨로지가 Competency questions에 적절하게 답을 제시하는 정도로 평가하였다.

제3절 연구 결과

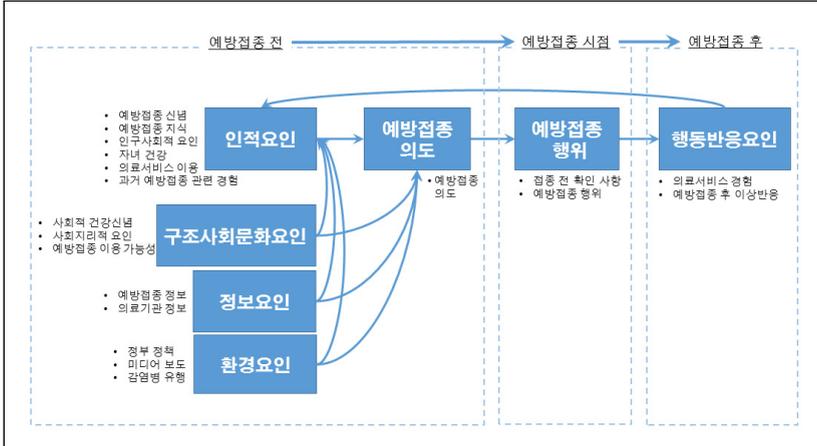
1. 어린이 예방접종 온톨로지

본 연구에서 개발한 어린이 예방접종에 대한 부모의 감정, 태도, 경험, 지식을 설명하는 온톨로지는 인적요인, 구조사회문화요인, 정보요인, 환경요인, 예방접종 의도, 예방접종 행위, 행동반응요인의 최상위 클래스로 구성된다. [그림 5-3]은 각 최상위 클래스 사이의 관계와 최상위 클래스에 포함되는 하부 클래스를 표현한 개념 틀을 제시하고 있다. [그림 5-3]에 제시된 어린이 예방접종 온톨로지 개념 틀에 따르면 자녀의 예방접종에 대한 부모의 감정, 태도, 경험, 지식은 예방접종 이전 단계에서부터 예방접종 시점과 예방접종 후까지의 모든 과정에 영향을 미치며 예방접종 후의 감정, 태도, 경험, 지식은 또다시 차기 예방접종에 영향을 미치는 반복적인 과정을 거친다. 어린이 예방접종에 대한 부모의 감정, 태도, 경험, 지식을 포함하고 있는 인적요인과 이에 영향을 미치는 구조사회문화요인, 정보요인, 환경요인이 예방접종 이전에 예방접종 의도에 영향을 미치고, 예방접종에 대한 의도는 예방접종 행위로 이어지며, 예방접종 후의 행동반응요인은 다시 인적요인에 영향을 미친다. 인적요인, 구조사회문화요인, 정보요인, 환경요인, 예방접종 의도, 예방접종 행위, 행동반응요인들에 대한 상세 설명과 참고한 문헌은 다음과 같다.

가. 인적요인

인적요인에는 예방접종 신념, 예방접종 지식, 인구사회적 요인, 자녀건강, 의료서비스 이용, 과거 예방접종 관련 경험이 포함된다.

[그림 5-3] 어린이 예방접종 온톨로지 개념 틀



예방접종에 대한 신념은 부모가 자녀의 예방접종에 대해 지니고 있는 생각과 감정으로, 건강신념모델을 이용하여 예방접종에 대한 부모의 인지를 확인한 여러 연구(임은실 외, 2006; 김금순 외, 2007; 차혜경 외 2012; 오현경, 2015)를 기초로 지각된 유익성, 지각된 심각성, 지각된 민감성, 지각된 장애 요인, 자기 효능감 개념을 추출하였다. 이들 개념을 설명하는 용어로 예방 효과, 감염병 위험성, 감염병 전파성, 예방접종 불신, 예방접종 복잡성 등을 포함하였다. 예방접종에 대한 부모의 지식은 예방접종을 결정하는 데 관련이 있는 지식의 수준, 그 지식의 정보원으로 구성된다(정미은, 2007; 진선미, 이석구, 2007; 최인영 외, 2007). 예방접종 지식 수준은 예방접종 대상 감염병, 예방접종 종류, 예방접종 시기, 예방접종 후 이상반응, 예방접종 주의 사항에 대해 부모가 스스로 인지하는 정도를 포함하였다. 예방접종에 대한 정보를 얻는 정보원으로는 정부에서 제공하는 예방접종 수첩, 예방접종 도우미 사이트 및 앱, 대중매체, 의료기관, 의료인, 가족 및 주변 이웃, 책을 포함하였다.

인구사회적 요인은 부모, 자녀, 가족의 인구사회적 요인으로 나뉘어 구

성된다. 예방접종의 영향요인을 파악한 연구(진선미, 이석구, 2007; 최윤경 외, 2008; 이무식 외, 2012; 이석구, 전소연, 2015)를 기초로 부모의 인구사회적 요인에 연령, 교육 수준, 직업을 포함하였고, 자녀의 인구사회적 요인에 성별, 연령, 출생 순서, 출생지, 주 양육자를 포함하였다. 가족의 사회적 요인에 자녀 수, 가족 형태, 혼인 형태, 소득 수준, 의료보장 형태를 포함하였다. 자녀의 건강은 예방접종을 결정하기 전 필수적으로 고려되는 요인으로 예방접종을 하지 않은 이유에 대한 조사에서 '질병을 앓고 있어서', '아이가 아파서'와 같은 아이의 건강 상태(이석구, 전소연, 2015)가 예방접종과 밀접한 관련이 있는 것으로 나타나 자녀의 건강에 자녀의 건강 상태와 건강에 대한 인식을 포함하였다. 건강 상태는 질병과 장애 여부 및 중증도로 구성하였고 건강에 대한 인식은 부모가 인지하는 자녀의 건강 상태와 정도로 구성하였다.

자녀 또는 가족 구성원의 평소 의료서비스 이용은 예방접종에 긍정적인 영향을 미친다는 연구(박명배 외, 2013) 결과에 따라 의료서비스 이용에 입원, 외래, 건강검진 이용을 포함하였다. 의료서비스 이용 대상을 자녀, 형제/자매, 부모로 나누었고 외래 이용은 최근 2주 이내를 기준으로 이용 여부와 횟수를, 입원은 최근 1년 이내를 기준으로 이용 여부와 일수를, 건강검진은 최근 2년 이내를 기준으로 시행 여부와 횟수를 포함하였다. 또한 부모가 각 의료서비스 이용 대상의 이용 정도에 대해 인식하는 정도도 포함하였다. 인적요인의 하위분류 및 용어체계는 <표 5-4>와 같다. 인적요인의 데이터모델링 예시는 <표 5-5>와 같다.

〈표 5-4〉 인적요인의 하위분류 및 용어체계

| 대분류 | 중분류 | 소분류 1 | 소분류 2 | |
|------|------------|--------------------|-----------------|-------|
| 인적요인 | 예방접종 신념 | 예방접종 필요성 | 필요성 | |
| | | 예방접종 유익성 | 예방 효과 | |
| | | | 감염병 유행 감소 | |
| | | 예방접종 유해성 | 예방접종 해로움 | |
| | | | 예방접종 부작용 발생 | |
| | | | 예방접종의 고통 유발 | |
| | | 예방접종 대상 감염병 심각성 | 감염병 위험성 | |
| | | | 감염병의 고통 유발 | |
| | | | 감염병 전파성 | |
| | | | 감염병 합병증 발생 | |
| | | 예방접종 장애 요인 | 예방접종 일정 잊어버림 | |
| | | | 예방접종 불신 | |
| | | | 바쁨 | |
| | | | 예방접종 복잡성 | |
| | | | 비용 부담 | |
| | | | 빠른 접종 시기 | |
| | | | 지정 의료기관과의 먼 거리 | |
| | | | 많은 예방접종 항목 | |
| | 예방접종 촉진 요인 | | 예방접종 알림서비스 제공 | |
| | | | 취학 시 접종 확인 | |
| | | 무접종 아이 기피 | | |
| | | 예방접종 수첩 사용 | | |
| | | 예방접종 일정 기억 | | |
| | | 예방접종 신뢰 | | |
| | | 예방접종 시간 여유 | | |
| | | 예방접종 간단함 | | |
| | | 예방접종 비용 지원 여부 | | |
| | | 적절한 접종 시기 | | |
| | 예방접종 지식 | 지식 수준 | 예방접종 대상 감염병 | |
| | | | 예방접종 종류 | |
| | | | 예방접종 시기 | |
| | | | 예방접종 이상반응 | |
| | | | 예방접종 주의 사항 | |
| 정보원 | | 정부 | | |
| | | 대중매체 | | |
| | | 의료기관 | | |
| | | 의료인 | | |
| | | 가족 및 주변 이웃 | | |
| | | 책 | | |
| | | 인구사회적 요인 | 부모의 인구사회적 요인 | 연령 |
| | | | | 교육 수준 |

| 대분류 | 중분류 | 소분류 1 | 소분류 2 | | | |
|---------------|---------------|-------------------|--------------------|------------|------------------|----|
| | | | 직업 | | | |
| | | | 성별 | | | |
| | | 자녀의 인구사회적 요인 | | 연령 | | |
| | | | | 출생 순서 | | |
| | | | | 출생지 | | |
| | | | | 주 양육자 | | |
| | | 가족의 사회적 요인 | | 자녀 수 | | |
| | | | | 가족 형태 | | |
| | | | | 혼인 상태 | | |
| | | | | 소득 수준 | | |
| | | | | 의료보장 형태 | | |
| | | | | 자녀건강 | 건강 상태 | 질병 |
| | | | | | | 장애 |
| | | | | | 건강 인식 | |
| | 허약하다 | | | | | |
| | 의료서비스 이용 | 이용 대상 | | 자녀 | | |
| | | | | 형제/자매 | | |
| | | | | 어머니 | | |
| | | | | 아버지 | | |
| | | | 이용 내용 | | 최근 2주 내 외래 이용 여부 | |
| 최근 1년 내 입원 여부 | | | | | | |
| | | | 최근 2년 내 건강검진 시행 여부 | | | |
| | | | 의료서비스 이용에 대한 인식 | | | |
| 과거 예방접종 관련 경험 | 과거 예방접종 시행 경험 | | 경험 대상 | | | |
| | | | 예방접종 경험 여부 | | | |
| | | 과거 예방접종 대상 감염병 경험 | | 이상반응 발생 여부 | | |
| | | | | 감염병 경험 대상 | | |
| | | | 감염병 종류 | | | |

〈표 5-5〉 인적요인의 데이터모델링 예시

| ENTITY | ATTRIBUTE | VALUE |
|--------------|-----------|--|
| 예방접종 필요성 | 정도 | 반드시 해야 한다/해야 한다/보통이다/하지 않아야 한다/전혀 하지 않아야 한다 |
| 예방 효과 | 정도 | 매우 효과 있다/효과 있다/보통이다/효과 없다/전혀 효과 없다 |
| 자녀의 인구사회적 요인 | 성별 | 남/여 |
| | 연령 | 0개월/1개월/2개월/4개월/6개월/12개월/15개월/18개월/24개월/36개월/만 4세/만 6세/만 11세/만 12세 |
| | 출생 순서 | 첫째/둘째/셋째 이상 |
| | 출생지 | 가정분만/조산원/병원 |

| ENTITY | ATTRIBUTE | VALUE |
|------------|-----------|------------------------------|
| 가족의 사회적 요인 | 주 양육자 | 부모/조부모/친척, 보모/보육시설 |
| | 자녀 수 | 독자녀/다자녀 |
| | 가족 형태 | 조손가족/한부모가족/일반가족/다문화가족/장애인 가족 |
| | 혼인 상태 | 미혼/동거/기혼/이혼/별거/사별 |
| | 소득 수준 | 저소득층/중산층/고소득층 |
| | 의료보장 형태 | 건강보험/의료급여 |

나. 구조사회문화요인

구조사회문화요인에는 사회적 건강신념, 사회지리적 요인, 예방접종 이용 가능성이 포함된다. 사회적 건강신념은 개인이 아닌 대중의 예방접종에 대한 일반적인 감정과 태도로, 예방접종을 결정하는 데 중요한 역할을 하는 것으로 나타났다(최인영, 2007; Brunson, 2012). 사회적 건강신념은 대중과 부모, 주변 사람의 신념으로 구성되는데 사회적 신념의 내용은 예방접종 필요성, 예방접종 효과, 예방접종 유해성, 예방접종 부작용 발생, 예방접종 신뢰, 예방접종 복잡성으로 구성된다. 주변 사람의 신념에는 사회적 신념에 포함된 내용에 더해 예방접종 추천 여부가 포함된다. 사회지리적 요인으로 거주지, 예방접종 기관의 거리에 대한 내용을 포함하였다. 거주지에 따라 예방접종에 대한 접근성 및 태도가 달라지므로(박명배 외, 2013; 이석구, 전소연, 2015) 거주지를 국내외와 도시 규모에 따라 대도시, 중소도시, 군 지역으로 구분하였다. 예방접종 기관과의 거리는 이동하는 데 소요되는 시간과 이동 거리를 포함하였다.

예방접종 이용 가능성에는 백신 이용 가능성, 경제적 접근성, 국가 예방접종 지정 의료기관을 포함하였다. 백신 이용 가능성은 원하는 백신을 맞으려 할 때 백신이 생산·유통되는지, 재고가 충분한지에 대한 내용이 포함되며 경제적 접근성은 예방접종에 드는 비용 지불 여부, 비용이 비싸

다고 여기는지, 비용 지원을 받을 수 있는지가 포함된다. 국가 예방접종 지정 의료기관은 거주지 근처에 이용할 수 있는 국가 예방접종 지정 보건소, 개인 병의원, 종합병원이 있는지, 걸어서 이용 가능한 정도인 거주지 주위 1km 내에 의료기관이 몇 개 있는지, 휴일에 진료 가능한지 여부를 포함하였다. 구조사회문화요인의 하위분류 및 용어체계는 <표 5-6>과 같다. 구조사회문화요인의 데이터모델링 예시는 <표 5-7>과 같다.

<표 5-6> 구조사회문화요인의 하위분류 및 용어체계

| 대분류 | 중분류 | 소분류 1 | 소분류 2 | |
|--------------|-------------|--------------------------|-------------------------------|-------|
| 구조사회문화 요인 | 사회적 건강신념 | 예방접종에 대한 사회적 신념 | 예방접종 필요성의 사회적 인식 | |
| | | | 예방접종 효과의 사회적 인식 | |
| | | | 예방접종 유해성의 사회적 인식 | |
| | | | 예방접종 부작용 발생의 사회적 인식 | |
| | | | 예방접종 신뢰의 사회적 인식 | |
| | | | 예방접종 복잡성의 사회적 인식 | |
| | | 예방접종에 대한 주변 사람의 신념 | 예방접종 필요성에 대한 주변 사람의 인식 | |
| | | | 예방접종 효과에 대한 주변 사람의 인식 | |
| | | | 예방접종 유해성에 대한 주변 사람의 인식 | |
| | | | 예방접종 부작용 발생에 대한 주변 사 람의 인식 | |
| | | | 예방접종 신뢰에 대한 주변 사람의 인식 | |
| | | | 예방접종 복잡성에 대한 주변 사람의 인식 | |
| | 사회지리적 요인 | 거주지 | 국내외 | |
| | | | 도시 규모 | |
| | | 예방접종 기관과의 거리 | 소요 시간 | |
| | | | 이동 거리 | |
| | | 예방접종 이용 가능성 | 백신 이용 가능성 | 백신 생산 |
| | | | | 백신 재고 |
| | 경제적 접근성 | 비용 지불 여부 | | |

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|-----|-----|-----------------|------------------------|
| | | | 비용 정도 |
| | | | 비용 지원 여부 |
| | | 국가 예방접종 지정 의료기관 | 이용 가능한 의료기관 |
| | | | 거주지에서 도보 가능한 지정 의료기관 수 |
| | | | 휴일 진료 여부 |

〈표 5-7〉 구조사회문화요인의 데이터모델링 예시

| ENTITY | ATTRIBUTE | VALUE |
|-----------------|------------------------|-----------------------|
| 거주지 | 국내외 | 국내/국외 |
| | 도시 규모 | 대도시/중소도시/군 지역 |
| 예방접종 기관과의 거리 | 소요 시간 | 걸어서 20분 이하/걸어서 20분 초과 |
| | 이동 거리 | 거주지 1km 이하/거주지 1km 초과 |
| | 거리에 대한 인식 | 가깝다/적당하다/멀다 |
| 백신 이용 가능성 | 백신 생산 여부 | 만든다/안 만든다 |
| | 백신 재고 여부 | 있다/없다 |
| 경제적 접근성 | 비용 지불 여부 | 유료/무료 |
| | 비용 정도 | 비싸다/적당하다/싸다 |
| | 비용 지원 여부 | 지원한다/지원하지 않는다 |
| 국가 예방접종 지정 의료기관 | 이용 가능한 지정 의료기관 | 보건소/개인 병의원/종합병원 |
| | 거주지에서 도보 가능한 지정 의료기관 수 | 0개/1개/2개/3개 이상 |
| | 휴일 진료 여부 | 가능하다/가능하지 않다 |

다. 정보요인

정보요인에는 예방접종과 관련된 정보와 예방접종을 하는 의료기관에 대한 정보가 포함된다.

예방접종 정보는 질병관리본부에서 제공한 ‘예방접종 대상 감염병의 역학과 관리’와 미국의 CDC(Centers for Disease Control), 영국의 NHS(National Health Service)에서 제공하는 예방접종 실무지침을 기

초로 예방접종의 종류, 예방접종 대상 감염병, 백신 구분, 투여 경로, 접종 비용 지원 여부, 원인 병원체, 투여 회차, 투여 시기, 투여 부위, 이상 반응 정보, 이상반응 대처법으로 구성하였다.

의료기관 정보는 의료기관별 정보, 의료진 정보, 의료기관 서비스 정보로 구성하였다. 의료기관별 정보에는 예방접종을 할 수 있는 의료기관의 종류, 해당 의료기관에서 접종 가능한 예방접종의 종류, 진료 과목, 국가 지정 의료기관 여부, 의료기관의 평판, 병원인증 획득 여부와 획득했다면 획득한 병원인증제의 종류를 포함하였다. 의료진 정보에는 의료진의 평판과 의료진이 예방접종을 권유하는지 아닌지가 포함된다. 의료기관 서비스 제공 정보에는 예방접종 기록, 예방접종 사전알림서비스, 예방접종 교육이 포함된다. 정보요인의 하위분류 및 용어체계는 <표 5-8>과 같으며, 정보요인의 데이터모델링 예시는 <표 5-9>와 같다.

<표 5-8> 정보요인의 하위분류 및 용어체계

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|------|---------|---------|---------------------|
| 정보요인 | 예방접종 정보 | 예방접종 종류 | 결핵 예방접종 |
| | | | B형간염 예방접종 |
| | | | 디프테리아/파상풍/백일해 예방접종 |
| | | | 폴리오 예방접종 |
| | | | b형 헤모필루스 인플루엔자 예방접종 |
| | | | 폐렴구균예방접종 |
| | | | 홍역/유행성이하선염/풍진 예방접종 |
| | | | 수두 예방접종 |
| | | | A형간염 예방접종 |
| | | | 일본뇌염 예방접종 |
| | | | 인플루엔자 예방접종 |
| | | | 로타바이러스 예방접종 |
| | | | 인유두종바이러스 예방접종 |
| | 예방접종 대상 | 결핵 | |

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|-----|-----|-------------|-------------|
| | | 감염병 | B형간염 |
| | | | 디프테리아 |
| | | | 파상풍 |
| | | | 백일해 |
| | | | 소아마비 |
| | | | 뇌수막염 |
| | | | 폐렴 |
| | | | 홍역 |
| | | | 유행성이하선염 |
| | | | 풍진 |
| | | | 수두 |
| | | | A형간염 |
| | | | 백신 구분 |
| | | 독감 | |
| | | 투여 경로 | 장염 |
| | | | 인유두종바이러스 감염 |
| | | | 피내주사 |
| | | | 경피주사 |
| | | 접종 비용 지원 여부 | 경구 |
| | | | 국가 예방접종 |
| | | 원인 병원체 | 기타 예방접종 |
| | | | 세균 |
| | | 투여 회차 | 바이러스 |
| | | | 1차 |
| | | | 2차 |
| | | | 3차 |
| | | | 4차 |
| | | | 5차 |
| | | | 6차 |
| | | | 매년 접종 |
| | | 투여 시기 | 생후 0개월 |
| | | | 생후 1개월 |
| | | | 생후 2개월 |
| | | | 생후 4개월 |

| 대분류 | 중분류 | 소분류 1 | 소분류 2 | |
|-----|----------------|-------------|----------|-------------------|
| | | | 생후 6개월 | |
| | | | 생후 12개월 | |
| | | | 생후 15개월 | |
| | | | 생후 18개월 | |
| | | | 생후 24개월 | |
| | | | 생후 36개월 | |
| | | | 만 4세 | |
| | | | 만 6세 | |
| | | | 만 11세 | |
| | | | 만 12세 | |
| | | | 투여 부위 | 허벅지 전외측 |
| | | 허벅지 근육 | | |
| | | 상완 외측면 | | |
| | | 삼각근 | | |
| | | 백신별 이상반응 정보 | 이상반응 대처법 | 발열 대처법 |
| | 접종 부위 이상반응 대처법 | | | |
| | 의료기관 정보 | 의료기관 | 의료기관 | 의료기관 종류 |
| | | | | 접종 가능한 예방접종 종류 |
| | | | | 진료 과목 |
| | | | | 국가예방접종 지정 의료기관 여부 |
| | | | | 의료기관 평판 |
| | | | | 병원인증 획득 여부 |
| | | | | 획득한 병원인증제 종류 |
| | | 의료진 | 의료진 | 의료진의 예방접종 권유 여부 |
| | | | | 의료진 평판 |
| | | 의료기관 서비스 | 의료기관 서비스 | 예방접종 기록 |
| | | | | 예방접종 사전알림서비스 |
| | | | | 예방접종 교육 |

〈표 5-9〉 정보요인의 데이터모델링 예시

| ENTITY | ATTRIBUTE | VALUE |
|------------|-----------------------|---|
| 결핵 예방접종 | 원인 병원체 | 결핵균 |
| | 대상 감염병 | 결핵 |
| | 백신 구분 | 생백신 |
| | 투여 경로 | 피내주사/경피주사 |
| | 투여 부위 | 상완외측면 |
| | 투여 회차 | 1차 |
| | 투여 시기 | 생후 0개월 |
| | 상품명 | 피내용 비시지/경피용 비시지 |
| | 접종 비용 지원 여부 | 지원 한다/지원 안 한다 |
| | 접종 비용 -경피용 비시지 | 5만 원 미만/5만 원/6만 원/7만 원/8만 원/9만 원 /10만 원/10만 원 초과 |
| | 발생 가능한 이상반응 | 접종 부위 통증/접종 부위 부종/접종 부위 발적/접종 부위 홍터/접종 부위 염증/화농성 림프절염/BCG 골 염/전신 파종성 BCG 감염증/무통성 괴양/켈로이드 |
| 보건소 | 국가 예방접종 지정 의료기관 여부 | 맞다/아니다 |
| | 접종 가능한 예방접종 | 결핵 피내용 예방접종/B형간염 예방접종/디프테리아· 파상풍·백일해 예방접종/폴리오 예방접종/b형 헤모필 루스 인플루엔자 예방접종/폐렴구균 백신/홍역·유행 성이하선염·풍진 예방접종/수두 예방접종/A형간염 예 방접종/일본뇌염 예방접종/인플루엔자 예방접종 |
| | 평판 | 매우 좋다/좋다/보통이다/나쁘다/매우 나쁘다 |

라. 환경요인

환경요인에는 정부의 예방접종 관련 정책, 예방접종에 대한 미디어 보도, 감염병의 유행이 포함된다. 정부의 예방접종 관련 정책은 예방접종 실행 정책, 정보 제공 정책, 재정 지원 정책, 예방접종 실태조사로 구성하였다. 예방접종 실행 정책에는 필수적으로 맞아야 하는 예방접종 종류와 시기가 포함되고, 정보 제공 정책에는 예방접종에 대한 일반적인 정보를 제공하는 예방접종 도우미 사이트 및 앱, 예방접종 일정을 잊어버리지 않

도록 문자로 알려주는 정기 예방접종 사전알림, 예방접종 지정 의료기관의 이름과 위치 같은 정보를 알려주는 의료기관 찾기 서비스가 포함된다. 재정 지원 정책에는 0세에서 12세에 접종해야 할 14종의 예방접종 비용을 지원하는 어린이 국가 예방접종 지원 사업이 포함되고, 예방접종 실태 조사에는 예방접종률 조사, 취약아동 예방접종 확인 사업, 예방접종 등록 사업, 예방접종에 대한 생각, 태도 등을 파악하는 예방접종 국민 의식 조사, 예방접종 인식 조사, 예방접종 지원 정책 만족도 조사가 포함된다.

미디어를 통한 예방접종 관련 보도는 부모의 예방접종에 대한 감정, 태도, 경험, 지식에 큰 영향을 미친다. 예방접종과 관련한 미디어 보도에 더해 예방접종에 대한 국민의 인식을 파악한 ‘예방접종 안전성 및 정보교환에 관한 국민 의식 조사’와 ‘예방접종 안전성 이슈에 대한 커뮤니케이션 전략 연구’ 보고서를 기반으로 미디어 매체 종류, 미디어의 보도 내용, 미디어 성향, 보도 내용의 심각성 및 보도 빈도를 포함하였다. 예방접종 대상 감염병 또는 신종 감염병의 유행에는 유행하는 감염병에 대해 인지하고 있는 심각성, 전파 정도, 유행 질환의 종류가 포함된다. 감염병의 심각성과 전파 정도에는 부모가 인지하는 수준을 포함하였고, 유행 질환의 종류에는 통계청에서 발표한 법정 감염병 발생 현황을 기초로 최근 유행한 감염병을 포함하였다. 환경요인의 하위분류 및 용어체계는 <표 5-10>과 같으며, 환경요인의 데이터모델링 예시는 <표 5-11>과 같다.

<표 5-10> 환경요인의 하위분류 및 용어체계

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|--------------|------------|------------|---------|
| 환경요인 | 정부 정책 | 예방접종 실행 정책 | 국가 예방접종 |
| | | | 기타 예방접종 |
| | 표준 예방접종 일정 | | |
| | 정보 제공 정책 | 예방접종도우미 | |
| 정기 예방접종 사전알림 | | | |

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|---------------|----------------|-------------------|--------------------|
| | | | 예방접종 지정 의료기관 정보 제공 |
| | | 재정 지원 정책 | 국가 예방접종 지원 사업 |
| | | 예방접종 실태조사 | 전국 예방접종률 조사 |
| | | | 취학아동 예방접종 확인 |
| | | | 예방접종 등록 사업 |
| | | | 예방접종 국민 의식 조사 |
| | | | 예방접종 인식 조사 |
| | | 예방접종 지원 정책 만족도 조사 | |
| | 미디어 보도 | 미디어 보도 매체 | TV |
| | | | 신문 |
| | | | 광고 |
| | | | 라디오 |
| | | 미디어 보도 내용 | 예방접종 관련 일반 정보 |
| | | | 예방접종 관련 질병 |
| | | | 예방접종 약, 백신 |
| | | | 예방접종 관련 정책 및 홍보 |
| | | | 예방접종에 대한 인식 |
| | | | 예방접종 안전성, 관련사건, 사고 |
| | 예방접종 현황 | | |
| | 예방접종 관련 인물, 기관 | | |
| | 예방접종 관련 해명 | | |
| | 미디어 보도 성향 | 긍정적 | |
| | | 중립적 | |
| 부정적 | | | |
| 미디어 보도 내용 심각성 | | | |
| 미디어 보도 빈도 | | | |
| 감염병 유행 | 유행 질환 | 독감 | |
| | | 홍역 | |
| | | 수두 | |
| | | 페렴 | |
| | | 결핵 | |
| | | A형간염 | |
| | | 백일해 | |
| | | 유행성이하선염 | |

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|-----|-----|---------|--------|
| | | | 신종플루 |
| | | | 에볼라 |
| | | | 메르스 |
| | | | 지카바이러스 |
| | | 전파 정도 | |
| | | 감염병 심각성 | |

〈표 5-11〉 환경요인의 데이터모델링 예시

| ENTITY | ATTRIBUTE | VALUE |
|------------|-----------|---|
| 예방접종 실행 정책 | 항목 | 국가 예방접종/기타 예방접종/표준 예방접종 일정 |
| | 홍보 정도 | 잘 알고 있다/알고 있다/보통이다/모른다/전혀 모른다 |
| | 만족도 | 매우 만족/만족/보통/조금 불만족/불만족 |
| 미디어 보도 | 보도 매체 | TV/신문/광고/라디오 |
| | 보도 내용 | 예방접종 관련 일반 정보/예방접종 관련 질병/예방접종 약, 백신/예방접종 관련 정책 및 홍보/예방접종에 대한 인식/예방접종 안전성, 관련 사건, 사고/예방접종 현황/예방접종 관련 인물, 기관/예방접종 관련 해명 |
| | 보도 내용 심각성 | 매우 심각하다/심각하다/보통이다/심각하지 않다/전혀 심각하지 않다 |
| | 보도 성향 | 긍정적/중립적/부정적 |
| | 보도 빈도 | 매우 자주 보도한다/자주 보도한다/보통이다/가끔 보도한다/전혀 보도하지 않는다 |
| 감염병 유행 | 전파 정도 | 매우 잘 전파된다/잘 전파된다/보통이다/가끔 전파된다/전혀 전파되지 않는다 |
| | 심각성 | 매우 심각하다/심각하다/보통이다/심각하지 않다/전혀 심각하지 않다 |
| | 유행 질환 | 장티푸스/홍역/유행성이하선염/B형간염/수두/말라리아/성홍열/쓰쓰가무시/A형간염/신증후군출혈열/독감 |

마. 예방접종 의도

예방접종 의도는 예방접종을 할 것인지 하지 않을 것인지에 대한 부모의 생각으로 어떤 예방접종의 몇 번째 회차를 어느 의료기관에서 언제 접종할 것인지, 접종하지 않을 것인지, 접종 여부를 결정하지 못하고 고민하고 있는지를 의미한다. 구성 요소로 예방접종 여부, 예방접종 항목, 예방접종 차수, 예방접종 기관, 예방접종 양상을 포함하였다.

예방접종 의도의 하위분류 및 용어체계는 <표 5-12>이며, 예방접종 의도의 데이터모델링은 <표 5-13>과 같다.

<표 5-12> 예방접종 의도의 하위분류 및 용어체계

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|---------|---------|---------------------|-------|
| 예방접종 의도 | 예방접종 여부 | 접종한다 | |
| | | 고민 중이다 | |
| | | 접종하지 않는다 | |
| | 예방접종 항목 | 결핵 예방접종 | |
| | | B형간염 예방접종 | |
| | | 디프테리아/파상풍/백일해 예방접종 | |
| | | 폴리오 예방접종 | |
| | | b형 헤모필루스 인플루엔자 예방접종 | |
| | | 페렴구균 백신 | |
| | | 홍역/유행성이하선염/풍진 예방접종 | |
| | | 수두 예방접종 | |
| | | A형간염 예방접종 | |
| | | 일본뇌염 예방접종 | |
| | | 인플루엔자 예방접종 | |
| | | 로타바이러스 예방접종 | |
| | | 인유두종바이러스 예방접종 | |
| | 예방접종 차수 | 1차 | |
| | | 2차 | |
| | | 3차 | |

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|-----|---------|-------|-------|
| | | 4차 | |
| | | 5차 | |
| | | 6차 | |
| | 예방접종 기관 | 보건소 | |
| | | 개인 병원 | |
| | | 종합병원 | |
| | 예방접종 양상 | 적기 접종 | |
| | | 지연 접종 | |

〈표 5-13〉 예방접종 의도의 데이터모델링

| ENTITY | ATTRIBUTE | VALUE |
|---------|-----------|---|
| 예방접종 의도 | 여부 | 접종한다/고민 중이다/접종하지 않는다 |
| | 항목 | 결핵 예방접종/B형간염 예방접종/디프테리아·파상풍·백일해 예방접종/폴리오 예방접종/b형 헤모필루스 인플루엔자 예방접종/페렴구균 백신/홍역·유행성이하선염·풍진 예방접종/수두 예방접종/A형간염 예방접종/일본뇌염 예방접종/인플루엔자 예방접종/로타바이러스 예방접종/인유두종바이러스 예방접종 |
| | 차수 | 1차/2차/3차/4차/5차/6차 |
| | 접종 기관 | 보건소/개인 병원/종합병원 |
| | 양상 | 적기 접종/지연 접종 |

바. 예방접종 행위

예방접종 행위는 예방접종을 하는 시점에 확인해야 할 사항과 예방접종 경험으로 접종 전 확인 사항과 예방접종 행위로 구성하였다. 예방접종 전 확인해야 할 사항은 예방접종에 대한 금기 사항이 있는지, 주의해야 할 사항이 있는지, 예방접종 당일의 건강 상태가 어떤지, 준비물은 무엇 인지를 포함한다. 예방접종 행위는 어떤 예방접종 항목의 몇 번째 회차를

어느 접종 기관에서 접종했는지, 하지 않았는지, 예방접종을 모두 시행했는지, 같은 날 몇 개의 예방접종을 했는지가 포함된다.

예방접종 행위의 하위분류 및 용어체계는 <표 5-14>와 같으며, 예방접종 행위의 데이터모델링 예시는 <표 5-15>와 같다.

<표 5-14> 예방접종 행위의 하위분류 및 용어체계

| 대분류 | 중분류 | 소분류 1 | 소분류 2 | | | |
|------------|---------------------|--------------------|-------------------------|------------|-----------|--|
| 예방접종 행위 | 접종 전 확인 사항 | 접종 금기 사항 확인 | 영구적 금기 사항 예방접종 주의 사항 | | | |
| | | 당일의 건강 상태 확인 | 건강 상태 인식 | 발열 | | |
| | | | 감기 | 기침 | | |
| | | | 콧물 | 가래 | | |
| | | | 중이염 | 설사 | | |
| | | | 준비물 확인 | 예방접종 수첩 지참 | | |
| | | | 예방접종 행위 | 예방접종 행위 여부 | | |
| | | | | 접종 항목 | 결핵 예방접종 | |
| | | | | | B형간염 예방접종 | |
| | | 디프테리아/파상풍/백일해 예방접종 | | | | |
| | 폴리오 예방접종 | | | | | |
| | b형 헤모필루스 인플루엔자 예방접종 | | | | | |
| | 페렴구균 백신 | | | | | |
| | 홍역/유행성이하선염/풍진 예방접종 | | | | | |
| | 수두 예방접종 | | | | | |
| A형간염 예방접종 | | | | | | |
| 일본뇌염 예방접종 | | | | | | |
| 인플루엔자 예방접종 | | | | | | |

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|-------------|-----|-------------|---------------|
| | | | 로타바이러스 예방접종 |
| | | | 인유두종바이러스 예방접종 |
| | | 접종 차수 | |
| | | 접종 기관 | 보건소 |
| | | | 개인 병의원 |
| | | | 종합병원 |
| | | 접종 완료 여부 | 완전 접종 |
| | | | 불완전 접종 |
| | | 접종 시기 준수 여부 | 적기 접종 |
| | | | 지연 접종 |
| 같은 날 접종한 개수 | | | |

〈표 5-15〉 예방접종 행위의 데이터모델링 예시

| ENTITY | ATTRIBUTE | VALUE |
|------------|-------------|---|
| 연구적 금기 사항 | 항목 | 백신 성분에 대한 아나필락시스 발생/백일해 백신 투여 7일 내 뇌증 발생 |
| | 여부 | 있다/없다 |
| 예방접종 주의 사항 | 항목 | 약이나 음식물(계란 포함)로 인한 알레르기 증 상/이전 예방접종 후 이상반응 발생/경련 경험 /암, 백혈병, 면역계 질환 진단/최근 3개월 내 스테로이드, 항암제, 방사선 치료 시행/최근 1 년 내 수술 또는 면역글로불린 투여 |
| | 여부 | 있다/없다 |
| 예방접종 행위 | 여부 | 접종했다/접종하지 않았다 |
| | 항목 | 결핵 예방접종/B형간염 예방접종/디프테리아· 파상풍·백일해 예방접종/폴리오 예방접종/b형 헤모필루스 인플루엔자 예방접종/폐렴구균 백 신/홍역·유행성이하선염·풍진 예방접종/수두 예방접종/A형간염 예방접종/일본뇌염 예방접 종/인플루엔자 예방접종/로타바이러스 예방접 종/인유두종바이러스 예방접종 |
| | 차수 | 1차/2차/3차/4차/5차/6차 |
| | 접종 기관 | 보건소/개인 병의원/종합병원 |
| | 접종 완료 여부 | 완전 접종/불완전 접종 |

| ENTITY | ATTRIBUTE | VALUE |
|--------|-------------|-------------------|
| | 접종 시기 준수 여부 | 적기 접종/지연 접종 |
| | 같은 날 접종한 개수 | 1개/2개/3개/4개/5개 이상 |

사. 행동반응요인

행동반응요인은 예방접종을 하고 난 후의 부모의 감정, 태도, 경험 등으로 예방접종 관련 의료서비스 경험, 예방접종 후 이상반응 경험이 포함된다. 예방접종과 관련된 의료서비스의 경험은 의료기관에 따른 경험, 의료진에 따른 경험으로 구성하였다. 의료기관에 대한 경험에는 의료기관인 보건소, 개인 병원, 종합병원의 이용 여부, 이용 만족도, 이용 대기 시간, 타인에 대한 추천 여부가 포함된다. 의료진에 대한 경험은 의사와 간호사의 친절도와 만족도가 포함된다. 예방접종 후 이상반응 경험에는 예방접종을 하고 난 후 이상반응의 발생 여부, 발생한 이상반응의 항목, 발생 시기, 지속 기간, 이상반응의 인지된 중증도, 이상반응 대처 여부가 포함된다. 행동반응요인의 하위분류 및 용어체계는 <표 5-16>과 같으며 행동반응요인의 데이터모델링 예시는 <표 5-17>과 같다.

<표 5-16> 행동반응요인의 하위분류 및 용어체계

| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|------------|----------------------|----------------|----------|
| 행동반응 요인 | 의료서비스 경험 | 의료기관 | 보건소 경험 |
| | | | 개인 병원 경험 |
| | | | 종합병원 경험 |
| | | 의료진 | 의사 경험 |
| | | 간호사 경험 | |
| | 예방접종 후 이상반응 경험 | 이상반응 발생 여부 | |
| 이상반응 항목 | | 결핵 예방접종 후 이상반응 | |

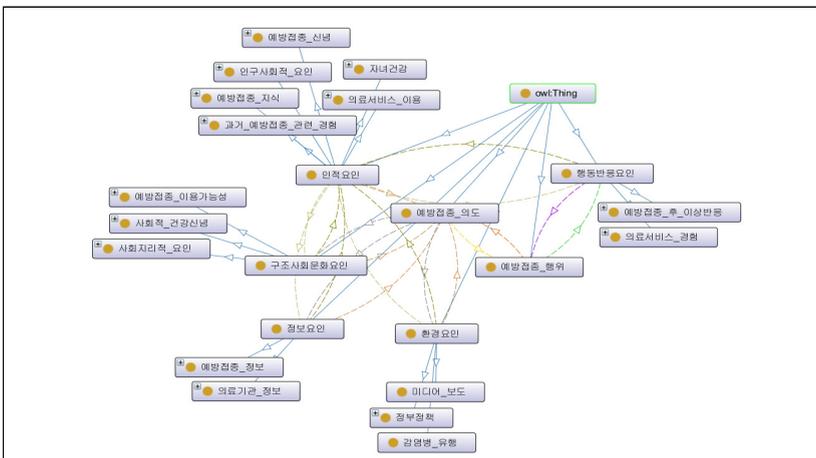
| 대분류 | 중분류 | 소분류 1 | 소분류 2 |
|-----|-----|---------------|----------------------------|
| | | | B형간염 예방접종 후 이상반응 |
| | | | 디프테리아/파상풍/백일해 예방접종 후 이상반응 |
| | | | 폴리오 예방접종 후 이상반응 |
| | | | b형 헤모필루스 인플루엔자 예방접종 후 이상반응 |
| | | | 폐렴구균 예방접종 후 예방접종 |
| | | | 홍역/유행성이하선염/풍진 예방접종 후 이상반응 |
| | | | 수두 예방접종 후 이상반응 |
| | | | A형간염 예방접종 후 이상반응 |
| | | | 일본뇌염 예방접종 후 이상반응 |
| | | | 인플루엔자 예방접종 후 이상반응 |
| | | | 로타바이러스 예방접종 후 이상반응 |
| | | | 인유두종바이러스 예방접종 후 이상반응 |
| | | 이상반응 발생 시기 | 접종 당일 |
| | | | 1일 후 |
| | | | 2일 후 |
| | | | 1주일 내 |
| | | 이상반응 지속 기간 | 1일 |
| | | | 2일 |
| | | | 1주일 내 |
| | | | 1주일 이상 |
| | | 이상반응 중증도 | |
| | | 이상반응 대처 가능 여부 | |

〈표 5-17〉 행동반응요인의 데이터모델링 예시

| ENTITY | ATTRIBUTE | VALUE |
|---------------------------|-----------------|---|
| 보건소 이용 경험 | 여부 | 있다/없다 |
| | 만족도 | 매우 만족/만족/보통/불만족/매우 불만족 |
| | 이용 대기 시간 | 길다/적당하다/짧다 |
| | 타인에 대한 추천 여부 | 추천한다/추천하지 않는다 |
| B형간염 예방접종 후 이상반응 경험 | 발생 여부 | 있다/없다 |
| | 항목 | 접종 부위 통증/접종 부위 부종/접종 부위 발적/접종 부위 발적/발열/피로감/두통/보챔/아나필락시스 |
| | 발생 시기 | 접종 당일/1일 후/2일 후/1주일 내 |
| | 지속 기간 | 1일/2일/1주일 내/1주일 이상 |
| | 중증도 | 매우 심각하다/심각하다/보통이다/심각하지 않다/전혀 심각하지 않다 |

[그림 5-4]는 본 연구에서 개발한 어린이 예방접종 온톨로지를 Protégé 5.0을 이용하여 최상위 클래스와 최상위 클래스의 관계뿐 아니라 최상위 클래스에 포함되는 하부 클래스를 제시한 것이다.

[그림 5-4] 어린이 예방접종 온톨로지



2. 온톨로지 평가

본 연구에서 개발한 어린이 예방접종 온톨로지의 범위와 영역은 개발 단계에서 작성한 17개의 Competency questions에 적절하게 답을 제시하는지로 평가하였다. 구체적인 평가 방법은 Protégé 5.0의 DL-Query를 이용하여 개념과 개념 간의 관계를 형식에 맞게 query로 입력한 후 그 결과를 확인하였다. Competency questions에 따른 DL-Query는 <표 5-18>과 같다.

<표 5-18> Competency questions와 DL-Query

| Competency questions | DL-Query |
|-----------------------------------|--|
| 예방접종을 결정하는 데 영향을 주는 요인은 무엇인가? | Is Influence Of some intention_to_vaccinate |
| 예방접종에 대해 어떤 감정을 가지고 있는가? | Is Sentiment Of some intention_to_vaccinate |
| 예방접종을 하는 데 고려하는 자녀의 건강 상태는 무엇인가? | Is Health State Of some intention_to_vaccinate |
| 예방접종에 대한 지식 수준은 어떠한가? | Is Knowledge Of some personal_factor |
| 어린이가 맞아야 하는 예방접종은 무엇인가? | Type Of some vaccination |
| 결핵 예방접종 후 발생 가능한 이상반응은 무엇인가? | Is Adverse Reaction Of some BCG_vaccination |
| 만 1세인 자녀는 어떤 예방접종을 해야 하는가? | Is Schedule Of some 1_years_old_children |
| 예방접종에 대한 정보는 어떤 매체를 통해 얻는가? | Is Information Source Of some personal_factor |
| 예방접종에 영향을 주는 자녀와 관련된 요인은? | Is Child Demographics Of some intention_to_vaccinate |
| 예방접종에 영향을 주는 과거 예방접종 관련 경험은 무엇인가? | Is Past Experience Of some vaccination |
| 예방접종 비용에 대해 어떻게 생각하는가? | Is Economic Affordability Of some intention_to_vaccinate |

| Competency questions | DL-Query |
|--------------------------------------|--|
| 디프테리아/파상풍/백일해 예방접종 후 경험한 이상반응은 무엇인가? | Is Experience Adverse Reaction Of some DTP_vaccination |
| 예방접종 후 발열 시 어떻게 해야 하는가? | Is Fevermanage Of some information_factor |
| 예방접종을 하러 간 의료기관에 대해 만족하는가? | Is Satisfaction Of some public_health_center |
| 의료진은 친절한가? | Is Kindness Of some staff |
| 정부의 예방접종 관련 정책은 무엇인가? | Is Policy Of some intention_to_vaccinate |
| 예방접종 관련 미디어의 성향은 어떠한가? | Is Media Tendency Of some intention_to_vaccinate |

평가 결과, 본 연구에서 개발한 어린이 예방접종 온톨로지가 작성된 17개의 Competency questions에 대해 모두 정확한 결과를 도출하였다. [그림 5-5]는 Competency questions 중 ‘예방접종을 결정하는 데 영향을 주는 요인은 무엇인가?’라는 질문에 대한 답을 찾는 쿼리(query)를 입력한 결과를 나타낸 예이다.

3. 어린이 예방접종 소셜 빅데이터 온라인 문서 현황

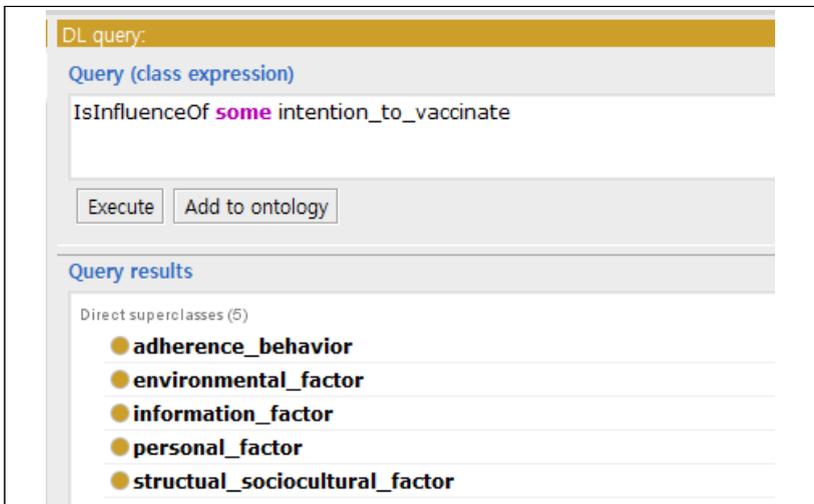
가. 분석 방법

1) 연구 대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜네트워크서비스(SNS), 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 대상으로 하였다. 본 분석에서는 99개의 온라인 뉴스 사이트, 3개의 블로그(네이버·다음·티스토리), 2개의 카페(네이버·다음), 1개의 SNS(트위터), 10개

의 게시판(네이버 지식인·네이트 지식·네이트톡·네이트판 등) 총 115개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서(버즈)를 소셜 빅데이터로 정의하였다. 지카바이러스 관련 토픽은 2016. 1. 1.~2016. 4. 15. 해당 채널에서 요일별, 주말, 휴일을 고려하지 않고 매시간 단위로 수집하였으며, 수집된 총 6만 5673건의 텍스트(Text) 문서를 본 연구 분석에 포함시켰다. 본 연구를 위한 소셜 빅데이터의 수집²³⁾은 크롤러(Crawler)를 사용하였고, 토픽의 분류는 주제분석(text mining) 기법을 사용하였다. 지카바이러스 토픽은 모든 관련 문서를 수집하기 위해 ‘지카바이러스’, ‘Zikavirus’, 그리고 ‘지카바이러스 감염’ 등을 사용하였다.

[그림 5-5] Competency questions 평가 과정 예시



23) 본 연구를 위한 소셜 빅데이터의 수집 및 토픽 분류는 '(주)SK텔레콤 스마트인사이트'에서 수행함.

2) 연구 도구

지카바이러스와 관련하여 수집된 버즈는 주제분석(text mining)과 요인분석(factor analysis)의 과정을 거쳐 <표 5-19>와 같이 정형화 데이터로 코드화하여 사용하였다.

<표 5-19> 어린이 예방접종 빅데이터 연구 도구(키워드)

| 속성 | 키워드 | |
|-------|--|---|
| 감정 | 긍정 | 가능, 가치, 각별, 간단, 간편, 감동, 감안, 값어치, 강력, 강조, 강추, 강화, 개선, 개편, 격려, 결심, 고급, 고려, 공감, 공급, 공유, 과감, 관심, 구입, 구조, 권고, 권유, 권장, 귀여움, 귀요미, 극복, 금상첨화, 긍정, 기대, 기쁨, 기원, 기적, 기특, 기회, 깔끔, 깜찍, 깨끗, 꾸준 등 |
| | 부정 | 갈등, 감소, 감시, 갑갑, 강제, 개고생, 거부, 거짓, 거짓말, 걱정, 검토, 결핍, 경각심, 경고, 경박, 경악, 고리타분, 고민, 고발, 고비, 고생, 고통, 곤혹, 골치, 공격, 공포, 과격, 과다, 과도, 과시, 과장, 귀차니즘, 규제, 규탄, 극심, 급락, 급증, 기만, 긴급, 긴장, 까기, 캄캄, 폼수, 꼼찍, 나쁨, 낙폭, 난동, 난리, 난해, 남용, 냉소 등 |
| 대상 | 의사, 간호사, 한의사, 치과 의사, 조산사, 닥터, 의사쌤, 의사분, 간호사쌤, 간호사분, 간호조무사, 물리치료사, 방사선사, 수의사, 약사, 영양사, 원장, 위생사, 의료기사, 의료인, 의료진, 의학자, 임상병리사, 작업치료사, 전문의, 치과기공사, 치과위생사, 피부과전문의, 한약사, 한약조제사, 감기환자, 감염자, 감염환자, 고위험군, 고혈압환자, 뇌혈관환자, 당뇨환자, 만성질환자, 보균자, 사망자, 시한부, 암환자, 외래환자, 위험군, 응급환자, 의심환자, 치매환자, 폐렴인, 확진자, 확진환자, 환자 | |
| 질병 | 감염병, 전염병, 합병증, 장애, 병, 질환, 결핵, B형간염, 디프테리아, 파상풍, 백일해, 폴리오, b형헤모필루스인플루엔자, 홍역, 유행성이하선염, 풍진, 소아마비, 뇌수막염, 폐렴, 볼거리, 수두, A형간염, 일본뇌염, 인플루엔자, 로타바이러스감염, 인유두종바이러스감염, 신종플루, 독감, 감기, 장염, 인플루엔자바이러스감염, 생식기사마귀, 에이형간염, 중이염, AI, avianinfluenza, C형간염, H3N2, H7N9형, INFLUENZA, 간경화, 간암, 간염, 간질환, 감염, 감염질환, 갑상선암, 건선, 고병원성, 고병원성AI, 고지혈증, 고콜레스테롤혈증, 골다공증, 관상동맥질환, 광우병, 구제역, 급성폐수종, 난치병, 노로바이러스, 뇌염, 뇌졸중, 뇌출혈, 뇌혈관, 뇌혈관질환, 당뇨, 당뇨병, 당뇨병성황반부종, 대사증후군, 대장암, 멧기열, 만성, 만성신부전, 만성질환, 만성콩팥질환, 만성폐쇄성질환, 만성호흡기질환, 말초동맥질환, 메르스, 발암성, 백혈병, 변형인플루엔자, 북부비만, 불면증, 비브리오패혈증, 비염, 사스, 성인병, 수두바이러스, 수인성, 수족구병, 식도암, 식중독, 신부전증, 신증후군, 심근경색, 심장병, 심장질환, 심장폐렴, 심혈관질환, 쓰쓰가무시증, 아토피, 알코올중독, 암, 에볼라, 에이즈, 우울증, 위암, 유방암, 인슐린저항성, 자궁경부암, 자궁암, 자폐증, 장티푸스, 전립선암, 조류인플루엔자, 조수증, 증동호흡기증후군, 증증감염, 지카, 지카바이러스, 천식, 축농증, 치매, 콜레라, 탈모, 패혈증, 폐암, 폐질환, 하지정맥류, 협심증, 호흡기질환, 홍콩독감, 황달, 황열 | |
| 원인병원체 | 세균, 결핵균, 디프테리아균, 파상풍균, 백일해균, b형헤모필루스균, 폐렴구균, 바이러스, B | |

| 속성 | 키워드 |
|-------|---|
| | 헝간염바이러스, 폴리오바이러스, 홍역바이러스, 유행성이하선염바이러스, 풍진바이러스, 수두대상포진바이러스, A형간염바이러스, 일본뇌염바이러스, 인플루엔자바이러스, 로타바이러스, 인유두종바이러스, 바리셀라조스터바이러스, A군, B군, 갑염성, 갑염증, 곰팡이, 구균, 구제역바이러스, 기생충, 내성균, 뇌혈전, 다이옥신, 대장균, 독, 독성, 독소, 로타, 박테리아, 발암, 발암물질, 벌레, 벤조피렌, 변형바이러스, 병원균, 병원체, 보툴리누스균, 보툴리눔, 비브리오, 살모넬라균, 생탄저균, 수막구균, 슈퍼박테리아, 식물바이러스, 식중독균, 알레르기성, 암세포, 야생진드기, 원인균, 진드기, 카드뮴, 콜레스테롤, 탄저균, 폐구균, 폐구균, 폐렴구균, 폐렴균, 혈청형 |
| 발병 원인 | 간접접연, 과다복용, 과다섭취, 과로, 과식, 과음, 노출, 농약, 니코틴, 담배, 물연변이, 매개모기, 맹독, 먼지, 면역력, 면역력저하, 모기, 물놀이, 미세먼지, 방부제, 방사능, 방사선, 방사성, 보형물, 본태성, 비만, 성관계, 스킨십, 식습관, 야식, 여행, 영양, 오염, 오염수, 운동부족, 유해물질, 음주, 이물질, 일교차, 자외선, 전염, 전염성, 전자담배, 증복흡연, 찬바람, 체중, 출국, 트랜스지방, 포용, 해외여행, 환경호르몬, 환절기, 황사, 흡연, 흡입독성, 흰줄슴모기 |
| 증상 | 후유증, 부작용, 이상반응, 통증, 부종, 발적, 흉터, 염증, 화농성립상결막, BCG골염, 전신파종성BCG감염증, 무통성결막, 켈로이드, 동통, 근육통, 아픔, 두통, 붓기, 부어오름, 팽진, 빨갛게, 발진, 빨개, 충창, 농양, 고름, 종기, 화농, 발열, 피로감, 보챔, 아나필락시스, 열, 미열, 불덩이, 계속자요, 권태감, 무력감, 알레르기쇼크, 어지러움, 식욕부진, 전신두드러기, 멍침, 단단함, 상완신경염, 길랑증후군, 바레증후군, 림프절염, 무력증, 과민반응, 경련, 압통, 소양감, 자반증, 알레르기, 점상출혈, 붉은반점, 경련, 혈소판감소증, 경기, 발작, 중추신경계이상, 뇌신경마비, 림프절염증, 뇌증, 뇌질환, 관절통, 혈관부종, 호르몬관, 저혈압성허탈, 중추신경계이상반응, 말초신경증, 눈호흡증후군, 천명, 급성염증성탈수초성다발성신경병증, 천명음, 구토, 토, 오바이트, 기절, 의식소실, 가려움증, 알러지, 대상포진, 두드러기, 고열, 돌연사, 마비, 몸살, 복통, 사망, 소화, 신경통, 오한, 의식불명, 의식장애, 재발, 피부발진, 감염성질환, 고령화, 고혈압, 관절염, 급성, 노화, 노화촉진, 동맥경화, 물질, 발기부전, 발병, 배탈, 변비, 빈혈, 신경장애, 심부전, 위산과다, 의심증세, 인후통, 재채기, 즉상경화, 증증질환, 출혈, 코막힘, 탈수, 폐경, 폐손상, 피부상처, 피부자극, 기침, 콧물, 가래, 설사 |
| 예방 | 수분섭취, 해열제, 청결, 냉찜질, 가슴기, 걸싸개, 격리, 금연, 긴바지, 긴소매, 다이어트, 마사지, 마스크, 모기장, 모자, 목욕, 목관리, 물수건, 방역, 방충망, 보청기, 비누, 비상방역체계, 산책, 살균, 살균제, 살충제, 생활습관, 샤워, 소독, 소식, 속싸개, 손수건, 손씻기, 수면, 숙면, 스트레스해소, 스트레칭, 식단, 식생활, 식생활안전가이드, 식이요법, 여가, 연구개발, 예방, 예방법, 운동, 위생, 위생관리, 유산소운동, 절식, 제습제, 콘돔, 퇴치제, 환기, 휴식, 황사마스크, 모기기피제, 위생상태, 파스, 세정제, 니코틴패치, 구강청결제, 피임, 손세정제, 급주, 인공눈물, 방충제, 낮잠, 물비누, 손소독제, 모기약, 단식, 세척제, 차폐예방, 흡연경고, 양치 살균기 |
| 예방교육 | 모자보건수첩, 산모수첩, 아기수첩, 예방접종도우미홈페이지, 예방접종도우미사이트, 예방접종도우미앱, 예방접종도우미플러스, 예방접종도우미애플리케이션, 수첩, 팜플렛, 안내책자, 교육자료, 안내문, 병원홈페이지, 병원사이트, 병원앱, 병원플러스, 병원애플리케이션, 예방접종등록시스템, 지방자치단체인터넷홈페이지, 지방자치단체홈페이지, 홈페이지, 집단교육, 방문교육, 예방접종도우미, 정기예방접종사전알림, 예방접종지정의료기관정보제공, MERS, 가이드라인, 교육, 메르스포털, 상담, 안전교육, 육아일기 |
| 예방검사 | 통원 외래진료, CT, 건강검진, 검사, 검진, 영유아건강검진, 영유아검진, 유전자검사, 입원, 진단, 진단서, 진료, 진찰, 피검사, 항체검사, 혈액검사, MRI, 내시경, 세균검사, 정밀검사, 진 |

| 속성 | 키워드 |
|--------|--|
| | 단키트, 방사능검사, 엑스레이, 원격진료, 원격의료, 안전검사, 대장내시경, X레이, CT촬영, 의사처방, 수면내시경 |
| 치료제 | CDC약물, DPT, 가정상비약, 감기약, 개량신약, 다당백신, 당뇨병치료제, 당뇨치료제, 두통약, 램시마, 램시마SC, 레이저, 리피토, 마취제, 만병통치약, 먹는약, 멀미약, 면역억제제, 무릎관절수술, 바스코, 바이알, 바이오시밀러, 바이오의약품, 발기부전치료제, 백신, 보충제, 보톡스, 복제약, 비아그라, 사후피임약, 상비약, 상처치료, 세포배양백신, 소독약, 소화제, 수면마취제, 수면제, 수술, 수액, 수혈, 시술, 식약, 식염수, 신약, 아스피린, 아토피크림, 알약, 약물, 약제, 약품, 에볼라백신, 에볼라치료제, 영양제, 예방백신, 우유주사, 의약, 의약품, 의약품, 인공관절수술, 일반약, 일반의약품, 임플란트, 전립선치료제, 전문의약품, 정력제, 제대혈, 제브타나, 줄기세포치료, 줄기세포치료제, 지엔티파마, 진통제, 천식약, 천연물신약, 첩분제, 초음파, 치료, 치료법, 치료약, 치료제, 카바수술, 타미플루, 타시그나, 타이레놀, 탈리도마이드, 특효약, 페니실린, 표적항암제, 플라시보, 피임약, 필러, 한약, 한약제, 항바이러스제, 항생제, 항암, 항암제, 항암치료, 항염증치료제, 항체, 항체치료제, 항정신성의약품 |
| 예방접종 | 결핵예방접종, B형간염예방접종, 디프테리아예방접종, 파상풍예방접종, 백일해예방접종, 폴리오예방접종, b형헤모필루스인플루엔자예방접종, 폐렴구균예방접종, 홍역예방접종, 유행성이하선염예방접종, 풍진예방접종, 수두예방접종, A형간염예방접종, 일본뇌염예방접종, 인플루엔자예방접종, 로타바이러스예방접종, 인유두종바이러스예방접종, BCG, 비씨지, 불주사, 디터피, 디피터, Dtap, Td, Tdap, DtapIPV, IPV, 엠엠알, MMR, Hib, PCV, 폐렴구균단백결합, 폐렴구균다당질, 일본뇌염생백신, 일본뇌염사백신, 독감예방접종, 자궁경부암예방접종, HPV, 생백신, 사백신, 근육주사, 피하주사, 피내주사, 경피주사, 경구용, 불활성화백신, 경구약, 국가예방접종, 기타예방접종, 폐렴구균백신, 결핵경피용예방접종, DTP, 결핵피내용예방접종, 경구, 국가필수예방접종, 독감백신, 독감예방주사, 독감접종, 독감주사, 로타릭스, 로타텍, 무료예방접종, 백신접종, 일본뇌염사백신예방접종, 일본뇌염생백신예방접종, 임시예방접종, 적기예방접종, 필수예방접종, 허브 |
| 예방접종기간 | 1차, 2차, 3차, 4차, 5차, 6차, 매년접종, 무접종, 미접종, 적기접종, 지연접종, 완전접종, 불완전접종, 1년, 1회, 2년, 2회, 3년, 3회, 4년, 4회, 5년, 5회, 6년, 6회, 네번째, 다섯번째, 두번째, 세번째, 여섯번째, 접종시기, 첫번째, 추가접종 |
| 신체 부위 | 다리, 허벅지, 팔뚝, 피부, 허반신, 허리, 머리, 머리카락, 몸통, 무릎, 발목, 발바닥, 손가락, 손톱, 얼굴, 엉덩이, 연골, 유방, 이마, 입술 |
| 전신 | DNA, 관절, 근력, 근육, 기관지, 기도, 내장, 동맥류, 림프구, 림프절, 말초혈관, 면역세포, 면역체계, 모발세포, 무릎관절, 배아줄기세포, 백혈구, 뱃살, 뱃속, 삼각근, 상완외측면, 세포벽, 성기능, 성장호르몬, 세포, 세포벽, 신경계, 신경세포, 신장, 신진대사, 심장, 심혈관, 압박골절, 요오드, 위장, 유익균, 유전자, 인체, 자궁, 자율신경, 장기, 중앙, 줄기세포, 지방, 청력, 체내, 체온, 포화지방, 허벅지근육, 허벅지전외측, 혈관, 혈당, 혈압, 혈액, 혈액순환, 혈중, 혈행, 호르몬, 호흡, 호흡기, 후관절, 몸속, 갑상선, 체지방, 전립선, 방광, 난자, 혈소판, 온몸, 피부세포, 골수, 부정맥, 콩팥, 염색체, 해마 |
| 성분 | 글루타민산나트륨, 글루타민산소듐, 나트륨, 단백질, 당분, 비타민, 비타민A, 비타민C, 비타민D, 아미노산, 영양소, 유산균, 유해성분, 저염식, 카페인, 칼슘, 탄수화물, 핵심성분, 당류, 인슐린, 영양성분, 마약성분, 식이섬유, 철분, 미네랄, 염산, 탄산, 글루텐, 카제인, 식물성, 트리카르산, 저나트륨, 포도당, 폴라렌, 농약성분, 김치유산균, 주성분, 비타민제, 칼륨, 염분, 저지방, 아세트아미노펜, 가공식품 |
| 식품 | 가공식품, 가공육, 간장, 갈치, 감, 건강기능식품, 건강식품, 건강음식, 계란, 고기, 고등어, 고 |

| 속성 | 키워드 |
|-----|---|
| | 래고기, 고추, 고추장, 고춧가루, 과일, 과자, 국물, 국물음식, 김밥, 김치, 나물, 날계란, 냉면, 노가리, 녹황색채소, 농산물, 농수산물, 농식품, 다시마, 달걀, 닭고기, 당근, 대추, 도시락, 독버섯, 돼지고기, 떡, 떡볶이, 라면, 레모나, 말고기, 먹거리, 명절식품, 명태, 목장치즈, 물엿, 미역, 밀, 밀가루, 바나나, 반찬, 밥, 배, 배추, 백수오, 버섯, 벌꿀, 베이킹소다, 보약, 보조식품, 복어회, 불나물, 분유, 불량식품, 빠가사리, 빵튀기, 사과, 사양물, 사탕, 산후식품, 산후조리식품, 삼계탕, 삼계탕, 생선, 석류액, 선호식품, 설탕, 성수식품, 소금, 소시지, 소탕, 쇠고기, 수산물, 수입식품, 순대, 시금치, 식빵, 식용, 식자재, 식재료, 식품, 쌀, 쌀국수, 아이스크림, 약과, 어린이식품, 어패류, 연어, 열대과일, 오리고기, 옥수수, 완전식품, 우수성과자, 육류, 울무, 음식, 이엽우피소, 이유식, 인삼, 족발, 중국산마늘, 참기름, 참외, 참치, 채소, 천일염, 초콜릿, 축산물, 치즈버거, 케이크, 케톤식, 코리아멜론, 콩, 토마토, 통조림, 푸드, 피자, 한과, 한우, 할랄식품, 해조류, 햄버거, 홍삼, 흑마늘 |
| 식음료 | 건강음료, 녹즙, 녹차, 두유, 막걸리, 맥주, 박카스, 사케, 생수, 생즙, 소주, 술, 알코올음료, 에너지음료, 오미자액, 와인, 우유, 웰빙술, 음료, 전통술, 주류, 주스, 차음차림, 커피, 커피믹스, 콜라, 탄산수, 탄산음료, 허브차 |
| 비용 | 접종비, 가격, 요금, 무료, 공짜, 무상, 비용, 경비, 자금, 금액, 병원비, 진료비, 국가비용, 무료지원, 5만원 미만, 5만원, 6만원, 7만원, 8만원, 9만원, 10만원, 10만원 초과, 오만원대, 육만원대, 칠만원대, 팔만원대, 구만원대, 십만원대, 100만원, 3만원, 500만원, 50만원, 가산세, 간병비, 고가, 교육비, 구만원, 보육료, 보험금, 보험료, 유료, 의료비, 반값, 비급여, 부담금, 보조금, 무료접종, 수당, 고액, 지원금, 신고포상금, 약제비, 수수료, 검사비, 건강증진부담금 |

3) 연구 결과

〈표 5-20〉과 같이 어린이 예방접종과 관련하여 긍정적인 감정은 63.9%, 부정적인 감정은 27.3%로 나타났다. 언급 채널은 카페 48.9%, 블로그 34.8%, 트위터 8.3% 등의 순으로 나타났다. 최초로 작성된 문서는 87.0%, 확산문서는 13.0%로 나타났다. 의료인의 언급이 52.1%, 환자의 언급이 47.9%로 나타났다. 어린이의 언급이 48.4%, 성인의 언급이 51.6%로 나타났다. 질병 유형은 제3군감염병, 질병, 제2군감염병, 감염증 등의 순으로 나타났다. 예방교육은 교육 방법, 교육 제공 사이트 등의 순으로 언급되었으며, 예방은 예방, 생활습관, 모기 예방, 식생활 등의 순으로 나타났다. 원인병원체는 감염 유발 원인, 세균, 바이러스 등의 순으로 언급된 것으로 나타났다. 발병 원인은 면역력 저하, 감염, 식습관 등의

순으로 나타났다. 증상은 부작용, 발열, 통증, 호흡기계, 출혈 등의 순으로 언급된 것으로 나타났다. 치료제는 백신, 치료제, 항암제, 항생제 등의 순으로 언급된 것으로 나타났다. 예방접종은 인플루엔자, 예방접종 분류, 폐렴구균, 홍역 등의 순으로 언급된 것으로 나타났다. 예방접종 기간은 접종 회차, 접종 시기 등의 순으로 나타났다. 신체 부위는 머리, 하반신 등의 순으로 나타났다. 식음료는 유제품, 차/커피, 음료 등의 순으로 나타났다. 전신(신체 내부)은 인체, 순환기계, 호흡기계, 세포 등의 순으로 언급된 것으로 나타났다. 비용은 무료, 비용, 실제 금액 등의 순으로 나타났다. 성분은 무기질, 영양소, 유익성분 등의 순으로 나타났다. 식품은 식품, 음식, 과일, 채소 등의 순으로 나타났다.

〈표 5-20〉 어린이 예방접종 관련 온라인 문서 현황

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) | |
|-----------------|-------------|--------------|---------------|-----------|-------------|-------------|
| 감정 | 긍정 | 46,985(63.9) | 원인 병원체 | 병원체 독성 등 | 1,608(7.2) | |
| | 보통 | 6,501(8.8) | | 유발 증상 | 2,197(9.9) | |
| | 부정 | 20,043(27.3) | | 감염 유발 원인 | 7,172(32.2) | |
| | 계 | 73,529 | | 세균 | 6,153(27.7) | |
| 채널 | 블로그 | 57,225(34.8) | | 바이러스 | 4,596(20.7) | |
| | 카페 | 80,324(48.9) | | 기생충 | 125(0.6) | |
| | 트위터 | 13,618(8.3) | | 콜레스테롤 | 392(1.8) | |
| | 게시판 | 6,444(3.9) | | 계 | 22,243 | |
| | 뉴스 | 6,687(4.1) | | 발병 원인 | 흡연 | 1,960(8.2) |
| | 계 | 164,298 | | | 면역력 저하 | 6,152(25.6) |
| | 순계정 | 최초 문서 | 143,015(87.0) | | 식습관 | 3,949(16.4) |
| 확산 문서 | | 21,283(13.0) | 모기 | | 1,094(4.5) | |
| 계 | | 164,298 | 감염 | | 4,080(17.0) | |
| 트위터 언급 형태 | Interactive | 97(0.7) | 민지 | | 775(3.2) | |
| | Propagation | 3,697(27.1) | 환경 | | 1,307(5.4) | |
| | Monologue | 2,480(18.2) | 여행 | | 2,409(10.0) | |
| | Reply | 854(6.3) | 유해물질 | | 941(3.9) | |
| | Link | 6,490(47.7) | 접촉 | | 985(3.3) | |
| | 계 | 13,618 | 체질 | 212(0.9) | | |
| 환자 구분 | 의료인 | 9,579(52.1) | 돌연변이 | 383(1.6) | | |
| | 환자 | 8,810(47.9) | 계 | 24,047 | | |
| | 계 | 18,389 | 증상 | 부작용 | 6,777(14.8) | |
| 어린이 구분 | 어린이 | 44,597(48.4) | | 통증 | 5,079(11.1) | |
| | 성인 | 47,625(51.6) | | 피부 문제 | 3,314(7.2) | |
| | 계 | 92,222 | | 감염 | 4,031(8.8) | |
| | | | | | | |

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) |
|--------|-------------|--------------|-------------|-------------|--------------|
| 질병 유형 | 질병 | 12,358(17.1) | 예방 | 발열 | 6,179(13.5) |
| | 제1군감염병 | 3,627(5.0) | | 피로 | 749(1.6) |
| | 제2군감염병 | 10,412(14.4) | | 국민 반응 | 1,592(3.5) |
| | 제3군감염병 | 15,834(21.9) | | 출혈 | 1,064(10.6) |
| | 제4군감염병 | 3,619(5.0) | | 중추 및 말초신경계 | 2,171(4.7) |
| | 지정감염병 | 596(0.8) | | 호흡기계 | 4,852(10.6) |
| | 감염증 | 8,908(12.3) | | 순환계 | 1,873(4.1) |
| | 암 | 4,587(6.3) | | 위장관계 | 3,060(6.7) |
| | 간질환 | 1,631(2.3) | | 사망 | 3,548(7.7) |
| | 순환기계질환 | 1,611(2.2) | | 노화 | 1,251(2.7) |
| | 호흡기계질환 | 2,323(3.2) | | 생식기계 | 282(0.6) |
| | 뇌혈관질환 | 1,006(1.4) | | 계 | 45,822 |
| | 대사증후군 | 2,419(3.3) | | 예방 | 10,679(29.4) |
| | 신장질환 | 968(1.3) | | 위생 | 5,746(15.8) |
| | 피부질환 | 640(0.9) | | 생활습관 | 7,175(19.7) |
| 신경정신질환 | 1,674(2.3) | 식생활 | 2,965(8.2) | | |
| 근골격계질환 | 176(0.2) | 금연 | 903(2.5) | | |
| 계 | 72,389 | 피인 | 309(0.8) | | |
| 예방 교육 | 교육 제공 사이트 | 3,709(34.1) | 호흡기 예방 | 2,395(6.6) | |
| | 예방접종 도우미 | 301(2.8) | 모기 예방 | 3,637(10.0) | |
| | 교육 방법 | 6,636(60.9) | 발열 대처 방법 | 2,333(6.4) | |
| | 메르스 교육 | 246(2.3) | 차매 예방 | 71(0.2) | |
| | 계 | 10,892 | 기타 | 147(0.4) | |
| 치료제 | 치료제 | 6,494(19.4) | 계 | 36,360 | |
| | 약제 | 2,567(7.7) | 유전자 | 580(2.6) | |
| | 수술 | 1,402(4.2) | 세포 | 1,900(8.5) | |
| | 시술 | 1,679(5.0) | 인체 | 3,166(14.1) | |
| | 백신 | 10,480(31.3) | 장기 | 1,139(5.1) | |
| | 감염병 치료제 | 446(1.3) | 체지방 | 1,160(5.2) | |
| | 마취제 | 112(0.3) | 근골격계 | 1,186(5.2) | |
| | 진통제 | 910(2.7) | 호흡기계 | 2,825(12.6) | |
| | 항암제 | 4,081(12.2) | 면역체계 | 1,708(7.6) | |
| | 항생제 | 2,884(8.6) | 생식기계 | 602(2.7) | |
| | 보충제 | 869(2.6) | 내분비계 | 1,097(4.9) | |
| | 기타 약제 | 1,547(4.6) | 뇌신경계 | 1,440(6.4) | |
| | 계 | 33,471 | 비노기계 | 678(3.0) | |
| 예방 접종 | 예방접종 | 407(1.3) | 순환기계 | 3,109(13.9) | |
| | 백신 상태 | 1,633(5.2) | 전신 | 1,505(6.7) | |
| | 예방접종 경로 | 542(1.7) | 질환 | 322(1.4) | |
| | 예방접종 분류 | 5,250(16.8) | 계 | 22,417 | |
| | 결핵 | 2,109(6.7) | 비용 | 비용 | 7,596(30.9) |
| | 간염 | 1,345(4.3) | | 무료 | 10,722(43.6) |
| | 디프테리아 | 1,762(5.6) | | 유료 | 692(2.8) |
| 폴리오 | 855(2.7) | 실제 금액 | | 2,901(11.8) | |
| | | 비용 정도 | | 480(2.0) | |

| 구분 | 항목 | N(%) | 구분 | 항목 | N(%) | |
|----------------|-------------------|-------------|----------|-------------|-------------|-------------|
| | b형 헤모필루스 인플루엔사 | 849(2.7) | | 지원금 | 500(2.0) | |
| | 페렴구균 | 2,069(6.6) | | 부과금 | 296(1.2) | |
| | 홍역 | 2,026(6.5) | | 보험료 | 505(2.1) | |
| | 수두 | 263(0.8) | | 교육비 | 899(3.7) | |
| | 일본뇌염 | 682(2.2) | | 계 | 24,591 | |
| | 인플루엔사 | 9,161(29.3) | | 성분 | 무기질 | 2,184(49.9) |
| | 로타바이러스 | 1,064(3.4) | | | 영양소 | 1,324(30.3) |
| | 인유두종바이러스 | 1,249(4.0) | | | 유익성분 | 651(14.9) |
| 계 | 31,266 | 유해성분 | 59(1.3) | | | |
| 예방 접종 기간 | 접종 회차 | 8,396(71.5) | 약물 | 157(3.6) | | |
| | 접종 상태 | 447(3.8) | 계 | 4,375 | | |
| | 접종 시기 | 1,092(16.2) | 식품 | 식품 | 4,348(25.0) | |
| | 접종 구분 | 1,004(8.5) | | 육류 | 905(5.2) | |
| 계 | 11,749 | 어패류 | | 409(2.4) | | |
| 신체 부위 | 머리 | 6,205(40.2) | | 해조류 | 277(1.6) | |
| | 상반신 | 2,406(15.6) | | 채소 | 1,608(9.3) | |
| | 하반신 | 3,627(23.5) | | 과일 | 1,709(9.8) | |
| | 기타 부위 | 3,209(20.8) | | 버섯류 | 305(1.8) | |
| | 계 | 15,447 | | 곡류 | 155(0.9) | |
| 식음료 | 음료 | 332(10.4) | | 계란 | 874(5.0) | |
| | 차&커피 | 939(29.5) | | 장&오일&조미료 | 881(5.1) | |
| | 유제품 | 1,042(32.8) | 한약제 | 193(1.1) | | |
| | 주류 | 413(13.0) | 과자 | 1,281(7.4) | | |
| | 생수 | 135(4.2) | 가공식품 | 197(1.1) | | |
| | 음료&주스 | 320(10.1) | 음식 | 4,230(24.3) | | |
| | 계 | 3,181 | 계 | 17,372 | | |

제4절 논의

본 연구에서는 소셜미디어상의 어린이 예방접종에 대한 부모의 감정, 태도, 경험, 지식과 관련된 용어를 수집하고 분석하기 위한 온톨로지를 개발하였고, 이 과정에서 용어체계와 EAV 모델을 함께 개발하였다.

어린이 예방접종에 대한 부모의 감정, 태도, 경험, 지식은 부모와 아이의 인적요인에서 사회문화적, 정책 및 환경적 요인까지 다양한 수준의 요인에 영향을 받으며, 예방접종을 결정하기 전부터 접종 시점, 접종 후의

전 기간에 걸쳐 영향을 미치는 것으로 나타나 온톨로지의 주요 개념을 예방접종 이전의 인적요인, 구조사회문화요인, 정보요인, 환경요인, 예방접종 의도, 예방접종 시점의 예방접종 행위, 예방접종 후의 행동반응요인으로 구성하였다. 각 요인에 해당하는 용어를 수집하기 위해 예방접종 관련 국내외의 실무지침과 연구논문을 고찰하였고 소셜 데이터를 분석하였다. 그 결과 전체 2,765개의 용어와 857개의 개념이 수집되었고, 이를 7개의 최상위 클래스와 208개의 하위 클래스로 구성하였다.

온톨로지에 포함된 최상위 클래스를 살펴보면, 인적요인에는 부모의 예방접종에 대한 신념 및 지식, 부모·자녀·가족의 인구사회적 요인, 자녀 건강, 의료서비스 이용, 과거 예방접종 관련 경험이 포함되며 구조사회문화요인에는 예방접종에 대한 사회적 신념과 부모 및 주변 사람들의 신념, 사회지리적 요인, 예방접종 이용 가능성이 포함된다. 정보요인에는 예방접종에 대한 정보와 의료기관에 대한 정보가 포함되고, 환경요인에는 정부 정책, 미디어 보도, 감염병 유행이 포함된다. 예방접종 의도에는 예방접종을 할 것인지가 포함되며, 예방접종 행위에는 접종 전 확인 사항, 예방접종에 대한 행위가 포함된다. 행동반응요인에는 의료서비스 경험과 예방접종 후 이상반응 경험이 포함된다.

이들 최상위 클래스 사이의 관계를 살펴보면 구조사회문화요인, 정보요인, 환경요인이 인적요인에 영향을 미치고, 인적요인은 예방접종 의도를 결정하는 데 영향을 주며, 예방접종 의도에 따라 예방접종 행위가 결정된다. 예방접종 행위 후의 행동반응요인은 또다시 인적요인에 영향을 미치며 다음번 예방접종을 결정하는 데 영향을 미친다.

본 연구에서 개발한 어린이 예방접종 온톨로지에 부모들의 감정, 태도 등이 포함되어 있는지 확인하기 위해 미국 부모들의 예방접종에 대한 우려 사항(Kennedy et al., 2011)에 포함된 내용과 비교해 보았다. 이 연

구에서 ‘한 번의 병원 방문에 너무 많은 예방접종을 하고 아이에게 고통을 줌’, ‘첫 2년 동안 너무 많은 예방접종을 함’, ‘백신은 아이에게 발열, 만성질환, 자폐증과 같은 장애를 일으킴’, ‘백신은 안전하지 않음’과 같은 우려 사항이 표현되어 있었는데, 이러한 내용이 본 연구에서 개발한 어린이 예방접종 온톨로지의 인적요인의 하위 요인인 예방접종에 대한 신념에 모두 포함되어 있음을 확인하였다. 본 연구에서 개발한 어린이 예방접종 온톨로지는 예방접종과 관련된 다양한 요인을 포함하고 있어 앞으로 소셜미디어상의 예방접종 관련 데이터를 분석해 부모의 예방접종에 대한 감정, 태도, 경험, 지식을 명확히 파악하는 데 활용될 수 있을 것이다.

제5절 결론

전 세계적인 감염병의 지속적인 유행으로 예방접종의 중요성은 날이 갈수록 강조되고 있으며, 이는 특히 면역력이 약한 어린이에게 더욱 중요하다. 우리나라 예방접종률은 감염병 퇴치 수준 이상으로 매우 높은 편이지만 예방접종률에 영향을 줄 수 있는 예방접종에 대한 부정적인 감정들이 소셜미디어를 통해 확산되고 있어 예방접종률을 유지하기 위해 예방접종에 대한 부모들의 감정, 태도, 경험, 지식을 정확히 파악할 필요가 있다. 소셜미디어에 이러한 예방접종에 대한 부모들의 부정적인 감정을 비롯한 다양한 감정 및 경험들이 게시되고 있어 소셜미디어가 부모들의 예방접종에 대한 감정, 태도, 경험, 지식을 파악하는 데 유용하게 활용될 수 있을 것이다. 그러나 소셜미디어의 게시 글은 비정형 데이터로, 이를 정확하게 분석하려면 개념을 정의하고 개념들 간의 관계를 나타낸 온톨로지가 필요하다. 본 연구에서는 Ontology development 101(Noy &

McGuinness, 2001)을 참고하여 어린이 예방접종에 대한 부모의 감정, 태도, 경험, 지식과 관련된 온톨로지를 개발하였다. 이 과정에 예방접종 실무지침과 예방접종 관련 연구논문을 고찰하여 인적요인, 구조사회문화요인, 정보요인, 환경요인, 예방접종 의도, 예방접종 행위, 행동반응요인과 같은 상위 클래스를 추출하였고 이들 클래스를 기본 틀로 활용하여 관련 용어들을 찾아 분류하고 관계를 정의하였다.

본 연구에서 개발한 어린이 예방접종 온톨로지는 부모와 자녀 간의 개인 관련 요인에서부터 구조적·사회적·문화적·환경적 요인을 모두 포함하고 있으며, 예방접종 이전부터 예방접종을 하는 시점, 예방접종을 하고 난 후의 모든 시기에 관련된 요인을 포함하고 있다. 따라서 본 연구에서 개발한 어린이 예방접종 온톨로지는 예방접종과 관련된 소셜미디어의 데이터를 체계적이고 포괄적으로 수집·분석하여 자녀의 예방접종에 대한 부모의 감정, 태도, 경험, 지식을 파악하는 데 활용될 수 있을 것이다. 본 연구에서 개발된 온톨로지가 소셜미디어에 게시되는 예방접종 관련 모든 용어나 개념들을 포함하고 있는지에 대한 포함력 평가는 시행하지 못하였다. 이러한 제한에도 불구하고 본 연구는 매일 생성되는 방대한 소셜미디어상의 데이터로부터 어린이 예방접종 관련 개념을 분석하는 분석틀을 제공했다는 점에서 의미가 있다.

참고문헌 <<

- 고영만. (2005). 온톨로지와 웹 온톨로지(Ontology and Web Ontology). 성균관대학교 문헌정보학과.
- 고재영, 최유석, 홍정익. (2015). 국가예방접종 정책 국민 인식 및 만족도 경향. 주간 건강과 질병, 8(28).
- 김금순, 김복자, 김경희, 권소희, 김수강. (2007). 학령전기 자녀를 둔 어머니의 추가접종 의도에 미치는 영향요인.
- 김학래. (2010). 시맨틱 웹과 의미적 연결성: 웹 사이언스를 위한 출발점. 정보과학회지, 28(3), pp. 31-40.
- 박명배, 김춘배, 주현실. (2013). 인플루엔자 예방접종률에 영향을 미치는 요인. 한국콘텐츠학회논문지, 13(4), pp. 300-311.
- 오현경. (2015). 만 3세 아동의 DTaP 4차 예방접종에 영향을 미치는 요인. 서울대학교 대학원 석사학위논문.
- 유은지, 김정철, 이춘열, 김남규. (2012). 시맨틱 텍스트 마이닝을 위한 온톨로지 활용 방안. 정보시스템연구, 21(3), pp. 137-161.
- 이무식, 김은영, 김건엽, 이진용, 장민영, 홍지영. (2012). 영유아의 예방접종 및 그 관련요인. 농촌의학·지역보건, 37(4), pp. 233-245.
- 이석구. (2013). 2013 전국 예방접종률 조사.
- 이석구, 전소연. (2015). DTaP, b형 헤모필루스 인플루엔자, 폐렴구균 추가(4차) 접종률에 영향을 미치는 요인. 한국모자보건학회지, 19(2), pp. 265-278.
- 이종구, 최원석. (2008). 우리나라의 백신정책. 감염과 화학요법, 40(1), pp. 15-22.
- 이현우. (2010). 예방접종 안전성 이슈에 대한 커뮤니케이션 전략 연구.
- 임은실, 이경자, 천의영, 임미란. (2006). 아동의 예방접종에 미치는 영향 요인. 지역사회간호학회지, 17(2), pp. 283-294.
- 임정우, 김창휘, 이원배, 강진한. (2006). 예방접종에 대한 부모의 인식에 관한 조사. 대한소아과학회지. 49(3), pp. 251-257.

- 정미은. (2007). 영유아 예방접종 인터넷 커뮤니티의 백신 위해도 정보습득 및 지식에 관한 연구, 가톨릭대학교 보건대학원 석사학위논문.
- 정인숙. (2004). 모자보건수첩보유 디피티 추가 예방접종에 대한, 인지와 디피티 추가 예방접종 실천간의 관계연구. *Community Nursing*, 15(1).
- 정혜실. (2016). 청소년 우울 관련 소셜데이터 수집과 분석을 위한 온톨로지 개발 및 평가. 서울대학교 대학원 석사학위논문.
- 진선미, 이석구. (2007). 도시지역 초등학교 입학생의 예방접종 실태와 적기 예방접종 관련 요인. *한국모자보건학회지*, 11(1), pp. 44-53.
- 질병관리본부. (2013). 예방접종 대상 감염병의 역학과 관리(제4판 수정판).
- 차혜경, 류언나, 박성희. (2012). 예방접종 불이행 부모의 예방접종 의도에 영향을 미치는 요인. *보건교육·건강증진학회지*, 29(1), pp. 89-96.
- 최윤경, 김건엽, 이무식, 나백주, 김은영, 심영빈, 김영택. (2008). 보건소 예방접종 영유아의 적기접종률 분석. *한국모자보건학회지*, 12(2), pp. 181-198.
- 최인영, 정미은, 최순, 김석일. (2007). 온라인 커뮤니티에 따른 영·유아 예방접종에 대한 정보습득 경로 및 지식수준 비교. *예방의학회지*, 40(4), pp. 291-296.
- 한국인터넷진흥원. (2015). 2015년 인터넷이용실태조사.
- Brunson, E. K. (2013). *The impact of social networks on parents' vaccination decisions*. *Pediatrics*, 131(5), e1397-e1404.
- Centers for Disease Control and Prevention. (2015). *Epidemiology and Prevention of Vaccine-preventable Diseases*(13th edition).
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2), pp.199-220.
- Grüniger, M., Fox, M. S. (1995). *Methodology for the Design and Evaluation of Ontologies*.
- He, Y., Cowell, L., Diehl, A. D., Mobley, H. L., Peters, B., Ruttenberg, A., Xiang, Z. (2009, July). *VO: vaccine ontology*. In The 1st International Conference on Biomedical Ontology (ICBO 2009)

Nature Precedings.

- Katz, I. T., Ware, N. C., Gray, G., Haberer, J. E., Mellins, C. A., Bangsberg, D. R. (2010). *Scaling up human papillomavirus vaccination: a conceptual framework of vaccine adherence*. *Sexual health*, 7(3), pp. 279-286.
- Kennedy, A., LaVail, K., Nowak, G., Basket, M., Landry, S. (2011). *Confidence about vaccines in the United States: understanding parents' perceptions*. *Health Affairs*, 30(6), pp. 1151-1159.
- Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M., Paterson, P. (2014). *Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007-2012*. *Vaccine*, 32(19), pp. 2150-2159.
- Noy, N. F., McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880*, Stanford, CA.
- Public Health England. (2013). *Immunisation Against Infectious Disease*.
- Pulse, U. G. (2014). *Understanding Public Perceptions of Immunisation Using Social Media*.
- Salathé, M., & Khandelwal, S. (2011). *Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control*. *PLoS Comput Biol*, 7(10), e1002199.
- Sturm, L. A., Mays, R. M., & Zimet, G. D. (2005). *Parental beliefs and decision making about child and adolescent immunization: from polio to sexually transmitted infections*. *Journal of Developmental & Behavioral Pediatrics*, 26(6), pp. 441-452.
- UNICEF. (2013). *Tracking anti vaccination sentiment in eastern european social media networks*.