



■ 연구보고서 2014-05

보건복지 빅데이터 효율적 관리방안 연구

송태민·진달래·박대순
박현애·안지영·김정선

【책임연구자】

송태민 한국보건사회연구원 연구위원

【주요저서】

빅데이터 분석방법론

한나레아카데미, 2013(공저)

보건복지연구를 위한 구조방정식 모형

한나레아카데미, 2012(공저)

【공동연구진】

진달래 한국보건사회연구원 연구원

박대순 한국보건사회연구원 선임전문원

박현애 서울대학교 간호대학 교수

안지영 인제대학교 교수

김정선 SK 스마트인사이트

연구보고서 2014-05

보건복지 빅데이터 효율적 관리방안 연구

발 행 일 2014년 12월 31일

저 자 송 태 민

발 행 인 최 병 호

발 행 처 한국보건사회연구원

주 소 (339-007)세종특별자치시 시청대로 370

세종국책연구단지 사회정책동 1F~5F

전 화 대표전화: 044)287-8000

홈페이지 <http://www.kihasa.re.kr>

등 록 1994년 7월 1일 (제8-142호)

인 쇄 처 대명기획

정 가 11,000원

© 한국보건사회연구원 2014

ISBN 978-89-6827-149-6 93510

발간사 <<

최근 스마트폰, 스마트TV, RFID, 센서 등의 급속한 보급과 모바일 인터넷과 소셜미디어의 확산으로 데이터량이 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 시장변동 예측과 신사업 발굴 등 경제적 가치창출 사례 및 효과를 제시하고 있다. 빅데이터는 미래 국가 경쟁력에도 큰 영향을 미칠 것으로 기대하여 국가별로는 안전을 위협하는 글로벌요인이나 테러, 재난재해, 질병, 위기 등에 선제적으로 대응하기 위해 우선적으로 도입하고 있다. 특히, 모바일 인터넷과 소셜 미디어의 확산으로 데이터량이 기하급수적으로 증가함으로써 경제적 자산이 될 수 있는 데이터를 분석하여 가치 있는 정보를 찾아내고 활용하는 일이 무엇보다 중요한 시대가 되었다.

한국은 최근 정부3.0과 창조경제의 추진과 실현을 위하여 현 정부의 주요정책과제를 지원하기 위하여 다양한 분야에 빅데이터의 활용가치가 강조되고 있다. 보건복지 빅데이터는 다른 분야 보다 수요자에게 직접적인 영향을 주는 것으로 규모가 방대할 뿐 아니라 효율적 관리와 활용을 통한 경제적 가치는 무한할 것으로 예측되고 있다. 특히, 저출산 고령화가 급속히 진전되고 보건복지서비스가 공급자 중심에서 수요자 중심의 맞춤형 서비스 체계로 변화됨에 따라 대국민 보건의료 서비스와 복지전

달체계를 개선하고 행정비용을 줄이기 위해서 보건복지 분야 빅데이터의 효율적 관리와 활용방안의 마련이 요구된다. 이에 본 연구는 국내외의 보건복지 관련 빅데이터 현황과 기술을 분석하고 수요공급의 예측과 분석 사례를 통하여 수요자 중심의 맞춤형 서비스를 위한 보건복지 빅데이터의 효율적 관리방안을 제시하였다.

본 연구는 한국보건사회연구원의 송태민 연구위원의 주관 하에 원내의 진달래연구원, 박대순 선임전문원, 서울대 간호대학 박현애 교수, 인제대학교 안지영 교수, SKT 스마트인사이트 김정선 부장의 참여로 수행되었다. 연구진은 본 연구에 귀중한 조언을 아끼지 않으신 서울대학교 보건대학 조성일 교수님과 인제대학원대학교 류시원 교수님, 원내의 정보기술 융합센터 정영철 센터장님과 통계정보연구실 장영식 초빙연구위원께 사의를 표한다. 그리고 빅데이터 수집과 분석에 참여한 (주)SKT스마트인사이트 임직원 분들에게 감사하며, 본 연구에 도움을 주신 학계 및 빅데이터 전문가 여러분께 감사한다. 끝으로 본 보고서에 수록된 모든 내용은 우리 연구원의 공식적인 견해는 아니며 어디까지나 참여한 연구진의 의견임을 밝힌다.

2014년 12월

한국보건사회연구원장

최 병 호

목 차

Abstract	1
요약	3
제1장 서론	9
제1절 연구의 배경 및 목적	11
제2절 연구목적	13
제3절 연구내용 및 연구방법	14
제4절 빅데이터의 개념	15
제2장 빅데이터 이론적 배경	19
제1절 빅데이터 전망	21
제2절 빅데이터 관련 기술	25
제3절 보건복지분야 빅데이터 활용	36
제4절 빅데이터 프라이버시 보호방안	45
제3장 국내외 공공 빅데이터 공개 현황	57
제1절 공공데이터 및 빅데이터 국내 관련 조직	59
제2절 빅데이터 관련 인력양성 현황	66
제3절 공개·개방된 빅데이터 현황 및 활용 방안	67
제4장 보건복지 이슈 분석 및 통계제공 현황	101
제1절 보건복지 이슈 분석	103
제2절 보건복지 통계제공 현황	127

제5장 보건복지 빅데이터 수요공급 예측 155

제1절 빅데이터 수용요인 분석 157

제2절 공공과 보건 의료 빅데이터 수요공급 예측 184

제3절 보건복지 정책 수요예측 212

제6장 소셜 빅데이터를 활용한 사회위험요인 예측 227

제1절 비만 빅데이터 수집을 위한 관련 주제 분류 230

제2절 소셜 빅데이터를 이용한 비만 담론 분석 252

제3절 소셜 빅데이터를 이용한 청소년 자살위험 예측요인 271

제7장 보건복지 빅데이터 활성화 전략 및 발전방안 289

제1절 보건복지 빅데이터 활성화를 위한 전략분석 291

제2절 보건복지 빅데이터 발전방안 294

참고문헌 317

부록 337

부록1 보건복지 공공데이터 현황 337

부록2 빅데이터 기술수용모형 조사표 383

부록3 소셜빅데이터 비만 담론 분석 속성별 키워드 분류 386

표 목차

〈표 1-1〉 빅데이터의 4가지 구성요소	16
〈표 2-1〉 국내 빅데이터 시장 현황 및 전망	23
〈표 2-2〉 일본 IT벤드의 빅데이터 이용 동향	24
〈표 2-3〉 빅데이터 기술 및 서비스 현황	25
〈표 2-4〉 빅데이터 처리 프로세스별 기술 영역	26
〈표 2-5〉 빅데이터 수집기술	29
〈표 2-6〉 빅데이터 저장기술	30
〈표 2-7〉 빅데이터 처리기술	31
〈표 2-8〉 빅데이터 분석기술	33
〈표 2-9〉 빅데이터 활용기술	35
〈표 2-10〉 빅데이터 기타기술	35
〈표 2-11〉 국내·외 보건분야 빅데이터 활용 사례	36
〈표 2-12〉 국내·외 복지분야 빅데이터 활용 사례	40
〈표 3-1〉 17부 공공데이터 담당자 및 업무현황	60
〈표 3-2〉 3처 공공데이터 담당자 및 업무현황	62
〈표 3-3〉 17청 공공데이터 담당자 및 업무현황	62
〈표 3-4〉 빅데이터 관련 부서 현황	65
〈표 3-5〉 지자체 빅데이터 관련 부서 현황	66
〈표 3-6〉 국내 빅데이터 관련 학과 현황	67
〈표 3-7〉 영국 Open data DB 현황	70
〈표 3-8〉 영국 Open data 분류체계	71
〈표 3-9〉 영국 정부 부처별 데이터 공개내용	74
〈표 3-10〉 미국 공공데이터 제공 사이트 Open DB 현황	76
〈표 3-11〉 미국 보건관련 공공데이터 제공 사이트 DB 현황	77
〈표 3-12〉 일본의 빅데이터 장기 추진계획 (2019~2021년)	79
〈표 3-13〉 뉴질랜드 Open data DB 현황	80

〈표 3-14〉 캐나다 Open data 주제별 분류에 의한 DB 현황	82
〈표 3-15〉 싱가포르 Open data 주제별 분류에 의한 DB 현황	84
〈표 3-16〉 싱가포르 Open data 제공 기관별 분류에 의한 DB 현황	85
〈표 3-17〉 독일 Open data 주제별 분류에 의한 DB 현황	87
〈표 3-18〉 독일 Open 소스 기반의 어플리케이션 현황	89
〈표 3-19〉 호주 Open data 주제별 분류에 의한 DB 현황	90
〈표 3-20〉 제공주체별 공공데이터 자료 현황	92
〈표 3-21〉 공공데이터 서비스 제공 유형 현황	93
〈표 3-22〉 공공데이터 서비스 제공 데이터 확장자 유형	94
〈표 3-23〉 「공공데이터 포털」에서 보건/복지 분야 Open API 자료 현황	95
〈표 3-24〉 Open API제공자료 중 보건의료분야 서비스 분류	95
〈표 3-25〉 Open API제공자료 중 식품의약품안전 서비스 분류	96
〈표 3-26〉 Open API제공자료 중 노인·청소년 서비스 분류	96
〈표 3-27〉 「공공데이터 포털」에서 공공데이터 활용사례	97
〈표 3-28〉 보건의료분야 활용사례	97
〈표 3-29〉 복지분야 활용사례	98
〈표 3-30〉 식품 의약품 분야 활용사례	99
〈표 4-1〉 연구자료 수집대상	108
〈표 4-2〉 연결정도 중심성 상위 30위 키워드	110
〈표 4-3〉 연도별 상위 키워드 변화	113
〈표 4-4〉 사회보장 부분 학술문헌 이슈분석	126
〈표 4-5〉 보건복지부 통계포털 제공통계 목록	128
〈표 4-6〉 보건복지 주요지표	130
〈표 4-7〉 e-나라지표 보건지표 제공 목록	134
〈표 4-8〉 e-나라지표 복지지표 제공 목록	136
〈표 4-9〉 통계청 보건분야 통계지표	138
〈표 4-10〉 통계청 보건분야 통계제공기관별 제공지표수 현황	139
〈표 4-11〉 통계청 복지분야 통계지표	141

〈표 4-12〉 통계청 복지분야 통계제공기관별 제공지표수 현황	142
〈표 4-13〉 통계청 사회분야 통계지표	143
〈표 4-14〉 통계청 사회분야 통계제공기관별 제공지표수 현황	144
〈표 4-15〉 국민건강보험공단 건강보험 통계DB 현황	145
〈표 4-16〉 건강보험심사평가원 건강보험 통계DB 현황	147
〈표 4-17〉 식품의약품안전처 통계DB 현황	148
〈표 4-18〉 국가암 정보센터 통계DB 현황	149
〈표 4-19〉 한국보건사회연구원 보건복지통계 보건분야 통계지표	151
〈표 4-20〉 한국보건사회연구원 보건복지통계 복지분야 통계지표	152
〈표 5-1〉 기술수용모델 국내 선행연구	161
〈표 5-2〉 연구모델의 개발도구	168
〈표 5-3〉 모형비교	175
〈표 5-4〉 주요 변인들의 기술통계	176
〈표 5-5〉 다중집단 구조모형의 적합도 비교	177
〈표 5-6〉 사용목적 수용자와 활용목적 수용자 집단 간의 구조모형 분석결과	179
〈표 5-7〉 외생변인과 TAM변인의 매개효과 검증	180
〈표 5-8〉 국가별 빅데이터 관련 분야 버즈 현황	190
〈표 5-9〉 산업별 빅데이터 관련 분야 버즈 현황	191
〈표 5-10〉 보건의료 및 공공 분야 빅데이터 버즈 현황	193
〈표 5-11〉 업무별 빅데이터 기술 활용 현황	194
〈표 5-12〉 업무별 빅데이터 기반 활용 현황	196
〈표 5-13〉 빅데이터 수요자와 공급자에 미치는 영향요인	197
〈표 5-14〉 보건의료 및 공공 분야 빅데이터 수요자와 공급자에 미치는 영향요인	198
〈표 5-15〉 빅데이터 수요공급 예측모형에 대한 이익도표(감성요인)	200
〈표 5-16〉 빅데이터 수요공급 예측모형에 대한 이익도표(기반요인)	202
〈표 5-17〉 빅데이터 수요공급 예측모형에 대한 이익도표(업무요인)	204
〈표 5-18〉 보건의료분야 빅데이터 수요공급 예측모형에 대한 이익도표(감성요인)	206
〈표 5-19〉 공공분야 빅데이터 수요공급 예측모형에 대한 이익도표(감성요인)	208

〈표 5-20〉 보건의료분야 빅데이터 수요공급 예측모형에 대한 이익도표(기반요인)	209
〈표 5-21〉 공공분야 빅데이터 수요공급 예측모형에 대한 이익도표(기반요인)	211
〈표 5-22〉 보건복지 관련 버즈 현황	219
〈표 5-23〉 공공기관별 보건복지 정책 및 이슈 버즈 현황	221
〈표 5-24〉 민간기관별 보건복지 정책 및 이슈 버즈 현황	222
〈표 5-25〉 보건복지 정책 수요에 미치는 영향요인	223
〈표 5-26〉 보건복지정책 수요 예측모형에 대한 이익도표	226
〈표 6-1〉 비만 영역 분류틀	236
〈표 6-2〉 인구학적 특성 영역 분류틀	238
〈표 6-3〉 위험요인 영역 분류틀	239
〈표 6-4〉 증상 및 징후 영역 분류틀	240
〈표 6-5〉 합병증 영역 분류틀	241
〈표 6-6〉 진단 영역 분류틀	242
〈표 6-7〉 치료 영역 분류틀	245
〈표 6-8〉 예방 영역 분류틀	247
〈표 6-9〉 비만 온톨로지 분류에 따른 영역수준	250
〈표 6-10〉 소셜 빅데이터 활용한 비만 담론 분석 조사설계	253
〈표 6-11〉 연도별 비만 부위별 언급 키워드 현황	260
〈표 6-12〉 연도별 비만 치료 방법(약물치료) 언급 키워드 현황	261
〈표 6-13〉 연도별 비만 치료 방법(운동치료) 언급 키워드 현황	262
〈표 6-14〉 연도별 비만 치료 방법(식사치료) 언급 키워드 현황	263
〈표 6-15〉 비만에 대한 긍·부정 감성분석 결과(키워드 분석)	264
〈표 6-16〉 다이어트에 대한 긍·부정 감성분석 결과(키워드 분석)	267
〈표 6-17〉 운동에 대한 긍·부정 감성분석 결과(키워드 분석)	268
〈표 6-18〉 다이어트로 인한 정신적인 증상에 대한 긍·부정 감성분석 결과(키워드 분석) ..	269
〈표 6-19〉 Descriptive Statistics of Factors	279
〈표 6-20〉 Multinomial Logistic of Suicide Causes	281
〈표 6-21〉 Predictive Performance according to Modeling Methods	281

〈표 6-22〉 Profit chart of Predictive models of Suicide	284
〈표 7-1〉 보건복지 빅데이터 서비스 활성화를 위한 SWOT 분석	293
〈표 7-2〉 ‘빅데이터 개인정보보호 가이드라인’의 주요내용	302

그림 목차

[그림 1-1] 빅데이터의 특성과 보건복지부 3.0 추진 전략	17
[그림 2-1] 빅데이터의 시대별 특징	21
[그림 2-2] 2012년도 가트너의 이머징 기술 하이프사이클(Hype Cycle)	22
[그림 2-3] 주요 시장조사 기관 빅데이터 세계시장 규모	23
[그림 2-4] 빅데이터의 진화단계	25
[그림 2-5] 빅데이터 플랫폼	27
[그림 2-6] 빅데이터 수집 기술	29
[그림 2-7] 지역보건의료정보시스템 구성도	40
[그림 2-8] 사회복지통합관리망 구성도	42
[그림 2-9] 사회서비스전자바우처시스템 구성도	43
[그림 2-10] 보육통합정보시스템 구성도	43
[그림 2-11] 개인에 관한 정보와 개인정보 및 프라이버시의 관계	48
[그림 2-12] 개인정보 식별성과 공개성	49
[그림 2-13] 개인정보와 프라이버시권에 해당하는 정보의 관계	50
[그림 3-1] 정부 3.0 목표	59
[그림 3-2] 17부 3차 17청 공공데이터 제공 현황	60
[그림 3-3] 국외 공공데이터 순위: 2013년	68
[그림 3-4] 영국의 빅데이터 전략위원회 조직 구성 체계도	69
[그림 3-5] 영국 Open data 보건분야 분류체계	71
[그림 3-6] 영국 Open data 포털 화면	71
[그림 3-7] 영국 Open data Institute 포털 화면	72
[그림 3-8] 영국 ODI 데이터 인증 페이지	73
[그림 3-9] 미국의 빅데이터 추진전략 및 체계	76
[그림 3-10] 미국 Open data 포털 화면	78
[그림 3-11] 뉴질랜드 Open data 포털 화면	81
[그림 3-12] 캐나다 Open data 포털 화면	83

[그림 3-13] 싱가포르 Open data 포털 화면	85
[그림 3-14] 독일 공공데이터 공개를 위한 '열린정부' 전략	86
[그림 3-15] 독일 공공데이터 공개를 위한 '열린정부' 구현 모델	87
[그림 3-16] 독일 Open data 포털 화면	88
[그림 3-17] 호주 Open data 포털 화면	91
[그림 4-1] 연구자료 수집대상	107
[그림 4-2] 보건의료분야의 키워드 네트워크 분석	112
[그림 4-3] 2009년 보건의료분야 키워드 네트워크 분석	115
[그림 4-4] 2010년 보건의료분야 키워드 네트워크 분석	117
[그림 4-5] 2011년 보건의료분야 키워드 네트워크 분석	118
[그림 4-6] 2012년 보건의료분야 키워드 네트워크 분석	120
[그림 4-7] 2013년 보건의료분야 키워드 네트워크 분석	121
[그림 4-8] 2014년 보건의료분야 키워드 네트워크 분석	123
[그림 4-9] 보건복지부 주요지표 제공기관 현황	133
[그림 4-10] 통계청 e-나라지표 보건지표 제공기관	135
[그림 4-11] 통계청 e-나라지표 복지지표 제공기관	137
[그림 4-12] 통계청 보건분야 통계제공기관별 제공지표 현황	140
[그림 4-13] 통계청 복지분야 통계제공기관별 제공지표 현황	142
[그림 4-14] 통계청 사회분야 통계제공기관별 제공지표 현황	144
[그림 4-15] 국민건강보험공단 건강보험통계DB 제공화면	146
[그림 4-16] 국가암정보센터 통계지표 제공화면	150
[그림 4-17] 한국보건사회연구원 통계지표 제공화면	153
[그림 5-1] 연구모형	175
[그림 5-2] 빅데이터 수요공급 예측모형(감성요인)	200
[그림 5-3] 빅데이터 수요공급 예측모형(기반요인)	202
[그림 5-4] 빅데이터 수요공급 예측모형(업무요인)	204
[그림 5-5] 보건의료분야 빅데이터 수요공급 예측모형(감성요인)	206
[그림 5-6] 공공분야 빅데이터 수요공급 예측모형(감성요인)	208

[그림 5-7] 보건의료분야 빅데이터 수요공급 예측모형(기반요인)	209
[그림 5-8] 공공분야 빅데이터 수요공급 예측모형(기반요인)	211
[그림 5-9] 보건복지 정책 수요 예측모형	225
[그림 6-1] 비만 온톨로지 인구학적 특성	248
[그림 6-2] 비만 관리 서비스 단계적 흐름	249
[그림 6-3] 비만 온톨로지 비만의 진단, 예방과 치료 프로세스	249
[그림 6-4] 비만 대상자의 상태에 따른 치료 또는 예방 진단	250
[그림 6-5] Ontology for obesity management	251
[그림 6-6] 연도별 버즈 현황: 2012-2013	254
[그림 6-7] 연도별 버즈 점유율 현황: 2012-2013	255
[그림 6-8] 2011년 버즈량과 순계정 비율	256
[그림 6-9] 2011년 전체 문서수와 문서작성계정 수	256
[그림 6-10] 2012년 버즈량과 순계정 비율	257
[그림 6-11] 2012년 전체 문서수와 문서작성계정 수	257
[그림 6-12] 2013년 버즈량과 순계정 비율	258
[그림 6-13] 2013년 전체 문서수와 문서작성계정 수	258
[그림 6-14] 연도별 트위터 유형 분석	259
[그림 6-15] 연도별 비만 부위별 언급 키워드 트렌드	260
[그림 6-16] 연도별 비만관련 운동치료 언급 키워드 트렌드	262
[그림 6-17] 비만에 대한 감성분석 결과	264
[그림 6-18] 비만에 대한 감성분석 결과(SNS 언급)	265
[그림 6-19] 다이어트에 대한 감성분석 결과(SNS 언급)	267
[그림 6-20] 운동에 대한 감성분석 결과(SNS 언급)	268
[그림 6-21] 다이어트로 인한 정신적인 증상에 대한 감성분석 결과	269
[그림 6-22] Predictive model of risk factors for Suicide	283
[그림 7-1] 보건복지통계 관리 전담조직(가칭 데이터연계센터)	304
[그림 7-2] 청소년 자살징후 포착을 위한 소셜 분석 시스템 구성도(안)	310
[그림 7-3] 소셜 빅데이터 분석 절차 및 방법(자살버즈 분석 사례)	312
[그림 7-4] 소셜 빅데이터와 정형 빅데이터 매칭	313

Abstract <<

Efficient management of Big Data on Health & Welfare

Big data which can play a role in improving our national power, were introduced to cope with serious incidents such as terrorism, a natural disaster, disease, and global factors which pose a threat to the national security of nation. In order to address social issues, big data were used in various countries. Big data were utilized by different technology in various regions such as the U.S., the U.K., Japan, Australia and the E.U. countries. All of them consider it as essential state policy for their future.

Emphasizing the importance of utility of big data related to supporting the Government's flagship policy and creative economy, this study aim to analyze method and current status of big data in the fields of health and welfare. The study suggested the way to manage big data effectively by analyzing case study and forecasting supply and demand of the data to provide consumer-directed Services.

Chapter 1 indicates the need of big data and the definition and concept of big data. Chapter 2 includes method and technology related with big data with classifying different field of

2 보건복지 빅데이터 효율적 관리방안 연구

big data into collecting, storing, processing, analyzing and utilizing. Chapter 3 explains the status of publishment of public big data. Chapter 4 includes the status of statistics data being provided and analysis of the issue related to health and welfare. Chapter 5 and 6 explain supply and demand of big data, and the societal risky factors by utilizing social big data. Chapter 7 suggests the strategies to promote to use big data and the ideas for improvement.

1. 연구의 배경 및 목적

빅데이터는 미래 국가 경쟁력에도 큰 영향을 미칠 것으로 기대하여 국가별로는 안전을 위협하는 글로벌 요인이나 테러, 재난재해, 질병, 위기 등에 선제적으로 대응하기 위해 우선적으로 도입하고 있다. 국내외에서 사회 전반의 다양한 문제에 빅데이터 기술을 활용하고 있다. 미국, 일본, 영국, 호주, 유럽연합 등 각 지역의 특징을 가지면서도 빅데이터 전략을 미래의 핵심적 국가 과제로 삼고 있다는 점에서 공통점을 가지고 있다. 샌프란시스코 경찰청은 범죄 예방 시스템을, 싱가포르의 테러 및 전염병을 관리하기 위한 스캐닝 시스템을, 영국은 The Foresight Horizon Scanning Centre를 설립·운영하여 비만대책 수립, 잠재적 위험관리(해안침식, 기후변화), 전염병 대응 등 사회 전반의 다양한 문제에 빅데이터 기술을 활용하고 있다. 국외 뿐 아니라 국내에서도 빅데이터 자료를 수집하여 체계적으로 대응할 수 있는 예측 시스템을 설계하였다. 한국인체자원은행네트워크는 16개 병원을 통해 36만명의 인체자원을 확보하여 질병지표 및 발굴 및 질병조기진단을 위해 활용하고 있다. DNA LINK는 질병관리 및 개인의 유전체 염기서열 분석으로 맞춤형 건강진단 서비스 제공하는 유전자 분석시스템 제공하고 있다. 한국정보화진흥원은 2012년 자살로 언급된 빅데이터를 수집하여 온라인 자살관련 Buzz의 발생 패턴에 따라 보다 체계적으로 대응할 수 있는 자살예방체계를 설계할 수 있다는 가능성을 보였고, 한국보건사회연구원(www.kihasa.re.kr)에서는 2012년 구글 검색트렌드를 활용하여 자살검색량과 자살률간의 관계를

밝힌바 있다.

본 연구에서는 국내외의 보건복지 관련 빅데이터 현황과 기술을 분석하고 수요공급의 예측과 분석사례를 통하여 수요자 중심의 맞춤형 서비스를 위한 보건복지 빅데이터의 효율적 관리방안을 제시하고자 한다.

2. 주요 연구결과

제1장 ‘서론’에서는 정부3.0과 창조경제의 추진과 실현을 위하여 현 정부의 주요정책 과제를 지원하기 위하여 다양한 분야에 빅데이터의 활용가치가 강조됨에 따라 국내외 빅데이터 관련 문헌 분석과 웹사이트 검색을 통한 빅데이터 도입 배경과 필요성, 국내외에서 정의하고 있는 빅데이터 개념에 대하여 설명하였다.

제2장 ‘빅데이터 이론적 배경’에서는 빅데이터의 시대별 특징, 빅데이터 세계 시장 규모를 통해 빅데이터가 산업에 미치는 파급효과를 전망하고, 정보 통신 기술과 함께 국가의 새로운 성장 동력이 될 것으로 예측하였다. 빅데이터의 관련 기술 분야를 수집기술, 저장기술, 처리기술, 분석기술, 활용기술로 구분하여 각 영역별 세부기술에 대하여 제시하였으며, 이러한 기술을 바탕으로 빅데이터를 활용한 국내·외 보건복지분야의 사례를 함께 제시하였다. 그리고 빅데이터는 새로운 비즈니스 기회를 창출하는데 없어서는 안되는 획기적인 자원으로 가치를 더해감에 따라 다양한 목적으로 활용이 되면서 개인정보 유출과 프라이버시에 침해에 대한 우려가 높아지고 있다. 빅데이터 활용에 있어서의 개인정보 보호를 위하여 정보보호관리체계 인증 기술, 영향평가 기술, 프라이버시 보호기술에 대하여 설명하였다.

제 3장 ‘국내외 공공 빅데이터 공개현황’에서는 공공데이터 출연배경

과 공공데이터 담당 부서 및 현황, 국내 빅데이터 관련 인력 양성 현황, 국내 공개·개방된 공공데이터의 지자체별 제공 유형과 국외 빅데이터 현황 및 활용 사례를 제시하였다. 제4장에서는 보건복지 영역별 이슈를 감지하여 미래 보건복지 전략을 수립하기 위한 기초자료로 활용하기 위하여 보건복지분야의 주요 학술지에 게재된 키워드를 분석한 결과를 제시하였다. 상위 키워드의 연도별 응집 구조를 분석한 결과 'Adolescents(청소년)'와 'Social support(사회적 지지)', 'Self-esteem(자존감)', 'Self-efficacy(자기효능감)', 'Smoking(흡연)', 'Suicide(자살)' 등이 서로 연결되어 연구되고 있었으며, 'Suicide(자살)'는 다시 'Depression(우울)', 'Stress(스트레스)', 'Social support(사회적 지지)', 'Self-esteem(자존감)' 등과 함께 연결되어 연구되고 있는 것으로 나타났다. 이는 청소년을 대상으로 한 다양한 방면에서의 연구가 지속적으로 수행되어 오고 있으며, 최근 대두되고 있는 노화, 다문화, 사회적 약자에 대한 연구 등 국내 보건의료 분야의 연구가 한국의 사회적 변화를 잘 반영하고 있다고 해석할 수 있다. 제 5장 '보건복지 빅데이터 수요 공급 예측'에서는 우리나라 기업이 빅데이터 기술을 수용하는 과정에서 발생하는 목적에 따른 차이점을 분석하여 기업에서 빅데이터 기술 도입 시 고려해야 할 사항을 제시하였으며, 우리나라 온라인 뉴스사이트, 블로그 카페, SNS, 게시판 등에서 수집한 소셜 데이터를 바탕으로 국내 공공과 보건의료 분야 빅데이터 수요공급을 예측할 수 있는 의사결정나무를 모형과 보건복지 주요정책에 대한 수요를 예측하여 결과를 제시하였다.

제 6장 '소셜 빅데이터를 활용한 사회위험 요인 예측'에서는 빅데이터 분석 방법론을 적용하여, 우리나라 사회위험요인인 비만, 청소년자살에 대한 위험요인을 예측하였다. 소셜 미디어 상의 비만 주제와 관련된 메시지를 수집하여 분석하기 위한 분석틀을 마련하기 위하여 비만 관리 관련

주제를 분류하고, 비만 관리 온톨로지와 용어체계를 개발하였으며, 이를 기반으로 소셜 미디어 상의 비만 언급에 대한 패턴과 원인 담론을 분석하였다. 분석 결과 ‘비만 및 다이어트’에 관한 버즈 발생은 특정 시즌이 아닌 1년 내내 나타나는 것으로 분석 되었으며, 3개년 동안 해마다 온라인상의 언급량이 증가한 것으로 나타났다. 3개년 동안 SNS의 언급량이 가장 높았으며, 그 다음이 블로그에서 다수 언급된 것으로 나타났다. 청소년들의 다이어트 방법으로는 꾸준한 노력보다 쉽게 살을 뺄 수 있는 약물, 수술 치료 등에 관심을 보이는 경향이 있었으며, 최근 3년간 비만 및 다이어트에 관한 긍정적인 인식의 비중이 지속적으로 높아지는 것으로 나타났다. 마지막으로 제 7장에서는 보건복지 빅데이터 활성화 전략 및 발전방안에 대해 SWOT 분석을 통한 세부 전략을 제안하였다.

3. 결론 및 시사점

빅데이터의 활성화를 위해서는 다양한 분야에서 생성되는 다양한 데이터를 언제 어디서나 활용 할 수 있도록 해야 하며 지식과 정보의 공유가 가능하도록 해야 한다.

우리나라 보건복지 빅데이터의 효율적인 관리를 위해서는 다음과 같은 접근이 필요할 것이다. 빅데이터를 통합적으로 관리하기 위한 범정부 차원의 (가칭)보건복지 빅데이터 관리위원회 운영되어야 한다. 비정형화된 데이터를 관리하고 민간기관과 협조체계가 마련되어야 한다. 국가 차원의 오픈 API(Open Application Programming Interface)의 제공이 지속적으로 이루어져야 한다. 보건복지 빅데이터를 분석 처리할 수 있는 관련 기술개발과 표준화가 마련되어야 한다. 구조화되지 않은 대규모 데이터 속에서 숨겨진 정보를 찾아내는 데이터 사이언티스트(Data

Scientist)의 인재 양성위한 전략이 마련되어야 한다. 보건복지 빅데이터의 개인정보와 기밀정보에 대한 보안정책이 마련되어야 한다.

*주요용어: 보건복지, 빅데이터, 사회위험예측, 데이터마이닝, 다변량분석



제1장 서론

제1절 연구의 배경 및 목적

제2절 연구목적

제3절 연구내용 및 방법

제4절 빅데이터 개념

제1절 연구의 배경 및 목적

최근 스마트폰, 스마트TV, RFID, 센서 등의 급속한 보급과 모바일 인터넷과 소셜미디어의 확산으로 데이터량이 기하급수적으로 증가하고 데이터의 생산, 유통, 소비 체계에 큰 변화를 주면서 데이터가 경제적 자산이 될 수 있는 빅데이터 시대를 맞이하게 되었다(송태민, 2012).

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 시장변동 예측과 신사업 발굴 등 경제적 가치창출 사례 및 효과를 제시하고 있다. The Economist(2010)는 빅데이터를 제대로 활용하면 전 세계가 직면한 환경, 에너지, 식량, 의료문제를 상당부분 해결할 것으로 전망하고 있다. Gartner(2012)는 2011년 이머징 기술 전망에서 빅데이터는 21세기 현재는 기술발생단계(technology trigger)지만 2012년 이후 가장 빠르게 성숙하는 신기술로 전망하고 있다. Mckinsey(2011)는 빅데이터의 활용에 따라 기업/공공분야의 경쟁력 확보와 생산성 향상, 사업혁신/신규사업 발굴의 차이가 생길 것이라고 보고 있으며, 유럽 공공분야에서 연 2,500억 달러의 경제적 효과가 있을 것으로 예측하고 있다. IDC(2014)가 최근 발간한 「국내 빅데이터 스토리지 시장 전망」보고서에 따르면 향후 5년간 연평균 성장률이 31.3%로 증가해 2018년 1087억원 규모에 이를 것으로 전망하고 있다. Policy Exchange(2012)는 영국에서 빅데이터 도입 시, 공공무분에서는 연간 160억~330억 파운드를 절감할 것으로

예측하고 있고, CEBR(2012)는 향후 5년간('12~'17년) 영국 산업전체에서 약 2,160억파운드(약 395조원)의 경제적 효과가 발생할 것으로 전망하고 있다.

특히, 빅데이터는 미래 국가 경쟁력에도 큰 영향을 미칠 것으로 기대하여 국가별로는 안전을 위협하는 글로벌요인이나 테러, 재난재해, 질병, 위기 등에 선제적으로 대응하기 위해 우선적으로 도입하고 있다. 샌프란시스코 경찰청은 범죄예방지역 및 시각을 예측하여 범죄를 미연에 방지하기 위한 범죄예방 시스템을 구축하였다. 싱가포르의 테러 및 전염병으로 인한 불확실한 미래를 대비하기 위해 2004년부터 빅데이터를 관리하기 위한 RAHS(Risk Assessment & Horizon Scanning)를 구축하여 운영하고 있다. 영국은 The Foresight Horizon Scanning Centre를 설립·운영하여 비만대책 수립, 잠재적 위험관리(해안침식, 기후변화), 전염병 대응 등 사회 전반의 다양한 문제에 빅데이터 기술을 활용하고 있다. OECD는 빅데이터를 비즈니스 효율성을 제공하는 새로운 자산으로 인식하여 제 15차 WPIIS 회의¹⁾에서 빅데이터의 경제학 측정을 의제로 채택하였다. 한국의 경우 한국정보화진흥원(www.nia.or.kr)은 2012년 자살로 언급된 빅데이터를 수집하여 온라인 자살관련 Buzz의 발생 패턴에 따라 보다 체계적으로 대응할 수 있는 자살예방체계를 설계할 수 있다는 가능성을 보였고, 한국보건사회연구원(www.kihasa.re.kr)에서는 2012년 구글 검색트렌드를 활용하여 자살검색량과 자살률간의 관계를 밝혔다.

한편, 한국은 최근 정부3.0과 창조경제의 추진과 실현을 위하여 현 정부의 주요정책과제를 지원하기 위하여 다양한 분야에 빅데이터의 활용가치가 강조되고 있다. 보건의료 부문에서는 바이오센싱, 영상진단, 유전체

1) OECD, 15TH MEETING OF THE WORKING PARTY ON INDICATORS FOR THE INFORMATION SOCIETY, 7-8 June 2011.

분석 등 데이터가 급속하게 증가함에 따라 보건의료 빅데이터의 자원화 및 가치창출을 위한 산업의 생태계 조성을 필요로 하고 있다. 사회복지 분야에서는 사회복지통합관리망 구축 운영에 따른 데이터의 폭증과 공공 부조, 사회보험, 사회서비스 등 사회보장 전반에 대한 데이터 기반 정책 수립과 현안에 대한 선제적 대응을 위해 빅데이터의 연계 및 분석·활용을 시도하고 있다. 보건복지 빅데이터는 다른 분야 보다 수요자에게 직접적인 영향을 주는 것으로 규모가 방대할 뿐 아니라 효율적 관리와 활용을 통한 경제적 가치는 무한할 것으로 예측되고 있다. 특히, 저출산 고령화가 급속히 진전되고 보건복지서비스가 공급자 중심에서 수요자 중심의 맞춤형 서비스 체계로 변화됨에 따라 대국민 보건의료 서비스와 복지전달체계를 개선하고 행정비용을 줄이기 위해서 보건복지 분야 빅데이터의 효율적 관리와 활용방안의 마련이 요구된다.

제2절 연구 목적

본 연구는 국내외의 보건복지 관련 빅데이터 현황과 기술을 분석하고 수요공급의 예측과 분석사례를 통하여 수요자 중심의 맞춤형 서비스를 위한 보건복지 빅데이터의 효율적 관리방안을 제시하는 것으로 구체적인 연구목적은 다음과 같다.

첫째, 국내외 빅데이터의 전망, 기술, 활용, 분석방법, 프라이버시 보호 방안 등을 분석하여 현황 및 문제점을 파악한다.

둘째, 주요국과 우리나라의 공공 데이터 공개현황을 분석하여 공공데이터의 효율적 관리방안을 제시한다.

셋째, 우리나라 보건복지 빅데이터 이슈분석과 통계 제공 현황을 분석한다.

넷째, 국내 보건복지 빅데이터 수요공급을 예측하기 위한 모형을 개발한다.

다섯째, 보건복지 분야 소셜 빅데이터 사례분석을 통하여 보건복지 정책개발과 위험요인 예측을 위한 방안을 제시한다.

여섯째, SWOT 분석을 실시하여 빅데이터 활용을 위한 세부전략을 도출하고 보건복지 빅데이터 활성화 전략 및 발전방안을 제시한다.

본 연구는 보건복지 분야의 빅데이터 기술과 정책 수요를 예측함으로써 실효성 있는 보건복지 빅데이터 정책수립을 가능하게 하고 빅데이터 연계를 통한 사전예측과 위험징후에 대한 선제적 대응체계의 구축이 가능할 것으로 본다. 또한 소셜 빅데이터의 다양한 분석방법론을 제시함으로써 관련 인재양성 및 서비스 산업 발굴을 통한 일자리 창출에 기여할 것으로 본다.

제3절 연구내용 및 연구방법

본 연구는 총 8개의 장으로 구성하였다.

제1장 ‘서론’에서는 국내외 문헌분석과 웹사이트 검색을 통하여 국내외 빅데이터의 도입 배경과 필요성 그리고 빅데이터의 개념에 대해 설명하였다. 제2장 ‘이론적 배경’에서는 문헌분석, 웹사이트 검색, 해외출장 등을 통하여 국내외 빅데이터 전망, 기술, 활용, 분석방법, 프라이버시 등 빅데이터의 전반적인 자원과 활용방안에 대해 살펴보았다. 제 3장 ‘국내외 공공 빅데이터 공개현황’에서는 문헌분석과 웹사이트 검색을 통하여 공공데이터 출연배경과 빅데이터 관련부서 현황, 국내외 공개·개방된 빅데이터 현황 및 활용방안에 대해 살펴보았다. 제4장 ‘보건복지 이슈분석 및 통계제공 현황’에서는 네트워크분석, 문헌분석, 기관별 보건복지 통계정보시스템 분석을 통하여 국내 주요학술지에 게재된 보건복지 관련 논문들의 키워드를 분류하여 주요 이슈를 분석하였고 통계제공 현황을

분석하여 보건복지 통계 관리방안을 제시하였다. 제5장 ‘보건보건 빅데이터 수요공급 예측’에서는 우리나라 빅데이터 수용요인 분석과 빅데이터 분석방법론을 적용하여 우리나라 온라인 뉴스사이트, 블로그 카페, SNS, 게시판 등에서 수집한 소셜 데이터를 바탕으로 국내 공공과 보건의료 분야 빅데이터 수요공급을 예측하였다. 그리고 소셜 빅 데이터를 활용하여 보건복지 정책 수요를 예측하였다. 제 6장 ‘소셜 빅데이터를 활용한 사회위험요인 예측’에서는 빅데이터분석방법론을 적용하여, 우리나라 사회위험요인인 비만과 청소년 자살에 대한 위험요인을 예측하였다. 제 7장 ‘보건복지 빅데이터 활성화 전략 및 발전방안’에서는 SWOT분석에서 도출된 세부전략에 대한 활성화 방안을 제시하였다.

제4절 빅데이터의 개념

빅데이터(Big Data)는 Wikipedia(2013. 5. 30.)에서 ‘기존 데이터베이스 관리도구로 데이터를 수집·저장·관리·분석의 역량을 넘어서는 대량의 정형 또는 비정형 데이터 세트 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술’로 정의하고 있다. Gartner(2012)는 더 나은 의사결정, 시사점 발견 및 프로세스 최적화를 위해 사용되는 새로운 형태의 정보처리가 필요한 대용량, 초고속 및 다양성의 특성을 가진 정보자산으로 정의하고 있으며, 국가정보화전략위원회에서는 ‘대용량 데이터를 활용, 분석하여 가치 있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술’이라고 정의하고 있다(국가정보화전략위원회, 2011). 삼성경제연구소는 ‘빅데이터란 수십에서 수천 테라바이트 정도의 거대한 크기를 갖고 여러 가지 다양한 비정형 데이터를 포함하고 있으며, 생성, 유통, 소비가 몇 초에서 몇 시간

단위로 일어나 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합으로 대규모 데이터와 관계된 인력, 조직, 기술 및 도구(수집, 저장, 검색, 공유, 분석, 시각화 등)까지 모두 포함하는 개념'으로 정의하고 있다(함유근·채승병, 2012). 이와 같은 정의를 살펴볼 때 빅데이터란 엄청나게 많은 데이터로 양적인 의미를 벗어나 데이터 분석과 활용을 포괄하는 개념으로 사용되고 있다(송태민, 2012).

한편, 정부3.0의 효과적인 추진과 생애주기별 맞춤형 보건복지 및 국민 행복 실현을 위한 보건복지분야 빅데이터의 효율적 활용방안 모색하기 위하여 보건복지분야 빅데이터 추진방안이 마련되었다.

정부 3.0은 공공정보를 적극 개방·공유하고, 부처 간 칸막이를 없애고 소통·협력함으로써 국정과제에 대한 추진동력을 확보하고 국민 맞춤형 서비스를 제공함과 동시에 일자리 창출과 창조경제를 지원하는 새로운 정부운영 패러다임을 의미하고 있다. 빅데이터의 주요 특성은 일반적으로 3V(Volume, Variety, Velocity)를 기본으로 2V(Value, Veracity)나 1C(Complexity)의 특성을 추가하여 설명하고 있다. 특히, 보건복지분야에서는 국민의 생명과 직결되는 정보를 다루고 있어 빅데이터에 있어 가치(Value)와 신뢰성(Veracity)은 매우 중요하다고 할 수 있다.

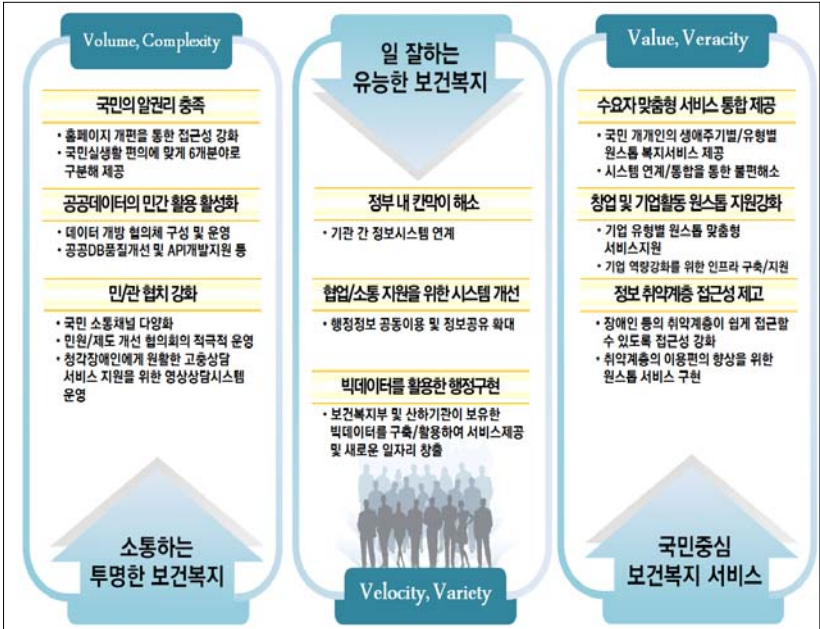
〈표 1-1〉 빅데이터의 4가지 구성요소

구분	주요내용
규모의(Volume) 증가	기술적인 발전과 IT의 일상화가 진행되면서 디지털 정보량이 기하급수적으로 폭증하여 제타바이트(ZB)시대로 진입
다양성(Variety) 증가	로그기록, 소셜, 위치, 소비, 현실데이터 등 데이터 종류의 증가와 멀티미디어 등 비정형화된 데이터 유형의 다양화
복잡성(Complexity) 증가	구조화되지 않은 데이터, 저장방식의 차이, 중복성 문제, 데이터 종류 확대, 데이터 관리 및 처리의 복잡성 심화
속도(Velocity) 증가	사물정보(센서, 모니터링), 스트리밍 정보 등 실시간 정보의 증가로 데이터 생성과 이동 속도가 증가, 대규모 데이터 처리와 정보의 활용을 위한 데이터 처리 및 분석 속도가 중요

자료: 정지선(2011), 新가치창출 엔진, 빅데이터의 새로운 가능성과 대응 전략.

빅데이터의 특성(5V, 1C)은 보건복지부 3.0의 추진 전략과 유기적인 연관성이 있다(오미애, 2014). [그림 1-1]과 같이 보건복지부 3.0의 ‘소통하는 투명한 보건복지’는 빅데이터의 이용 활성화를 위해 공공 데이터를 적극 개방함으로써 활용 가능한 자료가 복잡하고(Complexity), 양이 매우 방대(Volume)해진다. 보건복지부 3.0의 ‘일 잘하는 유능한 보건복지’는 빅데이터를 활용한 과학적 행정 구현으로 다양한(Variety) 정보의 결합이 가능하고, 정부운영시스템 개선으로 인한 자료의 축적 속도(Velocity)가 빠르다. 보건복지부 3.0의 ‘국민중심 보건복지 서비스’는 빅데이터 분석결과를 기초로 수요자 맞춤형 서비스 통합을 제공함으로써 신뢰성 있는(Veracity) 새로운 가치(Value)를 창출한다.

[그림 1-1] 빅데이터의 특성과 보건복지부 3.0 추진 전략





제2장

빅데이터 이론적 배경

제1절 빅데이터 전망

제2절 빅데이터 관련 기술

제3절 보건복지분야 빅데이터 활용

제4절 빅데이터 프라이버시 보호방안

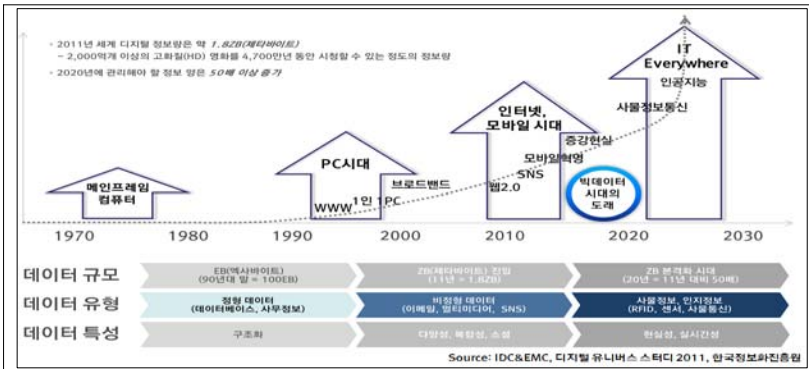
2

빅데이터 이론적 배경 <<

제1절 빅데이터 전망

[그림 2-1]과 같이 스마트폰, 스마트TV, RFID, 각종 센서의 급속한 보급과 모바일 인터넷, 소셜미디어, u-Health서비스의 확산으로 데이터가 폭발적으로 증가함에 따라 혼돈과 잠재적인 가능성이 공존하는 빅데이터 시대가 도래하게 되었다(정지선, 2011).

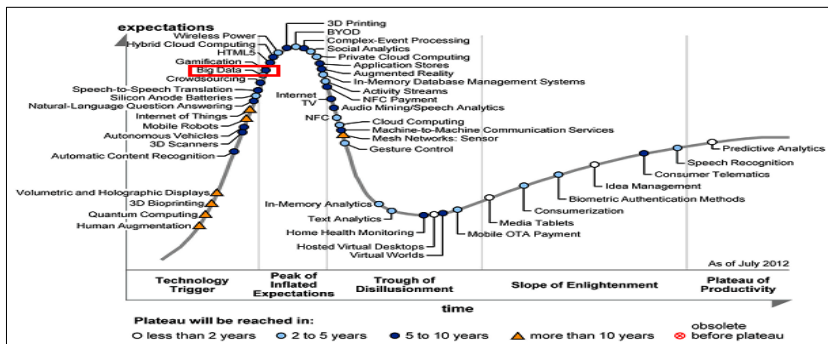
[그림 2-1] 빅데이터의 시대별 특징



빅데이터는 국내외 금융, 유통, 통신, 제조, 서비스, 공공, 의료 등 다양한 분야의 산업에서 활용되고 있다. 그러나 국내에서는 아직 관련기업의 인식수준은 낮은 편이며 실제로 활용하고 있는 사례도 드문 것으로 나타나고 있다(테이코산업연구소, 2013). 미국은 기업의 70%정도는 빅데이터 분석을 통해 수요자, 시장, 파트너, 비용, 인력 운용 측면에서 새로운

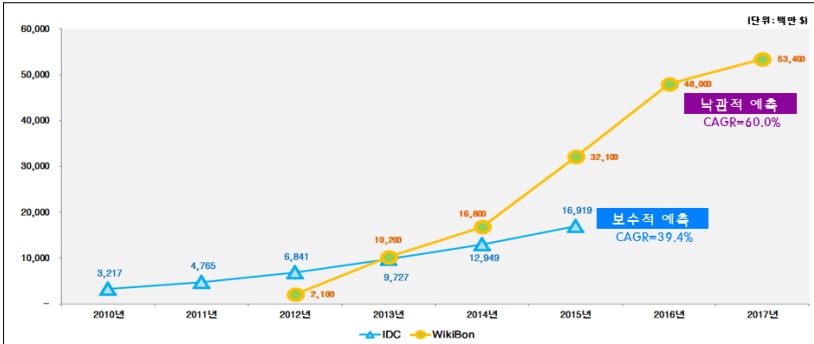
사실들을 발견할 수 있을 것으로 보고 있다(TDWI, 2011). 일본은 2012년 현재 빅데이터는 발전단계이며 빅데이터를 이용하는데 있어 문제가 되는 요인으로 인재의 확보를 들고 있고, 획기적인 분석방법이 개발되면 크게 성장할 가능성이 있다고 보고 있다(独立行政法人 情報処理推進機構, 2012). 가트너는 2011년 이머징 기술 전망에서 ‘빅데이터’를 새롭게 포함시키고 앞으로 주목해야할 기술로 소개하면서 빅데이터는 현재 기술 발생단계이며(Technology Trigger) 향후 2-5년 후에 성숙할 것이며, 2012년 가장 빠르게 성숙하는 신기술로 전망하고 있다(Gatner, 2011/2012)(그림 2-2 참조).

[그림 2-2] 2012년도 가트너의 이머징 기술 하이프사이클(Hype Cycle)



전 세계 빅데이터 시장은 2010년 32억달러에서 2015년 169억~321억 달러로 향후 5년간 연평균 39%~60% 성장을 예측하고 있다(KISTI, 2013)(그림 2-3 참조). <표 2-1>은 국내 빅데이터 시장 현황 및 전망에 관한 것으로 국내 빅데이터 시장은 2015년 약 2억 6,320만달러, 2020년 약 9억달러(한화 1조원)에 이를 것으로 예상되며, 향후 빅데이터 기술이 전 산업 영역으로 확장 적용될 것을 감안한다면 빅데이터 관련 시장규모는 예상치를 상회할 것으로 전망하고 있다(KISTI, 2013).

[그림 2-3] 주요 시장조사 기관 빅데이터 세계시장 규모



〈표 2-1〉 국내 빅데이터 시장 현황 및 전망

(단위: 백만 \$)

구분	2010	2011	2012	2013	2014	2015	연평균성장률
Servers	0.6	0.8	0.9	1.1	1.5	2.0	27.6%
Storage	3.8	6.8	12.5	18.9	24.0	29.0	50.0%
Networking	3.9	4.6	7.1	9.8	12.2	14.4	25.3%
Software	16.9	20.4	29.1	37.8	47.3	65.6	22.9%
Services	42.8	46.2	69.7	95.7	123.2	152.2	23.6%
합계	68.0	78.8	119.3	163.3	208.2	263.2	25.1%
세계시장비중	2.1	1.7	1.7	1.7	1.6	1.6	

자료: 방송통신 산업 통계연보(2012), ITSTAT(정보통신산업진흥원), IDC(2012)자료를 기반으로 KSTI추경(방송통신위원회 및 한국정보통신협회, 지식경제부 및 한국전자통신산업진흥회가 제공하는 IT 산업 통계(ITSTAT)를 기반으로 국내 서버(중대형 컴퓨터), 스토리지(저장장치), 네트워크 장비, 소프트웨어, IT 서비스 시장 규모를 조사 한 후 빅데이터 부문 예산 진입 비율을 적용하여 산출함)

국내외를 막론하고 많은 IT업체가 빅데이터의 활용과 지원서비스에 주력하고 있다. 미국의 IT벤더는 2009년 이후 빅데이터 관련 기업을 인수 합병하거나 제품과 서비스를 계속해서 투입하고 있다. 일본은 2010년 이후 주요 IT사업자가 Hadoop의 통합(intergration) 사업을 전개하고 있고, 복합 이벤트 처리기술의 개발에 경쟁을 벌이고 있다(송태민, 2013).

24 보건복지 빅데이터 효율적 관리방안 연구

〈표 2-2〉 일본 IT벤드의 빅데이터 이용 동향

회사명	시기	주요 시책
히타치제작소	2011년 11월	클라우드 컴퓨팅사업 ‘HarmoniousCloud’ 서비스에 ‘빅데이터 활용 서비스’와 ‘스마트 인프라 서비스’ 메뉴를 추가했다.
일본IBM	2011년 11월	일본 기업의 글로벌 진출과 현지 비즈니스 강화를 지원하기 위해, 고객의 방대한 정보를 경영전략에 활용하기 위한 기반을 제공하는 클라우드 서비스를 개시했다. 분석대상은 중국어, 영어, 스페인어, 불어, 독일어 등 11개 언어이다.
NEC	2011년 11월	센서 등에서 수집된 빅데이터를 실시간으로 분석하여 사용자에게 제공함과 동시에 분석 시스템에서 사용하는 분산 스토리지의 소비전력을 기존 대비 2/3로 줄일 빅데이터 처리 기반기술을 개발했다.
일본HP	2011년 12월	기업의 빅데이터 활용을 위한 IT 인프라 구축을 지원하는 솔루션인 ‘HPHadoopHBase 서비스’를 발표했다. 이 솔루션은 ‘Hadoop’에 각종 서비스를 원스톱으로 제공한다.
후지츠	2012년 1월	빅데이터를 활용하기 위한 기반 클라우드 서비스인 ‘데이터 활용 기반 서비스’를 제공하기 시작했다. 대량의 센싱 데이터를 수집·축적·통합하고 실시간으로 처리하고 일괄 처리에 의한 장래예측 등을 실시하는 ‘정보관리 통합서비스’, ‘통신제어 서비스’ 등을 제공한다.
NTT데이터	2012년 1월	비정형 데이터를 활용하는 분석 솔루션을 제공하기 위하여 미국 MarkLogic사와의 협력에 합의했다. 이 제휴로 비정형 데이터 분석에 효과적인 ‘네이티브형 XML 데이터베이스’를 활용한 빅데이터 활용 솔루션을 제공할 수 있게 되었다.

자료: 独立行政法人 情報処理推進機構(2012). くらしと経済の基盤としてのITを考える研究会報告書, つながるITがもたらす豊かなくらしと経済. p. 136.

이와 같이 많은 국가에서 빅데이터가 산업에 미치는 파급효과를 전망함에 따라 빅데이터는 정보통신기술과 함께 국가의 새로운 성장동력이 될 것으로 보여진다(김정선 외, 2014).

제2절 빅데이터 관련 기술

정보통신기술 주도권이 인프라, 기술, SW 등에서 데이터로 이동됨에 따라 빅데이터의 역할은 분석과 추론(전망)의 방향으로 진화되어 가치창출의 원천요소로 작용하고 있다(정지선, 2011).

[그림 2-4] 빅데이터의 진화단계



자료: 정지선(2011), 「新가치창출 엔진, 빅데이터의 새로운 가능성과 대응 전략」.

국내 빅데이터 시장은 <표 2-3>와 같이 데이터를 수집, 저장, 관리, 표현하는 인프라 SW제품으로 하둡과 맵리듀스 기술 등이 있다. 통계, 데이터마이닝, 기계학습, 패턴인식 등을 위한 분석 SW제품으로 IBM, 솔트룩스 등에서 개발한 기술이 있다. 서버, 스토리지, 운영체제, 데이터베이스 등 여러 소프트웨어와 하드웨어를 통합하는 어플라이언스는 EMC, IBM, Oracle 등에서 개발된 기술이 있다. 빅데이터 분석, 인프라, 서비스를 동시에 제공하는 서비스 플랫폼으로 Google, Amazon, MS, SKT 등에서 개발된 기술이 있다(KISTI, 2013).

<표 2-3> 빅데이터 기술 및 서비스 현황

구분	인프라 SW제품	분석 SW제품	어플라이언스	서비스 플랫폼
주요 내용	데이터 수집, 저장, 관리, 표현 등 빅데이터 처리 분석 플랫폼을 위한 인프라가 되는 단일 소프트웨어 제품	통계, 데이터 마이닝, 기계학습, 패턴인식 등을 통한 분석을 제공하는 실시간 분석 소프트웨어 제품	여러 가지 빅데이터 쿼리 소프트웨어가 최적화되어 설치된 통합 제품	빅데이터 분석을 위한 인프라와 서비스 동시 제공

구분	인프라 SW제품	분석 SW제품	어플라이언스	서비스 플랫폼
주요 제품 기술	Hadoop, Hbase, Cassandra, NoSQL, MapReduce	EMC Greenplum, IBM MPP Data Warehouse, Oracle Big Data Appliance, Teradata Aster Data, 솔트룩스 트루스토리, 사이람 NetMetrica, 넥스알 NDAP, 그루터 BAAS	EMC Greenplum, IBM MPP Data Warehouse, Oracle Big Data Appliance, Teradata Aster Data,	Google BigQuery, Amazon AWS, MS Azure, social metrics, 다음소프트, SKT 스마트인사이트

자료: 2012 중소기업 기술로드맵, 「빅데이터 처리분석 플랫폼」, 한국과학기술정보연구원

빅데이터 기술은 ‘생성→수집→저장→분석→표현’의 처리 전 과정을 거치면서 요구되는 개념으로 분석기술과 인프라는 <표 2-4>과 같다.

<표 2-4> 빅데이터 처리 프로세스별 기술 영역

구분	영역	개요
소스	내부데이터	database, file management system
	외부데이터	file, multimedia streaming
수집	크롤링(crawling)	검색엔진의 로봇을 이용한 데이터 수집
	ETL(Extraction, Transformation, Loading)	소스데이터 추출, 전송, 변환, 적재
저장	NoSQL database	비정형 데이터 관리
	storage	빅데이터 관리
	servers	초경량 서버
처리	mapreduce	데이터 추출
	processing	다중업무처리
분석	NLP(Neuro Linguistic Programming)	자연어처리
	machine learning	기계학습을 통해 데이터 패턴 발견
	serialization	데이터 간의 순서화
표현	visualization	데이터를 도표나 그래픽으로 표현
	acquisition	데이터 획득 및 재해석

자료: Pete Warden(2011). *Big Data Glossary*: O'Reilly Media.

빅데이터 분석기술은 통계, 데이터마이닝, 기계학습, 자연어처리, 패턴 인식, 소셜네트워크 분석, 비디오·오디오·이미지 프로세싱 등이 해당된다. 빅데이터의 활용, 분석, 처리 등을 포함하는 인프라에는 BI, DW, 클라우드 컴퓨팅, 분산데이터베이스(NoSQL), 분산 병렬처리, 하둡(Hadoop)²⁾ 분산파일시스템(HDFS), MapReduce 등이 해당된다(Pete Warden, 2011; 장상현, 2012; 송태민 2012에서 재인용). 그리고 다양한 데이터 소스에서 수집된 빅데이터를 처리·분석하여 지식을 추출하고 이를 기반으로 지능화된 서비스를 제공하기 위해서는 [그림 2-5]과 같이 빅데이터 플랫폼이 필요하다(황승구 외, 2013).

[그림 2-5] 빅데이터 플랫폼



빅데이터 관련기술은 크게 수집기술, 저장기술, 처리기술, 분석기술, 활용기술, 기타기술로 정의할 수 있다(김정선 외, 2014). 본 연구에서는 수집기술로는 ‘크롤링, 카산드라, 로그수집기, 센싱, RSS, Open API’로 정의하였다. 저장기술은 ‘데이터웨어하우스, RDB, 클라우드, X86, 디스

2) 하둡은 대용량 데이터 처리 분석을 위한 대규모 분산 컴퓨팅 지원 프레임워크로 하둡 분산파일시스템(HDFS)과 분산처리를 위한 맵리듀스가 핵심요소이며 그외 분산DB인 Hbase, 검색엔진(Nutch), 쿼리 언어(Pig) 등을 포함한다.

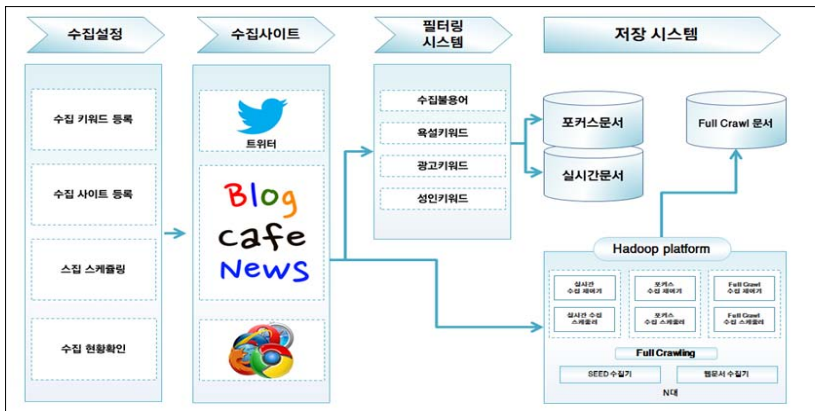
크, 스토리지, NoSQL, SAN'으로 정의하였다. 처리기술은 '가상화, 맵리듀스, 스케일아웃, 어플라이언스, 하둡, 인-메모리, Hbase, R, 드레멜, 퍼콜레이터, 너치, 인텍싱, 스톰, 하둡분산파일시스템'으로 정의하였다. 분석기술은 'NDAP, 기계학습, 네트워크, 시멘틱웹, 온톨로지, 패턴인식, EDW, 데이터마이닝, 텍스트마이닝, 오피니언마이닝, 웹마이닝, 현실마이닝, 소셜네트워크분석, 클러스터분석, 통계적분석, 음성인식, 영상인식, AR, AI, Mahout, ETL, 알고리즘, 프리젤'로 정의하였다. 활용기술은 'BI, 인포그래픽스, SaaS, PaaS, IaaS, DaaS'로 정의하였으며 '아파치, 자바, HTML5'는 기타기술로 정의하였다.

1. 수집기술

기존 검색포털이나 통신사에서 수행하고 있는 빅데이터 수집기술[(주)SKT 스마트 인사이트 빅데이터 수집 방식]은 [그림 2-6]과 같다. 빅데이터 수집기술은 크게 데이터 크롤러인 웹로봇이 매일 뉴스와 트위터를 방문하여 데이터를 수집하여 해당 기관의 DB에 보관하는 Full Crawl 방식이 있고, 새로운 토픽(예를 들면: 자살)이 설정되면 수집조건에 따라 웹로봇이 설정된 사이트를 방문하여 데이터를 수집하는 Focus Crawl 방식이 있다. 데이터 수집에는 ETL(Extraction Transformation Loading)과 크롤링엔진(Crawling Engine)을 사용한다. ETL은 다양한 소스시스템으로부터 필요한 데이터를 추출하여 변환작업을 거쳐 저장하거나 분석을 담당하는 시스템으로 전송 및 적재하는 모든 과정을 포함한다. 크롤링엔진은 로봇이 거미줄처럼 얽혀있는 인터넷 링크를 따라다니며 방문한 사이트의 모든 페이지의 복사본을 생성함으로써 문서를 수집한다. 하둡 분산파일시스템을 이용하여 수천 대의 노드들을 연결하여 수 페타바이트급의 저장용량을 제공하고, 다이나모(Dynamo) 분산데이터 관

리시스템은 데이터를 분할하여 노드들을 배치함으로써 대용량의 데이터를 관리할 수 있다(황승구 외, 2013, p86; 송태민·송주영, 2013에서 재인용). 인메모리컴퓨팅은 데이터베이스 자체를 메모리에 올려서 입출력을 빠르게 하여 데이터의 분석과 저장, 제공을 빠르게 지원한다.

[그림 2-6] 빅데이터 수집 기술



빅데이터 수집기술로는 <표 2-5>과 같이 크롤링, 카산드라, 로그수집기, 센싱, RSS 등이 있다.

<표 2-5> 빅데이터 수집기술

영역	내용
크롤링 (Crawling)	주로 검색엔진의 웹 로봇을 이용하여 SNS, 뉴스, 웹 정보 등 조직 외부 측, 인터넷에 공개되어 있는 웹 문서(정보) 수집 [1]
카산드라 (Cassandra)	분산 시스템에서 방대한 분량의 데이터를 처리할 수 있도록 디자인 된 오픈 소스 데이터베이스 관리 시스템 [2]
로그수집기 (Log collector)	조직 내부에 존재하는 웹서버의 로그 수집, 웹 로드, 트랜잭션 로그, 클릭 로그, DB 로그 데이터 등을 수집 [1]
센싱 (Sensing)	각종 센서를 통해 데이터 수집 [1]

영역	내용
RSS (Really Simple Syndication or Rich Site Summary)	데이터의 생산, 공유, 참여 환경인 웹2.0을 구현하는 기술로 필요한 데이터를 프로그래밍을 통해 수집 [1]
Open API (Open Application Program Interface)	서비스, 정보, 데이터 등을 어디서나 쉽게 이용할 수 있도록 운영체제와 응용프로그램 사이의 통신에 사용되는 언어나 메시지 형식의 개방된 API로 데이터 수집 방식을 제공 [1]

자료: [1] 신신애·김근은(2013). 빅데이터 기술분류 및 현황. 한국정보화진흥원 빅데이터전략센터. 2013.12

[2] 최성곤·오진태·장종수. 미래 지식 서비스를 위한 빅데이터 처리의 기술적 요구사항. 한국정보기술학회지 제10권 제3호, pp.5-10.

2. 저장기술

빅데이터 저장기술로는 <표 2-6>과 같이 데이터웨어하우스, RDB, 클라우드, X86, 디스크(스토리지), NoSQL, SAN 등이 있다.

<표 2-6> 빅데이터 저장기술

영역	내용
데이터웨어하우스 (DW, Data Warehouse)	사용자의 의사 결정 지원을 위하여 다양한 운영 시스템에서 추출, 변환, 통합, 요약된 데이터베이스 [4]
RDB (Relational DataBase)	관계형데이터를 저장, 수정하고 관리할 수 있는 데이터베이스로 SQL 문장을 통하여 데이터베이스의 생성, 수정 및 검색 등 서비스 제공 [5]
클라우드 (Cloud)	인터넷 기반의 컴퓨팅 기술로 무형의 형태로 하드웨어, 소프트웨어 컴퓨팅 자원을 제공하는 기술 [6]
X86	인텔이 개발한 마이크로프로세서 계열을 이르는 말로 이들과 호환되는 프로세서들에서 사용한 명령어 집합 구조들을 통칭 [7]
디스크, 스토리지 (Disk, Storage)	데이터와 명령어를 저장하기 위해 사용하는 장치로 빠른 속도 및 안정성을 강화한 하드디스크 기반의 저장장치
NoSQL (Not only SQL)	클라우드 환경에서 발생하는 빅데이터를 효과적으로 저장, 관리하는 비정형 데이터 관리를 위한 데이터 저장 기술 [1]

영역	내용
SAN (Storage Area Network)	대규모 네트워크 사용자들을 위하여 디스크 어레이, 테이프 라이브러리, 옵티컬 주크박스 등과 같은 서로 다른 종류의 데이터 저장장치를 관련 서버와 함께 연결하는 특수목적용 스토리지 전용 네트워크 [8]

자료: [4] http://ko.wikipedia.org/wiki/데이터_웨어하우스(검색일자: 2014. 9. 29)

[5] 양현철·김진철·신신애·김배현. 빅데이터 활용 단계별 업무절차 및 기술활용 매뉴얼 V1.0. 미래창조과학부·한국정보화진흥원. 2014.05.

[6] http://ko.wikipedia.org/wiki/클라우드_컴퓨팅(검색일자: 2014. 9. 29)

[7] <http://ko.wikipedia.org/wiki/x86>(검색일자: 2014. 9. 29)

[8] <http://www.terms.co.kr/SAN.htm>(검색일자: 2014. 9. 29)

3. 처리기술

빅데이터 처리기술로는 <표 2-7>과 같이 가상화, 맵리듀스, 스케일아웃, 어플라이어스, 하둡, 인메모리, 에이치베이스, R, 드레멜, 퍼클레이터, 너치, 인텍싱, 스톱, 하둡분산파일시스템 등이 있다.

<표 2-7> 빅데이터 처리기술

영역	내용
가상화 (Virtualization)	물리적 컴퓨팅 자원을 논리적으로 나누어 사용자에게 서로 다른 서버, 운영체제 등의 장치로 보이게 하는 기술 [9]
맵리듀스 (MapReduce)	분산 시스템 상에서 데이터 추출, 대용량 데이터를 병렬처리로 지원하기 위하여 구글이 제안한 분산처리 소프트웨어 프레임워크 [10]
스케일아웃 (Scale Out)	서버의 대수를 늘려 트래픽을 분산, 처리 능력을 향상시키는 방식
어플라이언스 (Appliance)	데이터웨어하우징을 위해 서버, 스토리지, 운영체제, 데이터베이스, BI, 데이터마이닝 등 여러가지 하드웨어(HW)와 소프트웨어(SW)가 최적화의 상태로 통합된 장비 [2]
하둡 (Hadoop)	분산 시스템 상에서 대용량 데이터 처리 분석을 위한 대규모 분산 컴퓨팅을 지원하는 자바기반 소프트웨어 프레임워크 [10]
인메모리 (In-Memory)	메모리 상에 필요한 데이터와 이의 인덱스를 메모리에 저장하여 처리하는 기법 [10]
에이치베이스 (Hbase)	컬럼 기반의 데이터베이스로 대규모 데이터 처리를 위한 분산 데이터 저장소 [11]

영역	내용
R	• 통계계산 및 시각화를 위해 R 언어와 개발환경을 제공하며 이를 통해 기본적인 통계 기법부터 모델링, 최신 데이터 마이닝 기법까지 구현 및 개선 가능한 오픈소스 프로젝트 [1]
드레멜 (Dremel)	• 빠른 속도로 쿼리를 수행하여 대용량 데이터를 분석하는 분산 처리가 지원되는 기술 [13]
퍼콜레이터 (Percolator)	• 구글의 검색 엔진에서 검색 인덱스를 작성하기 위해 채택된 기술 [14]
너치 (Nutch)	• 자료와 정보를 검색하는 크롤러 [15]
인덱싱 (Indexing)	• 대량의 데이터를 유형이나 연관성 등 일정한 순서에 따라 체계적으로 정리하여 특정 정보를 쉽게 발견하기 위한 기법 [16]
스톡 (Storm)	• 다양하게 분산되어 있는 정보들로부터 메타 데이터의 추출, 통합, 저장, 관리 및 활용을 위한 기반 구조, 응용 프레임워크, 개발 방법론을 제공하는 데이터 추론 플랫폼 [17]
하둡 분산 파일 시스템 (HDFS, Hadoop Distributed File System)	• 이기종 간의 하드웨어로 구성된 클러스터에서 대용량 데이터 처리를 위하여 개발된 분산 파일 시스템으로 분산된 서버의 로컬 디스크에 파일을 저장하고 파일의 읽기, 쓰기 등과 같은 연산을 운영체제가 아닌 API를 제공하여 처리 [10]

자료: [9] <http://ko.wikipedia.org/wiki/>(검색일자: 2014. 9. 29)

[10] 정지선. 성공적인 빅데이터 활용을 위한 3대 요소: 자원, 기술, 인력. IT & Future Strategy 제3호. 한국정보화진흥원. 2012.04.

[11] 유민석. 빅데이터를 위한 EDW(Enterprise Data Warehouse) 고도화 방안. 정보처리학회지 제19권 제2호. pp54-61. 2012.03.

[13] 서정옥·오영일·김현준·심탁길. 대용량 데이터 병렬 쿼리 엔진 성능 비교 분석. 2012년 한국경영정보학회 추계학술대회. pp.387-391.

[14] <http://blog.acronym.co.kr/394>(검색일자: 2014. 9. 29)

[15] <http://nblack.tistory.com/>(검색일자: 2014. 9. 29)

[16] 김혜선·강부중·양지수·임을규. 침입 탐지 시스템에서 증가형 접두사 인덱싱을 사용한 효율적인 시그니처 탐지 방법. 보안공학연구논문지, 제9권 제1호. pp.15-30. 2012.02.

[17] 김정준·한성근·신인수·한기준. 공간 Hbase:공간 빅 데이터를 위한 Hbase 확장. 정보과학회논문지: 데이터베이스 제40권 제5호. pp.295-304. 2013.10.

4. 분석기술

빅데이터 분석기술로는 <표 2-8>과 같이 NDAP, 기계학습, 네트워크, 시멘틱웹, 온톨로지, 패턴인식, EDW, 데이터마이닝, 텍스트마이닝, 오픈니언마이닝, 웹마이닝, 현실마이닝, 소셜네트워크분석, 클러스터분석, 통계적분석, 음성인식, 영상인식, 증강현실, 인공지능, Mahout, ETL, 알고리즘, 프레젠테이션 등이 있다.

〈표 2-8〉 빅데이터 분석기술

영역	내용
NDAP (NexR Data Analytics Platform)	• 데이터 형태와 관계없이 모든 데이터의 수집, 처리, 저장, 분석 등과 관련한 모든 엔드 투 엔드 서비스를 제공하는 플랫폼
기계 학습 (Machine Learning)	• 인공지능의 한 분야로 패턴 인식 등 컴퓨터가 학습할 수 있도록 알고리즘과 기술을 개발 [18]
네트워크 (Network)	• 개인 또는 집단이 하나의 노드가 되어 각 노드들 간의 상호 의존적인 관계에서 만들어지는 관계 구조
시맨틱 웹 (Semantic Web)	• 분산 환경에서 리소스에 대한 정보와 자원 사이의 관계 및 의미 정보를 온톨로지 형태로 표현하고 이를 자동화 처리 가능토록 하는 프레임워크 [19]
온톨로지 (Ontology)	• 도메인 내에서 공유되는 데이터들을 개념화하고 명시적으로 정의한 기술 [20]
패턴인식 (Pattern Recognition)	• 데이터로부터 중요한 특징이나 속성을 추출하여 입력 데이터를 식별할 수 있는 부류로 분류하는 기법 [21]
EDW	• 기존 DW(Data Warehouse)를 전사적으로 확장한 모델 [22]
데이터 마이닝 (Enterprise Data Warehouse)	• 대용량의 데이터, 데이터베이스 등에서 감춰진 지식, 기대하지 못했던 경향, 새로운 규칙 등의 유용한 정보를 패턴인식, 인공지능 기법 등을 이용하여 데이터간의 상호 관련성 및 유용한 정보를 추출하는 과정 [10]
텍스트 마이닝 (Text Mining)	• 자연어로 구성된 비정형 텍스트 데이터에서 패턴 또는 관계를 추출하여 가치와 의미있는 정보를 찾아내는 기법 [10]
오피니언 마이닝 (Opinion Mining)	• 웹 서버 내의 데이터베이스에 저장되어있는 어떤 주제 혹은 특정한 대상에 대한 사람의 의견을 포함하고 있는 텍스트 속에서 의미를 추출, 분석하여 의미있는 정보를 찾는다 [23]
웹 마이닝 (Web Mining)	• 인터넷 상에서 수집된 정보를 데이터 마이닝 방법으로 분석 통합하는 기법 [10]
현실 마이닝 (Reality Mining)	• 사람들의 행동패턴을 예측하기 위해 현실에서 발생하는 사회적 행동과 관련된 정보를 휴대폰, GPS 등 기기를 통해 얻고 분석하는 기법 [10]
소셜 네트워크 분석 (Social Network Analytics)	• 소셜 미디어를 언어분석 기반 정보추출을 통해 이슈를 탐지하고 흐름이나 패턴 등의 향후 추이를 분석하는 기법 [23]
클러스터 분석 (Cluster Analysis)	• 관심과 취미에 따른 유사성을 정의하여 비슷한 특성의 개체들을 합쳐 서로 다른 그룹의 유사 특성을 발굴하는 기법 [23]
통계적 분석 (Statistical Analysis)	• 전통적인 분석 방법으로 주로 수치형 데이터에 대하여 확률을 기반으로 어떤 현상의 추정, 예측을 검증하는 기법

34 보건복지 빅데이터 효율적 관리방안 연구

영역	내용
음성인식 (Speech Recognition)	• 사람의 음성을 컴퓨터가 해석하여 그 내용을 문자 데이터로 전환하는 처리 기법
영상인식 (Vision Recognition)	• 객체 인식 기술을 이용하여 추출된 전자적 데이터를 여러 목적에 따른 알고리즘을 적용하여 정보를 처리하는 기법
증강현실(AR) (Augmented Reality)	• 가상 현실(VR, Virtual Reality)의 한 분야로 현실세계와 가상세계를 중첩하여 사용자에게 보다 나은 현실감을 제공하는 기법 [26]
인공지능(AI) (Artificial Intelligence)	• 인간의 지능을 연구하고 이를 모델링 하여 이론적 체계를 세우고 다양한 응용 시스템으로 구현하는 기법 [5]
Mahout (Apache Mahout)	• 분산처리가 가능하고 확장성을 가진 기계학습을 기반으로 비슷한 속성분류를 처리하는 기법의 라이브러리
ETL (Extraction, Transformation, Loading)	• 필요한 소스데이터 추출 후 변환을 거쳐 시스템에서 시스템으로 데이터를 이동(추출, 전송, 변환, 적재)시키는 기법 [10]
알고리즘 (Algorithm)	• 컴퓨터 혹은 디지털 대상이 과업을 수행하는 방법에 대한 설명이자 명확히 정의된 한정된 개수의 규제나 명령들의 집합 [29]
프레젠타 (Pregel)	• 정점 중심의 메시지 전달 방식을 이용하는 그래프 처리에 적합한 클라우드 기반 분산 처리 프레임워크 [30]

자료: [18] <http://ko.wikipedia.org/wiki/>(검색일자: 2014. 9. 29)

[19] <http://ko.wikipedia.org/wiki/>시맨틱 웹

[20] 황미녕·이승우·조민화·김순영·최성필·정한민. 연구 개발 트렌드 분석을 위한 기술 지식 온톨로지 구축. 한국콘텐츠학회논문지 '12 Vol. 12 No. 12. pp.35-45.

[21] http://www.aistudy.com/pattern/pr_definition.htm/(검색일자: 2014. 9. 29)

[22] <http://www.dbguide.net/knowledge.db?cmd=view&boardUid=125292&boardConfigUid=19&categoryUid=57/>(검색일자: 2014. 9. 29)

[23] 최성우·김호영·김영국. 빅데이터 기술 및 분석 기법의 연구 동향. 2012년 가을 학술발표 논문집 Vol. 39, No.2(A). pp.194-196

[26] 반경진·류남훈·김경옥·한재정·김웅곤. 증강현실을 위한 가상 공간좌표 생성. 한국콘텐츠학회 2009 춘계종합학술대회. pp.765-769

[29] 오세욱·이재현. 소프트웨어 '페이스북'의 알고리즘 분석-행위자 네트워크 관점-. pp.136-183. 2013.02.

[30] 신은옥·이성민·정연돈. Pregel을 이용한 독립집합 문제의 병렬 분산 알고리즘. 2012 한국컴퓨터종합학술대회 논문집 Vol.39, No.1(C). pp.104-106.

5. 활용기술

빅데이터 활용기술로는 <표 2-9>과 같이 BI, 인포그래픽스, SaaS, PaaS, IaaS, DaaS 등이 있다.

〈표 2-9〉 빅데이터 활용기술

영역	내용
BI (Business Intelligence)	• 신속하고 정확한 비즈니스 의사결정에 필요한 데이터의 수집, 저장, 처리, 분석하는 일련의 기술과 응용시스템 기술의 집합 [10]
인포그래픽스 (Infographics)	• 복잡한 정보, 자료 또는 지식의 시각적 표현 [31]
SaaS (Software as a Service)	• 클라우드 환경에서 동작하는 온라인 오피스 등 서비스로서의 소프트웨어
PaaS (Platform as a Service)	• 클라우드 환경에서 동작하는 어플리케이션이나 서비스가 실행되는 환경을 제공하는 서비스로서의 플랫폼
IaaS (Infrastructure as a Service)	• 클라우드 환경에서 동작하는 OS 및 응용프로그램을 포함한 IT 자원(서버, 스토리지, DB 등)을 제공 가능한 서비스로서의 인프라
DaaS (Desktop as a Service)	• 클라우드 환경에서 동작하는 데스크톱으로 PaaS, SaaS를 결합한 데스크톱 서비스

자료: [31] <http://ko.wikipedia.org/wiki/>(검색일자: 2014. 9. 29)

6. 기타기술

빅데이터 기타 기술로는 〈표 2-10〉과 같이 아파치, 자바, HTML5 등이 있다.

〈표 2-10〉 빅데이터 기타기술

영역	내용
아파치 (Apache)	• 클라이언트 요청을 처리하기 위해 모듈화된 접근을 사용하여 주요 소스코드를 변경하지 않고 서버측 기능을 구현하는 유연하고 확장 가능한 웹 서버
자바 (JAVA)	• 객체 지향적 프로그래밍 언어로 보안성이 뛰어나며 컴파일한 코드는 다른 운영체제에서 사용할 수 있도록 클래스(class)로 제공
HTML5 (Hyper Text Mark-up Language 5)	• 복잡한 애플리케이션까지 제공할 수 있는 웹 애플리케이션 플랫폼으로 진화한 HTML을 개선한 마크업 언어

제3절 보건복지분야 빅데이터 활용³⁾

국내·외 보건분야 빅데이터 활용 사례는 <표 2-11>과 같다.

<표 2-11> 국내·외 보건분야 빅데이터 활용 사례

구분		내용
국 외	미국 국립보건원	- 질병에 대한 관리 및 예측 시스템 구축 (www.1000genomes.org) - Pillbox 프로젝트(pillbox.nlm.nih.gov)
	미국 퇴역군인국	퇴역군인의 전자의료기록(Electronic Health Records) 분석을 통한 맞춤형 의료서비스를 지원하는 빅데이터 분석을 위해 2년간 25개의 DW를 배치하여 퇴역군인에게 의료서비스 제공
	싱가포르 PA	맞춤형 복지서비스 구축을 위한 인종, 나이, 문화, 소득, 연령에 따른 주민 데이터를 수집·분석하여 개인별 맞춤형 서비스 제공
	캐나다 온타리오 공과대병원	인큐베이터 내 미숙아에 대한 데이터 분석하여 병원균 감염 예측 할 수 있는 시스템 개발
	IBM/Wellpoint	- 의사와 다른 진료진들이 환자치료에 이용 할 수 있는 어플리케이션(왓슨)을 개발하여 제공 - 임상실험 및 우수치료사례, 과거 데이터 분석을 통해 환자에게 가장 적절한 치료 방법을 제공하고 최신 정보를 제시
	구글	- 독감 예보 서비스(www.google.org/flutrends/) - 다양한 사용자의 검색어 분석을 통한 정확한 정보 실시간제공
국 내	한국인체자원 은행네트워크	16개 병원을 통해 36만명의 인체자원을 확보하여 질병지표 및 발굴 및 질병조기진단을 위해 활용(kbn.cdc.go.kr)
	분당 서울대학교병원	- 빅데이터 도입을 통해 업무효율성 및 생산성 향상을 위한 임상 의사결정지원 시스템 개발 - 자연어 검색 지원, 의약품 처방 및 조제시 의약품 안전성과 관련 정보 실시간 제공하여 부적절한 약물 사용 사전에 검사할 수 있도록 확대

3) 본 절은 일부 내용은 ‘한국정보화진흥원(2012), 「빅데이터로 진화하는 세상(Big Data 글로벌 선진 사례)」를 분석, 재정리하였으며, ‘송태민, “우리나라 보건복지 빅데이터 동향 및 활용방안”, 과학기술정책, 통권 제192호, 2013, pp. 56-73.’의 내용을 수정·보완한 것임을 밝힌다.

구분		내용
	DNA Link	• 질병관리 및 개인의 유전체 염기서열 분석으로 맞춤형 건강 진단 서비스 제공하는 유전자 분석시스템 제공
	연세대학교 의료원	• u-Health를 활용하여 언제 어디서나 질병 예방에서부터 진단, 치료가 가능한 후(H∞H)헬스케어 시스템 제공
	한국보건복지 정보개발원	• 보건복지부 가족건강사업, 검진사업, 구강보건사업, 노인보건사업, 한의약 공공보건사업, 영양 개선 사업, 건강생활 실천 통합 서비스, 보건소 방문건강관리사업, 만성질환관리사업, 정신보건사업, 감염병 관리사업을 포함하여 총 29종의 보건사업 정보를 관리함

보건 분야의 빅데이터 국외 활용 사례로 미국 국립보건원은 다양한 질병을 연구하기 위해 유전자 데이터를 공유 분석할 수 있는 유전자 데이터 공유를 통한 질병치료체계를 마련하여 주요 관리 대상에 해당하는 질병에 대한 관리 및 예측을 실시하고 있다. 현재 1,700명의 유전자 정보를 아마존 클라우드에 저장하여 누구나 데이터를 이용 가능하게 구축하였다(www.1000genomes.org/). 미국 국립보건원 산하 국립의학도서관에서는 사용자가 요구하는 다양한 약에 대한 정보를 제공하고 제조사와 사용자 간의 쌍방향 상호작용을 통해 약의 정보를 제공하는 Pillbox 프로젝트를 통한 의료개혁을 추진하고 있다. Pillbox 서비스(pillbox.nlm.nih.gov/)로 미국 국립보건원에 접속되는 알약의 기능이나 유효기간을 문의하는 민원 수는 100만 건 이상으로 평균 한 건당 확인하는 소요 비용 50달러를 감안하면 연간 5,000만 달러의 비용 절감효과가 있는 것으로 전망하고 있다.

미국 퇴역군인국(U.S. Department of Veterans Affairs)에서는 퇴역군인의 전자의료기록 분석을 통한 맞춤형 의료서비스를 지원하는 빅데이터 분석을 위해 2년간 25개의 DW를 배치하여 2,200만 퇴역군인에게 의료서비스를 제공하고 있다. 퇴역군인 전자의무기록(EHR)을 분석하여

의사가 개별 환자를 쉽게 진료할 수 있도록 지원하고 있다.

싱가포르 PA(People's Association)는 1,800개 이상의 주민위원회 센터(커뮤니케이션 센터)에서 진행되는 다양한 활동들을 공유하기 위해 주민위원회센터 네트워크 기반의 맞춤형 복지사회를 구현하였다. 싱가포르 PA는 빅데이터 처리를 위하여 다양한 인종, 나이, 문화, 소득, 연령에 따른 주민의 데이터를 수집·분석하여 개인별 맞춤형 서비스를 제공하고 있다.

캐나다 온타리오 공과대병원인 인큐베이터 내 미숙아에 대한 다양한 데이터를 분석하여 병원균 감염을 예측할 수 있는 시스템을 개발하여 미숙아 모니터링을 통한 감염예방 및 예측, 감염징후 등을 조기에 발견하고 다른 미숙아 등에 대한 감염을 예방하며 퇴원 후 무선센서를 이용하여 병원 밖에서도 환자들을 실시간으로 체크를 할 수 있는 시스템을 구축하였다.

IBM과 미국 건강보험회사인 웰포인트(Wellpoint)는 의사와 다른 의료진들이 진단과 환자치료에 이용할 수 있는 애플리케이션(왓슨)을 개발하여 제공하고 있다. 왓슨은 임상실험 및 우수 치료사례 등 과거 데이터를 분석하여 환자에게 가장 적절한 치료방법을 제공하고 최신 정보를 과학적인 방법으로 제시하고 있다.

구글(Google)은 감기와 관련된 검색어 분석을 통하여 독감예보 서비스 제공하고 있다. 구글 독감예보 서비스(구글 플루 트렌드; www.google.org/flutrends/)는 다양한 사용자의 검색어 분석을 통하여 사용자에게 다시 유의미한 데이터로 가공하여 정확한 정보를 실시간으로 제공하고 있다.

보건 분야 국내 활용 사례로 질병관리본부에서 운영하는 한국인체자원은행네트워크(kbn.cdc.go.kr/)는 16개 병원을 통해 36만 명의 인체자

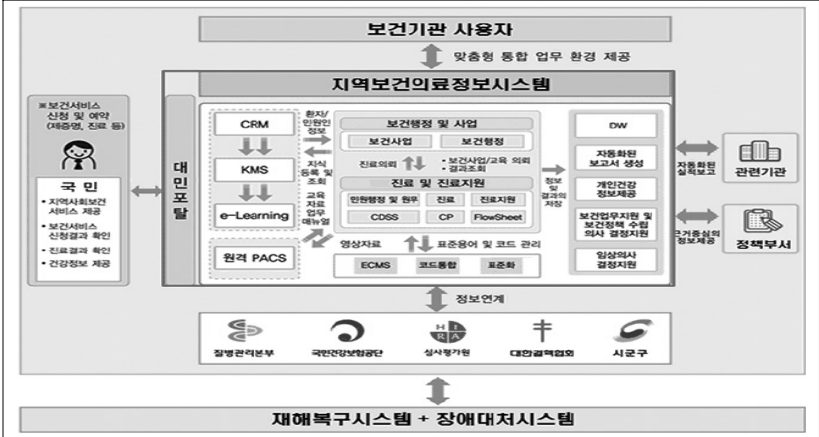
원 확보하여 질병지표 발굴 및 질병조기 진단을 위해 활용하고 있다. 한국인체자원은행네트워크는 생명연구자원의 체계적 수집과 정보 표준화, 정보공유를 통하여 질병의 예방과 진단, 맞춤치료, 신약-신기술을 위한 미래 바이오산업의 신성장동력으로서 기반을 마련하고 있다.

분당서울대병원은 빅데이터 도입을 통해 업무효율성 및 생산성 향상을 위한 임상 의사결정지원시스템을 개발하였다. 임상 의사결정지원시스템은 환자 개인의 특이사항을 입력하여 임상적 의사결정을 지원하기 위한 서비스로 시스템이 도입된 후, 부적절한 용량의 신독성 약물 처방률이 30.6%로 감소하는 효과를 가져왔다. 임상 의사결정지원시스템은 빅데이터를 분석하여 자연어 검색을 지원하고 의약품의 처방과 조제 시 의약품 안정성과 관련된 정보를 실시간으로 제공하여 부적절한 약물사용을 사전에 검사할 수 있도록 확대하고 있다.

DNA Link(dnalink.com/)에서는 질병관리 분석과 개인의 유전체 염기서열 분석으로 맞춤형 건강진단 서비스를 제공하는 유전자 분석시스템을 제공하고 있다. 연세대학교의료원에서는 u-Health를 이용하여 언제 어디서나 질병예방, 진단, 치료가 가능한 후(H∞H) 헬스케어 시스템을 제공하고 있다.

한국보건복지정보개발원에서는 보건복지부의 가족건강사업, 검진사업, 구강보건사업, 노인보건사업, 한의약공공보건사업, 영양개선사업, 건강생활실천통합서비스, 보건소방문건강관리사업, 만성질환관리사업, 정신보건사업, 감염병 관리사업을 포함하여 총 29종의 보건사업 정보를 관리하고 있다(ETRI, 2014)

[그림 2-7] 지역보건의료정보시스템 구성도



국내·외 복지분야 빅데이터 활용 사례는 <표 2-12>와 같다.

<표 2-12> 국내·외 복지분야 빅데이터 활용 사례

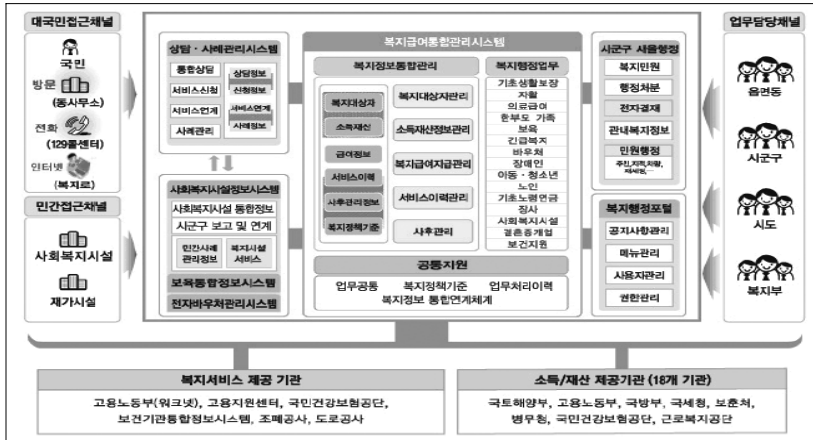
구분		내용
국 외	싱가포르 국가위험 관리시스템	• 2004년부터 빅데이터를 기반으로 위험관리 계획 추진 • 국가위험관리시스템을 구축하여 질병, 금융위기 등의 국가적 위험사항을 수집 및 분석하고 있음(hsc.gov.sg)
	FBI	• 유전자 정보은행 CODIS(Combined, DNA Index System)은 350만개의 DNA분석표가 내장되어 있으며, 유전자 색인 시스템을 활용하여 단시간 범인을 검거하는 체계 구축
	샌프란시스코 경찰청	• 범죄발생지역 및 시간을 예측하여 범죄를 미연에 방지하기 위한 범죄예방시스템을 구축(www.crimemapping.com)
국 내	보건복지부	• 사회복지통합관리망(행복e음)개발하여 수요자 중심의 복지서비스 구현 • 사회서비스전자바우처 시스템 구축
	근로복지공단	• 고객관계관리(CRM)을 구축하여 '찾아가는 서비스'를 통한 맞춤형 서비스를 제공
	한국정보화 진흥원	• 빅데이터 국가전략 분석팀에서 2012년 1월부터 10월 18일까지 자살로 언급된 소셜 데이터를 수집하여 분석하여 청소년 자살과 관련한 패턴 분석과 자살예방체계 설계할 수 있는 가능성을 보임

복지 분야 국외 사례로는 주로 안전과 관련한 빅데이터 활용이 주를 이루고 있다. 싱가포르에서는 국가위험관리시스템(Risk Assessment Horizon Scanning)을 구축하여 질병, 금융위기 등 모든 국가적 위험을 수집 및 분석을 하고 있다. RAHS(hsc.gov.sg)는 2004년부터 빅데이터를 기반으로 한 위험관리 계획을 추진하여 수집된 정보는 시뮬레이션, 시나리오 기법을 통해 분석하여 사전 위험 예측 및 대응 방안을 모색하고 있다. FBI는 유전자 색인 시스템 활용하여 단시간 범인을 검거하는 체계를 구축하고 있다. FBI의 유전자 정보은행 CODIS(Combined, DNA Index System)은 미제사건 용의자 및 실종자에 대한 DNA 정보 1만 3,000건을 포함하여 12만 명의 범죄자 DNA 정보가 저장되고 매년 2,200만 명의 DNA 샘플을 추가하여 범죄 수사에 활용하며 약 350만 개의 DNA 분석표가 내장되어 있다.

샌프란시스코 경찰청은 범죄발생지역 및 시간을 예측하여 범죄를 미연에 방지하기 위한 범죄예방시스템을 구축하였다(www.crimemapping.com). 범죄예방시스템은 과거 범죄를 분석하여 효율적으로 경찰을 배치하고 과거 범죄자 및 범죄유형을 SNS를 통해 지속적으로 관찰함으로써 그와 관련된 조직 및 범죄에 대한 예방을 하고 있다.

복지 분야 국내 활용현황으로는 보건복지부가 사회복지통합관리망(행복e음)을 개발하여 수요자중심의 복지서비스 구현하였다. 사회복지통합관리망은 각종 복지 급여 및 서비스 지원 대상자의 자격 및 이력에 관한 정보를 통합관리하고 지자체 복지업무 처리를 지원하는 중앙집중형 정보시스템으로 크게 복지급여통합관리시스템, 상담사례관리시스템, 사회복지시설정보시스템으로 구성되어 있다. 사회복지통합관리망은 총 123종의 보건복지사업 수혜자 이력 및 정보를 관리하고 있으며, 앞서 정의된 범부처 보건복지사업 296종 중 118종의 사업정보를 관리하고 있다(ETRI, 2014).

[그림 2-8] 사회복지통합관리망 구성도



보건복지부(보건복지정보개발원)의 사회서비스전자바우처시스템은 바우처 사업 대상자 및 제공기관을 효율적으로 관리하기 위해 구축된 정보 시스템으로 크게 바우처 포털시스템, 업무지원시스템, 임신출산진료비시스템으로 구성되어 있다. 사회서비스전자바우처시스템은 보건복지부의 노인돌봄서비스, 장애인활동지원사업, 산모신생아도우미사업, 가사간병 방문사업, 지역사회서비스투자사업, 임신출산진료비사업, 장애아동재활 치료사업, 언어발달지원사업 총 8개 분야의 사회서비스 사업 및 700여종의 지역사회서비스 정보를 관리하고 있다(그림 2-9 참조).

그리고 보육통합정보시스템은 보육전자바우처 아이사랑카드를 관리하는 바우처관리시스템, 지자체 공무원을 위한 행정지원시스템, 어린이집 지원시스템, 아이사랑보육포털, 예약자금관리시스템으로 구성되어 있다. 보육사업 관련 22종의 세부 사업 정보가 보육통합정보시스템을 통해 관리되고 있고, 296종의 범정부 보건복지사업 중 8종의 사업정보가 보육통합정보시스템을 통해 관리되고 있다(ETRI, 2014)(그림 2-10 참조).

[그림 2-9] 사회서비스전자바우처시스템 구성도



[그림 2-10] 보육통합정보시스템 구성도



근로복지공단은 공공부문 고객관계관리(CRM)를 구축하여 '찾아가는 서비스'를 통한 맞춤형 서비스를 제공하고 있다. 한국정보화진흥원 빅데이터 국가전략포럼 분석팀에서는 2012년 1월부터 10월 18일까지 자살

로 언급된 빅데이터 자료를 뉴스(온라인에서 게재되는 214개 웹사이트), 블로그(네이트, 네이버, 이글루스, 다음, 티스토리, 야후), 카페(네이버, 다음, 뽐뿌, 카드고릴라, SLR 클럽), SNS(트위터, 미투데이), 게시판(네이버 지식인, 네이트 지식, 다음 신지식 등) 등에서 수집하여 청소년이 작성했다고 추정되는 6만 9,886건을 분석하였다. 이를 통해 청소년들은 자살과 관련하여 온라인상에 많은 Buzz를 생성하고 있으며 Buzz의 발생패턴에 따라 보다 체계적으로 대응할 수 있는 자살예방체계를 설계할 수 있다는 가능성을 보였다. 한편, 2013년 현재, 보건의로 분야 빅데이터 시범사업으로 진행 중인 과제는 다음과 같다. 첫째, 국민건강 주의예보 시범서비스 구축사업은 국민건강공단의 건강보험 DB와 SNS 정보를 융합하여 홍역, 조류독감, SAS 등 감염병 발생 예측 모델링을 개발하고, 이를 상시 모니터링하여 위험징후 시 주의예보 서비스를 제공한다. 둘째, 빅데이터 기반 의약품 안전성 조기경보 서비스는 유해사례DB, 치료기록, SNS 등을 연계 분석하여, 유의의약품을 추출하고, 이들의 위험도를 예측하여 병의원, 제약회사 및 유관기관 등과 정보공유를 하는 사업이다. 셋째, 보건의로 빅데이터 활용 시범사업은 포털과 질병관리본부 등과 협의된 데이터 외 병원자체 데이터를 활용하여 독감유행 예측, 심실부정맥 예측, 입원 병상배정 최적화, 신종 마약류 인지 및 감시 서비스를 제공한다(송태민, 2013).

제4절 빅데이터 프라이버시 보호방안⁴⁾

빅데이터의 등장과 함께 공공 및 민간기관에서는 다양한 목적으로 폭넓게 활용 되면서 개인의 정보유출과 프라이버시 침해에 대한 우려는 높아지고 있다. 개인에 관련된 정보를 정부나 공공기관 혹은 서비스 제공자가 실시간으로 감시하는 행위는 사회전체의 안전과 편의라는 공통의 가치를 위해 개인은 자신의 프라이버시를 부분적으로 포기한다고 암묵적으로 동의한 것으로 여겨지고 있으나 실제적으로는 개인의 가치를 다양하게 반영한 선택적인 계약에 의한 것이 아니라 일률적인 점이 문제가 되고 있다. 빅데이터의 긍정적 활용을 위해서는 우선 빅데이터의 이용으로 인해 발생할 수 있는 부작용과 이에 대한 대책을 검토하고 개개인의 존중을 기본으로 하는 사회전체의 공정한 규칙을 마련해야 할 것이다. 그 중에서 가장 시급한 문제 중의 하나가 빅데이터의 활용에 있어서의 프라이버시의 보호 방안이다. 본 연구에서는 우리나라와 사회 제도면에서 유사한 일본을 중심으로 빅데이터의 활용을 위한 프라이버시 보호방안에 대해 살펴보고자 한다.

1. 개인정보와 프라이버시 개요

가. 개인정보개요

개인정보의 법률상의 정의로는 ‘개인정보법(제2조 제1호)’과 ‘정보통신망 이용촉진 및 정보보호 등에 관한 법률 제2조 제6호’에서 “살아있는

4) 본 절은 ‘송태민, ‘일본의 빅데이터 프라이버시 보호방안’, 보건복지포럼 통권 210호, 2014. 4’의 내용을 수정·보완한 것임을 밝힌다.

개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다.)를 말한다”로 규정되어 있다.

OECD는 개인정보보호지침(Guidelines on the Protection of Privacy and Transborder Flows of Personal Data)에서 개인데이터(Personal Data)는 식별되거나 식별될 수 있는 개인에 관한 모든 정보를 지칭한다고 정의하고 있다. EU도 개인정보 보호지침(Directive95/46EC)에서 개인데이터(Personal Data)는 식별되거나 식별될 수 있는 자연인에 관한 모든 정보를 지칭 한다고 규정하고 식별가능한 개인이란 직접 또는 간접적으로 신원확인번호, 신체적, 생리적, 정신적, 경제적, 사회적동일성(identity)을 나타내는 요소를 참조하여 그 신원이 확인될 수 있는 사람을 지칭한다고 정의하고 있다.

한편, 미국의 프라이버시법(Privacy Act, 1974)에서의 개인정보는 행정기관이 보유하는 개인기록(Record)에서 개인에 관한 정보(information about an individual)의 개개항목 또는 그 집합을 의미하며 여기에는 개인의 이름, 식별번호, 부호, 지문, 성문 등이 있고 특정 개인과 연결지을 수 있는(Linkable) 정보가 포함되어 있다. 최근에는 스마트폰의 ID와 같이 해당 개인의 고유한 식별자(identifying particular)까지 확대되어 개인 정보로 취급되고 있다. 일본의 ‘개인정보의 보호에 관한 법률’ 상에 정의된 개인정보는 ‘생존하는 개인에 관한 정보로서 해당 정보에 포함된 성명, 생년월일과 그외 개인을 식별할 수 있는 것(타의 정보와 쉽게 결합하여 그에 의해 특정의 개인을 식별할 수 있는 것을 포함한다.)’이라고 규정되어 있어 우리나라의 개인정보법에서와 표현이 유사하다. 위에서 보는 바와 같이 각국에서 개인정보의 법적인 정의는

비슷하며 우리나라 기준으로 개인정보를 열거하면 ‘성명, 주민등록번호를 비롯하여 주소, 성별, 본적, 가족관계와 여권번호, 운전면허증번호 등 신상에 관한 기본정보와 신장, 체중, 장애정도, 생체인식정보, DNA, 혈액형, 지병등과 같이 신체의 특징을 나타내거나 건강의료에 관련된 정보, 소득, 재산, 보험가입현황, 신용정보, 채권, 채무 등의 경제관계 정보, 병역, 직업경력, 사회활동경력, 전과기록 및 법률 위반기록 등의 사회경력 정보, 친구 선후배 애인 등의 인적관계정보, 종교, 취미, 사상, 신조, 가치관, 정치적 성향 등의 내면정보를 비롯하여 기타 통신내역, 위치정보, 음주, 흡연량’ 등의 정보가 포함된다.

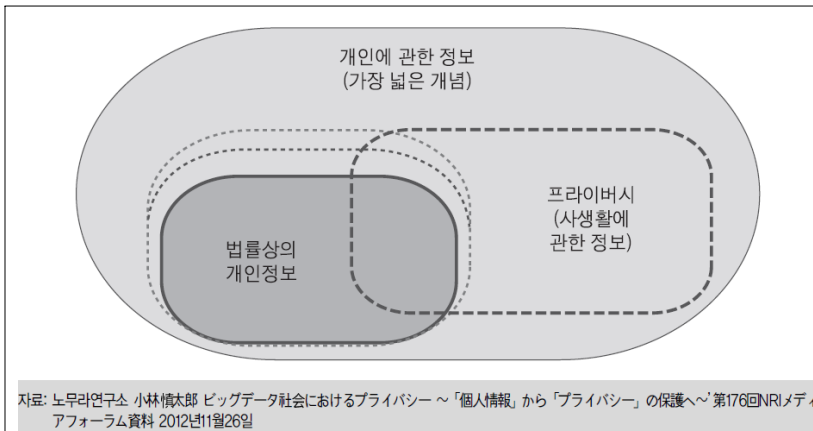
개인정보는 본인 확인 정보로서의 식별정보와 본인에 부속된 속성정보로 대별할 수 있다. 식별정보에는 성명, 주민등록번호, 여권번호와 얼굴사진, 지문, 성문, 홍채, 유전자 등의 생체정보가 있다. 이는 그 자체로서 개인을 식별하거나 특정 지을 수 있는 정보이다. 성명의 경우 동명이인이 존재하는 경우도 있지만 사회통념상 독립적으로 개인을 특정하는 데 사용되고 있음으로 식별정보로서 분류된다. 한편 주소, 생년월일, 성별, 인종, 국적 등은 단독으로 특정 개인을 식별할 수는 없지만 조합에 의해 본인을 식별할 수 있는 준식별 정보이다. 속성정보에는 건강의료정보, 경제상황정보, 개인의 신용정보, 학력과 경력정보 등이 있다. 반면 신용정보와 학력, 경력정보에 포함된 신용카드번호나 학번, 사번과 같은 개인ID 정보는 일반적으로 기본식별 정보와 준식별 정보를 근거로 발행되어 개인을 특정하는 데 사용됨으로 식별정보로 분류할 수 있다.

법률상에서 정의된 ‘개인정보’와 이를 기술하는데 사용된 ‘개인에 관한 정보’를 비교해 보면 후자가 전자를 포함하는 개념임을 내포하고 있다(그림 2-11 참조).

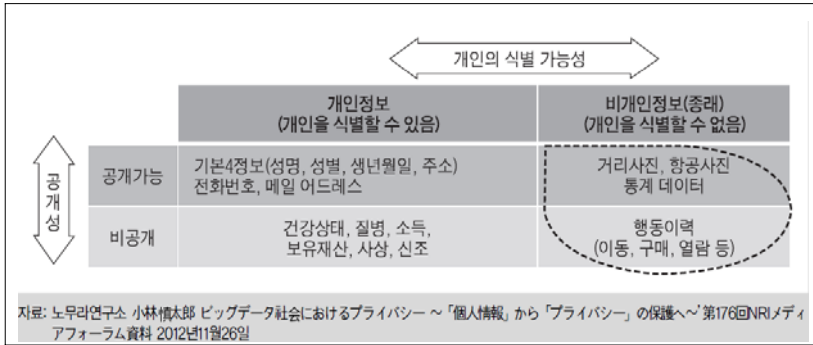
즉, 개인에 관한 정보가 보다 넓은 개념의 집합으로 그 중 일부가 개인

정보인 것이다. 개인에 관한 정보 중에는 개인의 식별로 연결되기가 어려워 법률적으로 개인정보에 포함되지 않는 것들이 있다. 따라서 이러한 정보는 개인에 관한 정보로서의 그 중요성이나 의미와 가치가 적은 것으로 취급되어 왔다. 경우에 따라서는 ‘개인에 관한 정보’에서 제외되었다. 그러나 정보통신 기술의 발전과 인터넷의 등장으로 지금까지 비개인정보로 취급되었던 사항들이 다른 정보와 결합하여 개인을 특정 지을 수 있게 사회가 변화됨에 따라 개인정보가 차지하는 영역이 점차 확대되어 가고 있는 상황이다. 과거에는 흔잡한 길거리의 사진에 촬영된 사람들의 정보는 ‘개인에 관한 정보’에서 제외되었고 도서관 자료의 열람정보와 같은 행동이력 등은 사회적인 상식의 범위에서 개인정보로 간주하지 않았던 사항들이었다(그림 2-12 참조).

[그림 2-11] 개인에 관한 정보와 개인정보 및 프라이버시의 관계



[그림 2-12] 개인정보 식별성과 공개성



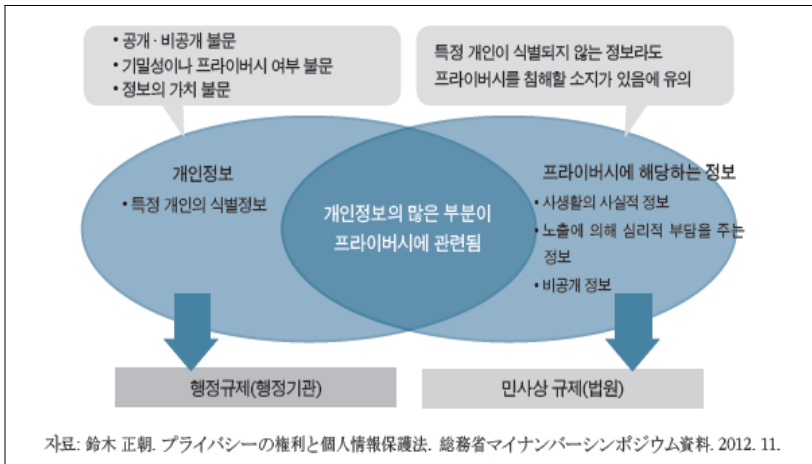
최근 들어 인터넷 상의 프라이버시 보호에 대한 논의와 함께 그러한 정보들이 개인정보의 범주에 포함 되고 있다. 특히, 데이터의 분석기술의 발달과 함께 그러한 경향은 더욱 가속되고 있으며 그동안 의미나 존재 가치가 희박했던 비개인정보들이 활용을 전제로 새로운 가치를 창출하는 빅데이터로서 재평가 받고 있는 것이다.

나. 프라이버시개요

개인에 관한 정보중에는 사생활에 관계되는 프라이버시(Privacy)의 부분이 있다(그림 2-13 참조). 프라이버시의 사전적인 의미로서는 개인의 사생활이나 사적인 일 또는 그것을 남에게 알려지지 않거나 간섭받지 않을 권리를 말하며 사생활(私生活)로 번역하기도 하지만 프라이버시는 권리를 포함한 더 큰 범주에 속한다. 프라이버시를 내용에 따라 분류하면 결정프라이버시(Decisional privacy), 공간프라이버시(Spatial privacy 혹은 Locational privacy), 의도프라이버시(Intentional privacy), 정보프라이버시(Informational privacy), 통신프라이버시

(Communicational privacy), 물리적정신적 프라이버시(Physical and psychological privacy) 등으로 나눌 수 있다.

[그림 2-13] 개인정보와 프라이버시권에 해당하는 정보의 관계



프라이버시는 초기에 ‘혼자 있을 권리(the right to be left alone)’로서 자기 자신에 관한 정보가 남에게 함부로 공개되거나 침해되지 않는 권리의 소극적인 개념에서 정보화의 물결과 함께 ‘자신의 정보를 통제할 수 있는 권리(the right to control information about oneself)’로 발전하게 되었다. 즉, 자신의 정보를 수집, 가공, 유통 및 제공 하는 데에 있어 접근권 및 통제권을 가지고 언제, 무슨 정보를 어느 범위까지 누구에게 유통 시키느냐를 스스로 결정하는 ‘정보의 자기결정권’을 의미하게 된 것이다. 이처럼 프라이버시 권리는 당초의 소극적인 개념에서 시작하여 정보사회에서는 ‘정보의 자기결정권’과 같은 적극적이고 능동적인 권리의 개념으로 바뀐 것이다. 프라이버시권의 법적지위는 유엔에 있어서의 세계인권선언을 비롯하여 각국은 헌법과 법률에 의해 국민의 기본권으로

보호하고 있다. 세계인권선언(Universal Declaration of Human Rights) 12조에는 '어느 누구도 자신의 사생활, 가족, 가정 또는 통신에 대하여 자의적인 간섭을 받지 않으며 자신의 명예와 신용에 대하여 공격을 받지 아니 한다. 모든 사람은 그러한 간섭과 공격에 대하여 법률의 보호를 받을 권리를 가진다'라고 규정하고 있다. 우리나라 헌법에서는 제16조 공간에 대한 프라이버시 제17조 사생활의 비밀과 자유로서의 프라이버시 제18조 통신에 대한 프라이버시를 국민의 기본권으로 보장하고 있다. 이를 침해하였을 때는 민법에서 불법적 행위에 의한 침해로서 제750조 불법행위의 내용에 대한 손해를 배상할 책임이 있고, 제751조(재산이외의 손해의 배상에서) 자유 또는 명예를 해하거나 기타 정신상의 고통에 대해 배상할 책임이 발생한다.

2. 빅데이터 활용에 있어서의 프라이버시 보호방안

빅데이터의 활용에 있어서의 가장 논란이 되고 있는 것은 개인의 각종 기록정보를 수집하여 분석함으로써 생길 수 있는 프라이버시의 침해에 대한 위험성의 문제이다. 이는 외부로부터의 침입에 의한 데이터의 유출의 위험성과 데이터 취급자의 부적절한 업무처리로 인한 무의식적인 공개와 같은 잠재적 위험성 그리고 내부자의 불법적인 열람행위 등에서 비롯되는 위험성을 내포하고 있다. 프라이버시가 노출되거나 개인정보가 악용된 경우 사회적으로 큰 파장이 일어 남은 물론 경제적인 면에서도 손해배상과 대책비용의 발생, 사업활동의 저숙 혹은 신용저하에 따른 매출 감소 등의 비용이 발생한다. 일본 네트워크 보안협회의 보고서에 따르면 2011년에 연간 약 1,551건의 개인정보 유출과 1,900억 엔 약 2조 원 이상의 손해배상이 있다고 보고되고 있다(日本ネットワークセキュリティ

協, 2012). 이러한 문제에 대해 적절히 대처하기 위해 사회적인 큰 틀에 있어서의 법제도적인 대책에서부터 비롯하여 조직적인 대책, 기술적인 대책 등이 검토되고 있다. 법제도의 측면에서는 한국과 일본에서는 개인정보법이 제정되어 있으며 미국에는 옵트아웃을 기반으로 한 ‘소비자 프라이버시 권리장전(Consumer Privacy Bill of Rights)’, 유럽의 옵트아웃을 기반으로 한 ‘EU 데이터보호규정(General Data Protection Regulation)’ 등이 정비되어 있다. 조직적인 대책으로는 법에서 규정한 안전관리 조치를 준수하는 차원에서 정보보호관리체계(ISMS)와 프라이버시 영향평가(PIA) 등의 시스템 운용과 관리상의 각종 기법이 도입되고 있다. 또한 기술적인면에서도 프라이버시보호 데이터마이닝(PPDM: Privacy-Preserving Data Mining) 등 여러 가지 기법이 연구 개발되어 적용되고 있다. 그러나 개인정보보호를 과도하게 중시하면 빅데이터의 활용을 저해하게 될 우려가 있다. 이는 서비스공급자와 이용자의 양자가 상호간에 다양한 효용성을 얻을 수 있는 기회를 놓치는 것이 된다. 이러한 점에서 법제도적, 기술적 대책을 적절히 운용하여 정보주체가 안심하고 자신의 데이터를 제공하게 하고 빅데이터가 적극적으로 활용되게 하는 일이 중요하다.

가. 정보보호관리체계 인증을 통한 프라이버시 보호

정보보호관리체계(ISMS: Information Security Management System)는 기업이나 조직이 정보보호 활동을 체계적이고 지속적으로 수행하기 위해, 보안정책을 수립하고 이에 근거한 계획의 작성, 실시, 운영, 그리고 일정 기간 후의 보안 방침과 계획의 재검토 등을 포함한 전체적인 위험관리를 지속적으로 수행하는 체계를 말한다⁵⁾. 기업이나 조직이

ISMS를 보유하고 유지하고 있는 지에 대해서는 ‘ISMS 인증제도’를 통하여 제도적으로 보증받는다. 이는 제삼자 기관에 의한 인증심사에 의해 정보보안 관리체계의 국제규격인ISO/IEC17799:2000 및 BS7799-2:1999에 입각한 평가를 받는 것이다. ISMS에 요구되는 범위는 ISO/IEC 15408 등이 정하는 기술적인 정보보안 대책의 레벨이 아니라 조직 전체에 있어서 보안관리 체제를 구축·감시하고 위기관리를 실시하는 것이다. ISMS는 개별적인 문제에 대한 기술 대책의 예도 조직관리의 일환으로서 스스로의 위험을 평가하여 필요한 보안레벨을 정하고 계획에 따라 자원을 배분해서 시스템을 운용하는 것이다. 조직이 보호해야할 정보자산에 대해서 기밀성, 완전성, 가용성을 균형있게 유지하고 개선하는 것이 ISMS의 기본 개념이다.

나. 프라이버시 영향평가를 통한 프라이버시 보호

프라이버시 영향평가(PIA: Privacy Impact Assessment)는 개인정보를 수집하여 취급하는 정보시스템의 기획, 구축, 보수, 유지 과정에 있어서 개인정보 제공자의 프라이버시에의 영향을 ‘사전’에 평가하는 일련의 프로세스를 말한다 이를 통하여 잠재적인 프라이버시 침해위험성을 분석하고 대안적인 방법이나 보호방안을 검증하는 등정보시스템의 구축운용을 적절하게 실시할 수 있도록 하는 과정을 포함한다. 설계 단계에서부터 프라이버시 보호대책을 검토함으로써 정보시스템 가동후의 프라이버시 리스크를 최소한으로 억제할 수 있어 향후 시스템의 개보수에 따르는 추가 비용의 발생을 막을 수 있다. PIA는 2008년 4월 ISO22307(Financial services Privacy impact assessment)로서 표준화 되었다. 부제의 타

5) <http://isms.kisa.or.kr/kor/intro/intro01.jsp>. 2013/09/13.

이들에서 의미하는 바와 같이 금융서비스를 제공하는 기업이 고객과 거래처 등의 재무데이터의 처리와 관련된 프라이버시 보호와 리스크에 대처하기 위한 방법론이 정의되어 있다. 그러나 그 내용이 금융관련 서비스에 특화한 것이 아니고 민간부문 및 공공부문을 불문하고 다양한 분야에 적용가능 한 것이다. 또한, 요구사항이 각국의 법체제나 사회제도에 의존하지 않는 최대공약수적인 내용이다. 실제적으로 법령상의 해석이나 의미명과 같은 기술적인 대책만으로는 개인정보의 이용을 정당화하는데 충분한 근거를 마련할 수 없는 상황이 생길 때가 많다. 이런 상황을 해소하는데 있어 PIA의 프로세스로서 활용하는 것이 유효할 수 있다. 즉 개인정보의 보호와 이용의 균형을 도모하기 위해 개인정보의 이용에 수반되는 프라이버시에의 영향 리스크를 평가해서 정보 이용에 의해 초래되는 편익과 비교한다. 이를 토대로 정보주체자를 비롯한 이해관계자들의 의향을 파악하여 이를 반영한 적절한 조치를 강구함으로써 정보이용에 대한 정당성을 확보하는 것이다.

다. 프라이버시 보호기술을 통한 프라이버시 보호

개인과 관계되는 정보를 유통함에 있어서 법제도적인 대책과 조직적인 대책을 완수하기 위해서는 기술적인 대책이 뒷받침되어야 한다. 프라이버시보호를 위해 정보유통에 강한 제약조건을 두게 되면 정보의 가치를 잃게 되거나 손상되어 사회적으로는 물론 정보주체자가 본인에게도 손실을 가져오게 된다. 서비스를 개인화하는데 있어 개인정보는 불가결한 요소이기 때문이다. 이러한 문제를 기술적으로보완하기 위하여 최근 개인정보의 이용과 프라이버시보호의 밸런스를 취하는 방안이 연구되고 있다. 그중에서 프라이버시보호 데이터마이닝에 관한 연구가 대표적으로

관심을 받고 있는 연구이다. PPDM는 개인정보나 기밀정보의 안전성을 유지하고 개인의 프라이버시를 보호하면서 대규모의 데이터로부터 특징이나 규칙성 등을 추출하고 새로운 지식을 발견하는데 활용하기 위한 기술을 총칭하는 것이다. PPDM에는 데이터가 가진 정보의 일부를 누락시키거나 개인을 특정할 수 있는 요소를 삭제·은폐하는 등의 익명화 수법과 통계학적인 처리에 의해 DB에 잡음을 첨가하여 통계적인 성질을 유지하면서 데이터의 누설을 막는 교란수법, 데이터가 가진 정보에 손상을 입히지 않고 당사자 간에 정보를 주고받는 암호화수법 등이 있다. 현재는 정보공개에 일반적으로 활용될 수 있는 익명화 기술이 많이 연구되고 있다. 그러나 데이터의 익명화는 정보의 손실을 초래하여 데이터마이닝(data mining)의 결과에 영향을 미치게됨으로 활용 용도에 따라 프라이버시 보호수준을 고려하여 익명화 처리에 특히 주의를 기울일 필요가 있다. 암호화 수법중에는 데이터를 복수의 그룹에 나누어서 분리 보관하고 그룹사이의 공개키 암호로 데이터를 암호화한 채로 필요에 따라 데이터마이닝(data mining)을 위한 계산을 실시하여 그 결과만을 전체 그룹 사이에 공유하는 비밀계산 기술도 최근 연구되고 있다.



제3장

국내외 공공 빅데이터 공개 현황

제1절 공공데이터 및 빅데이터 국내 관련 조직

제2절 빅데이터 관련 인력 양성 현황

제3절 공개·개방된 빅데이터 현황 및 활용방안

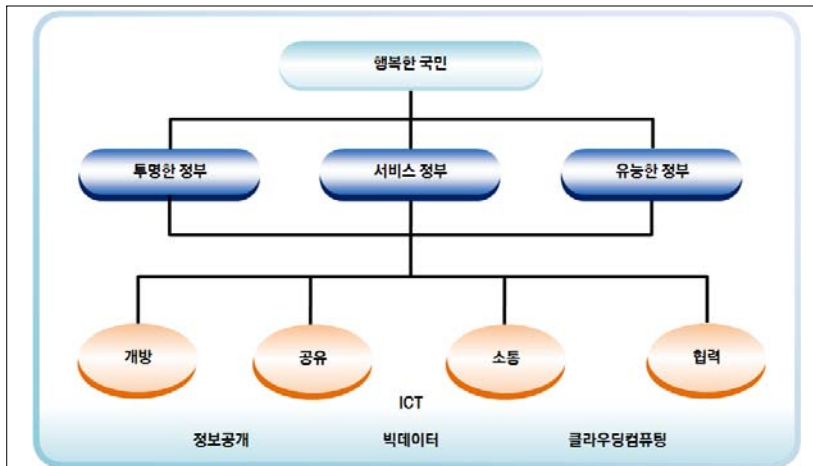
3

국내외 공공 빅데이터 << 공개 현황

제1절 공공데이터 및 빅데이터 국내 관련 조직

공공정보를 적극적으로 개방하고 공유하며 부처 간의 칸막이를 없애 소통하고 협력함으로써 국민 맞춤형 서비스를 제공하고 동시에 일자리 창출과 창조경제를 지원하는 정부 운영 패러다임의 일환으로 정부 3.0이 출범하였다. 정부 3.0이 제기된 배경으로는 정보통신기술(ICT)의 환경 변화로 인하여 정부 내부 및 정부 간 또는 민간과 개방·공유·소통·협력을 통하여 개인 맞춤형 서비스를 제공하고자 하는 것으로 기하급수적으로 증가하는 데이터의 양과 IT발달로 인한 빅데이터의 등장 이유를 들 수 있다.

[그림 3-1] 정부 3.0 목표



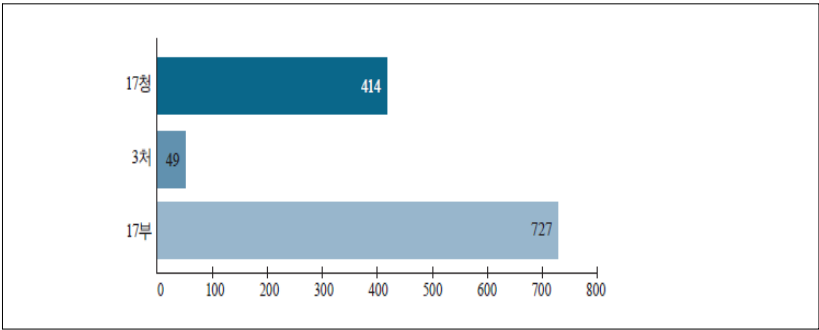
자료: 안전행정부(2014), 정부 3.0 길라잡이

1. 정부 공공데이터 담당부서 및 업무현황

현 정부가 출범하면서 정부조직법에 의하여 정부조직을 15부 2처 18청에서 17부 3처 17청으로 개편되었다. 각 부처별 공공데이터 담당자 및 업무 현황은 <표 3-1>, <표 3-2>, <표 3-3>과 같다.

[그림 3-2]은 17부 3처 17청의 공공데이터 제공 현황이다. 기획재정부, 미래창조 과학부를 비롯한 17개 부처에서는 총 727건의 공공데이터를 서비스 하고 있으며, 법제처, 국가보훈처, 식품의약품안전처에서는 총 49건의 공공데이터를 서비스 하고 있다. 국세청, 관세청, 조달청을 비롯한 17개 청사에서는 총 414건의 공공데이터를 홈페이지를 통하여 서비스 하고 있다.

[그림 3-2] 17부 3처 17청 공공데이터 제공 현황



* 2014.03.20. 기준

<표 3-1> 17부 공공데이터 담당자 및 업무현황

관련부처	부서	연락처	담당업무	제공 Data
기획재정부	기획조정실	044-215-2009	책임관	34
	정보화담당관	044-215-2614	실무담당	
미래창조과학부	정보화담당관	02-2110-2260	책임관	6
		02-2110-2263	실무담당	

관련부처	부서	연락처	담당업무	제공 Data
교육부	기획조정실	044-203-6323	책임관	21
	교육통계과	044-203-6325	실무담당	
		044-206-6321		
외교부	외교정보화담당관	02-2100-7102	책임관	41
		02-2100-7158	실무담당	
		02-2100-7614		
통일부	기획조정실	02-2100-5700	책임관	7
	창조행정담당관	02-2100-5690	실무담당	
법무부	기획조정실	02-2100-3003	책임관	13
	정보화담당관	02-2100-3159	실무담당	
국방부	기획조정실	02-748-6122	책임관	3
	정보체계통합담당관	02-748-5931	실무담당	
안전행정부	공공정보정책과	02-2100-1887	공공데이터	101
		02-2100-1872	이용활성화 사업	
문화체육관광부	기획조정실	044-203-2200	책임관	3
	정보통계담당관	044-203-2260	실무담당	
농림축산식품부	정책기획관	044-201-1301	책임관	11
	정보화담당관	044-201-1402	실무담당	
산업통상자원부	기획조정실장	044-203-5200	책임관	113
	정보화관리담당관	044-203-5552	실무담당	
		044-203-5556		
보건복지부	정보화담당관	044-202-2226	공공데이터 정보화 관련 등	42
		044-202-2227	보건복지 공공데이터 활성화 및 보건복지 빅데이터 추진	
	질병정책과	044-202-2517	암 공공데이터 수립 및 지원 사업	
환경부	기획조정실	044-201-6320	책임관	91
	정보화담당관	044-201-6411	실무담당	
고용노동부	기획조정실	(국번없이)1350	책임관	58
	정보화기획팀	(국번없이)1350	실무담당	
여성가족부	법무감사	02-2075-4560	책임관	25
	정보화담당관	02-2075-4601	실무담당	
국토교통부	정보화통계담당관	044-201-3293	공공데이터 확대 및 활성화 총괄	123
		044-201-3271	공공데이터 개방 및 공유	
해양수산부	정책기획관		책임관	35
	정보화담당관	044-200-5176	실무담당	
계				727

* 2014.03.20. 기준

〈표 3-2〉 3차 공공데이터 담당자 및 업무현황

관련부처	부서	연락처	담당업무	제공 Data
법제처	법령정보정책관	02-2100-2600	책임관	25
	법제정보과	02-2100-2600	실무담당	
국가보훈처	정부화담당관	044-202-5270	책임관	-
	전산주사	044-202-5284	실무담당	
식품의약품 안전처	기획조정관	043-719-1609	책임관	24
	정보화통계담당관	043-719-1615	실무담당	
계				49

* 2014.03.20. 기준

〈표 3-3〉 17청 공공데이터 담당자 및 업무현황

관련부처	부서	연락처	담당업무	제공 Data
국세청	기획조정관	02-397-1281	책임관	31
	통계기획담당관	02-397-1893	실무담당	
관세청	기획조정관	042-481-7640	책임관	89
	창조기획재정담당관	042-481-7660	실무담당	
조달청	기획조정관	040-4056-7056	책임관	20
	운영지원과	010-4056-7026	실무담당	
통계청	통계정보국	042-481-2310	책임관	4
	정보화기획과	042-481-2313	실무담당	
	정보화기획과	042-181-2314		
검찰청	기획조정부장	02-3480-2023	책임관	38
	운영지원과	02-3480-2455	실무담당	
병무청	병역자원국	042-481-2920	책임관	27
	정보기획과	042-481-2634	실무담당	
		0423-481-2646		
방위사업청	재정정보화기획관	02-2079-6700	공공데이터 책임관	12
	정보화기획담당관	02-2079-6782	공공데이터 제공 실무 담당자	
경찰청	경무담당관	02-3150-2623	공공데이터 책임관	10
	경무담당관실	02-3150-0751	공공데이터 제공 실무 담당자	
소방방재청	운영지원과	02-2100-5056	공공데이터 책임관 및 제공실무	26

관련부처	부서	연락처	담당업무	제공 Data
문화재청	기획조정관	042-481-4610	공공데이터 책임관	24
	정보화담당관	042-481-4750	공공데이터 공개 담당자	
	기록연구사	042-481-4762	공공데이터 공개 실무자	
농촌진흥청	기획조정관	031-229-2240	공공데이터 책임관	-
	지식정보화담당관	031-299-2371	공공데이터 제공 실무 담당자	
산림청	운영지원과	042-481-4010	공공데이터 책임관	16
		042-481-4011	공공데이터 제공 실무 담당자	
중소기업청	기획조정관	042-481-4301	공공데이터 책임관	39
	고객정보화담당관	042-481-4327	공공데이터 제공 실무 담당자	
		042-481-4591	공공데이터 제공 실무 담당자	
특허청	정보고객지원국	042-481-5817	공공데이터 책임관	1
	정보활용팀	042-481-5079	공공데이터 제공 실무 담당자	
기상청	기상자료과	02-2181-0233	공공데이터 담당관리	41
행정중심 복합도시건설청	지식정보팀	044-200-3262	공공데이터 제공 실무 담당자	15
해양경찰청	장비기술국	032-835-2270	공공데이터 책임관	21
	정보통신과	032-835-2081	공공데이터 실무 부서장	
		032-835-2482	공공데이터 실무 담당자	
계				414

* 2014.03.20. 기준

2. 국내 빅데이터 부서 현황

공공기관의 빅데이터 관련 조직을 개설한 기관은 <표 3-4>와 같다.

건강보험심사평가원 의료정보지원센터는 건강보험심사평가원이 보유하고 있는 다양한 빅데이터를 기반으로 한 IT 인프라 및 정보 활용을 적극 지원하고 보건의료 산업발전에 기여하는 것을 목적에 두고 있다. 향

후, 공개용 데이터베이스 및 사용 유저수를 확대하기 위한 인프라 확장 및 사용자 편의성 마련을 위한 포털 시스템을 구축할 예정이며, 빅데이터 인력 양성을 위한 교육 프로그램도 마련할 계획 중에 있다.

국민건강보험공단 빅데이터 운영실은 2013년 9월 9일 건강보험 빅데이터 운영센터를 개소하여 전국민 건강정보와 다양한 비정형 데이터를 융합한 빅데이터를 바탕으로 개인별 평생 맞춤형 건강서비스를 제공하고 관련 정보를 공개·개방을 하고 있다. 국민건강보험공단은 전국민 5천만 명의 출생에서부터 사망까지 자격 및 보험료 자료, 병의원 이용내역과 건강검진 결과, 가입자의 희귀난치성 질환 및 암등록정보 등 10년 동안 축적된 1조 3,034억 건의 빅데이터를 보유하고 있다. 건강보험공단의 빅데이터 운영실에서는 대용량의 정형·비정형 데이터를 처리할 수 있도록 지역별, 질환별, 연령군별, 사업장별 다량의 건강정보를 가공 구축할 수 있도록 하였으며, 개인별, 인구집단별, 다양한 맞춤형 건강관리 서비스를 제공하고 있다.

한국전자통신연구원에서는 빅데이터 소프트웨어 연구소를 개설하여 빅데이터 관련 분야 소프트웨어 컴퓨팅 산업 원천 기술 개발 사업을 활용한 업무를 담당하고 있다.

한국정보화진흥원은 빅데이터 전략센터를 개설하였으며, 빅데이터 기획부, 미래전략연구부, 지식자원 활용부로 총 3개 부서로 구성되어 있으며, 빅데이터 사업 기획 및 정책 수립, 빅데이터 컨설팅, 데이터 기반 아젠다 발굴 등의 업무를 담당하고 있다. 뿐만 아니라 빅데이터 분석활용센터를 개설하여 빅데이터를 분석 할 수 있는 인프라를 구축하고 다양한 분야에서 생산되는 데이터를 활용할 수 있도록 다양한 서비스를 제공하고 있다.

〈표 3-4〉 빅데이터 관련 부서 현황

관련부처	부서명	담당업무
건강보험심사평가원	의료정보 지원센터	- 보건의료정보 공개 포털시스템 구축 운영 - 보건의료 빅데이터 전문 인력 발굴 및 양성 확대
국민건강보험공단	빅데이터 운영실	- 빅데이터 체계 및 구축, 데이터 관리 - 개인별 평생 맞춤형 건강서비스 제공
한국전자통신연구원 (ETRI)	빅데이터 소프트웨어 연구소	- 빅데이터 관련 분야 소프트웨어 컴퓨팅 산업 원천 기술 개발 사업 잔행 등
한국정보화진흥원 (NIA)	빅데이터 전략센터	- 사업기획 및 정책수립 거버넌스 체계구축 운영 건설 팅 및 시범사업 발굴/추진 - 빅데이터 분석/유통기반 구축 - 국가DB관련정책 및 기본계획 수립지원, 법제도 연구 - 국가DB확충, 품질, 성과관리 및 활용촉진 국가DB관 련 방법론, 표준보급 등
	빅데이터 분석활용센터	- 빅데이터 활용 스마트 서비스 시범사업 - 빅데이터 산업 생태계 조성 등 여건조정 - 빅데이터 인력 양성 지원

〈표 3-5〉는 지자체 빅데이터 관련 부서 현황이다. 경기도청에서는 공공데이터 개방 추진 및 빅데이터 활성화 추진을 위해 정보화기획담당관실에서 운영하고 있다. 부산시청에서는 U-City 정보 담당관 부서를 개설하여 정부 3.0관련 공공데이터를 총괄 관리 담당하고 있으며, 빅데이터 과제 개발을 발굴하고 담당하고 있다. 뿐만 아니라 부산 해운대구청에서는 빅데이터 분석팀을 개설하여 빅데이터 관련 자료 수집 및 과제 개발을 하고 있으며, 타 기관과의 빅데이터 시스템 연계 추진을 위한 업무도 담당하고 있다. 광주광역시청에서도 빅데이터 관련 과제 발굴 및 현황조사 업무를 하고 있으며, 충청남도청 안전자치 행정국에서도 충청남도의 빅데이터 관련 과제발굴을 위한 업무를 담당하고 있다.

〈표 3-5〉 지자체 빅데이터 관련 부서 현황

지자체	부서명	팀원	담당업무
경기도	정보화기획관 정보서비스담당관	2	- 공공데이터 개방 추진 - 빅데이터 활용 및 활성화 추진
부산 해운대	빅데이터 분석팀	4	- 빅데이터적용업무 발굴 추진 및 수집 - 분석 정책방향 설정 및 빅데이터 시스템 구축 및 운영 - 타 기관 시스템 연계추진 - 통계전반의 통계연보 발간 및 행정지도 등
부산시	u-City정보담당관	1	- 정부 3.0부 총괄 (빅데이터 과제)
충청남도	안전자치행정국 정보화지원과	1	- 충남 공공데이터 관련 Open API개발 - 공공데이터 보유현황 파악관리 및 시스템 연계사업 - 빅데이터 관련 과제발굴 및 실과지원 업무
광주 광역시	기획조정실 정보화담당관	1	- 빅데이터 과제발굴 및 현황조사

제2절 빅데이터 관련 인력양성 현황

빅데이터 관련 시장의 지속적 성장과 더불어 대규모 데이터 속에서 새로운 가치를 창출하기 위하여 빅데이터 전문 인력 양성이 중요한 국가 경쟁력 중 하나로 부각되고 있다. 빅데이터 인력 양성을 위하여 기반 역량, 기술역량, 분석역량, 사업역량으로 구분하여 전문적인 교육을 통한 인력 양성을 기대하고 있다. 〈표 3-6〉은 국내 빅데이터 관련 인력을 양성하는 교육기관 현황이다. 국내에서는 총 13개 교육기관(가천대학교, 경북대학교, 국민대학교, 동국대학교, 서강대학교, 서울과학기술대학교, 서울대학교, 숙명여자대학교, 숭실대학교, 연세대학교, 울산과학기술대학원, 충북대학교, 한국과학기술원)에서 데이터 과학자 양성을 위한 빅데이터 커리큘럼을 개설하여 운영하고 있다.

〈표 3-6〉 국내 빅데이터 관련 학과 현황

대학교	학과
가천대학교	컴퓨터공학과
경북대학교	빅데이터 분석과
국민대학교	경영분석: 빅데이터통계전공, 빅데이터 경영 MBA
동국대학교	글로벌 소셜 창의 인재양성 교육과정
서강대학교	정보통신대학원 정보서비스트랙
서울과학기술대학교	데이터사이언스 석사과정
서울대학교	빅데이터 MBA
숙명여자대학교	빅데이터융합전공
숭실대학교	소프트웨어공학과
연세대학교	정보대학원 매경-연세대 빅데이터 학과
울산과학기술대학원	비즈니스 분석 과정
충북대학교	비즈니스데이터 융합과정
한국과학기술원	지식서비스공학 전공

자료: 한국정보화진흥원(2014), 빅데이터 커리큘럼 참조 모델 Ver 1.0.

제3절 공개·개방된 빅데이터 현황 및 활용 방안

‘공급자 위주’에서 ‘국민 중심’ 정보 공개로 패러다임이 전환됨에 따라 공공정보가 민간의 창의성 및 혁신적인 아이디어와 결합하여 새로운 비즈니스를 창출할 수 있는 생태계 조성을 위해 많은 나라에서 공공정보의 공개를 추진하고 있다. 우리나라는 2011년 7월부터 정부와 공공기관이 보유한 데이터를 대대적으로 개방하여, 기관 간 공유는 물론 국민과 기업이 상업적으로 자유롭게 활용할 수 있도록 공공데이터 개방을 추진하고

있다. 공공데이터는 각 기관이 전자적으로 생성 또는 취득하여 관리하고 있는 모든 데이터베이스(DB) 또는 전자화된 파일로 범정부 차원에서 영리·비영리적 목적에 관계없이 개발·활용을 촉진하고 있다. 공공데이터 개방과 관련하여 2013년 10월 ‘공공데이터의 제공 및 이용 활성화에 관한 법률’이 제정·시행됨에 따라 각 부처별로 분야별 공공데이터의 공개와 효율적 활용방안을 모색하고 있다. 월드와이드웹(World wide web Foundation)⁶⁾에서 발표한 전 세계 77개국 대상 공개데이터(open data) 현황보고서에 따르면 1위 영국, 2위 미국, 3위 스웨덴, 4위 뉴질랜드, 5위 노르웨이, 덴마크, 7위 호주, 8위 캐나다, 9위 독일, 10위 프랑스, 네덜란드, 12위 한국 순으로 나타나고 있다. 2013년 10월 현재 공개되고 있는 공공데이터는 영국 15,768건, 미국 90,565건, 한국 7,822건으로 보고하고 있다.

[그림 3-3] 국외 공공데이터 순위: 2013년



자료: <http://www.opendataresearch.org/dl/odb2013/Open-Data-Barometer-2013-Global-Report.pdf>

본 고에서는 주요국과 우리나라의 공공 데이터 공개 현황에 대해 살펴 보고 효율적 활용방안을 제시코자 한다.

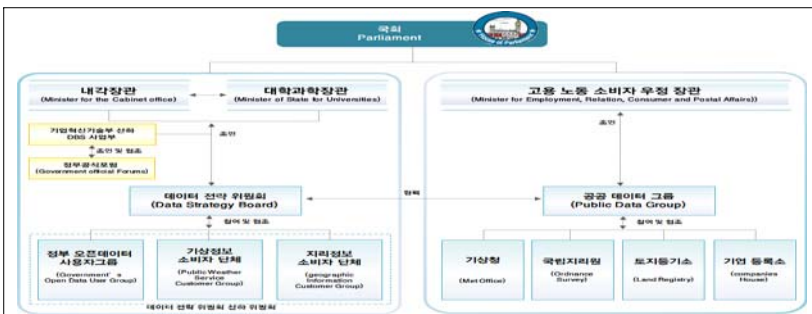
6) 월드와이드웹(World wide web Foundation)은 월드와이드웹을 고안한 팀 버너스 리가 인터넷의 가치를 증진시키기 위하여 설립한 비영리 재단

1. 해외 공공데이터 현황

가. 영국

영국은 내각사무처에서(Cabinet office)는 공공데이터 공개를 통해 데이터에 대한 용이한 접근과 데이터 활용 기회 증대하기 위하여 「오픈 데이터 연구소(Open Data Institute)」를 설립하여, 데이터 접근 및 활용을 통하여 비즈니스 모델 개발 및 데이터 구축을 지원하고 있으며, 빅데이터 활용의 기반이 되는 공공 부문의 정보공유 및 활용에 따른 가치창출을 위한 데이터 공개 및 공유 중심의 정책을 추진하고 있다. [그림 3-4]와 같이 영국 정부는 내각사무처에서 의사결정을 담당하고 있으며, 기업 혁신기술부에서 추진기구 및 참여부처를 전담하여 담당하고 있다. 기업 혁신기술부는 공공정보 공개 및 데이터 가치 창출을 위하여 데이터 전략위원회(Data Strategy Board)를 설립하여, 데이터 공개 여부 판단, 데이터 활용을 통한 서비스 발굴 가능성 여부 등을 검토하여 공개 및 활용 등에 대한 제언을 담당하고 있다(윤미영, 2013).

[그림 3-4] 영국의 빅데이터 전략위원회 조직 구성 체계도



자료: 윤미영(2013), 주요국의 빅데이터 추진전략 분석 및 시사점, 한국정보화진흥원 과학기술정책 23(3)

영국의 공공데이터 포털(data.gov.uk)에서는 정부 지출, mapping, 사회, 건강, 정부, 환경, 교육, 비즈니스 및 경제, 범죄 및 사법, 도시와 관련된 공공정보를 포털을 통하여 서비스 하고 있다. 전체 15,769건의 데이터 중 보건 분야는 1,686건(10.7%), 사회분야는 1,896건(12.0%)을 서비스 하고 있다(표 3-7 참조).

〈표 3-7〉 영국 Open data DB 현황(2014년 2월 21일 기준)

분류	DB 구축 현황
정부지출	2,462 (15.6)
매핑	2,049 (13.0)
사회	1,896 (12.0)
보건	1,686 (10.7)
정부	1,607 (10.2)
환경	3,282 (20.8)
교육	858 (5.4)
비즈니스 및 경제	561 (3.6)
범죄 및 사법	493 (3.1)
도시	875 (5.5)
계	15,769 (100.0)

자료: 영국 공공데이터 포털(data.gov.uk) 참조(2014년 2월 21일 기준)

제공 데이터 중 출판된 데이터 셋(published-data set)은 1,522건, 출판되지 않은 데이터 셋(unpublished-data set)은 164건을 서비스 하고 있다(표 3-8 참조).

〈표 3-8〉 영국 Open data 분류체계(2014년 2월 21일 기준)

분류	DB 구축 현황
Published 데이터셋	1522 (90.3)
unpublished 데이터셋	164 (9.7)
계	1,686 (100.0)

자료: 영국 공공데이터 포털(data.gov.uk) 참조(2014년 2월 21일 기준)

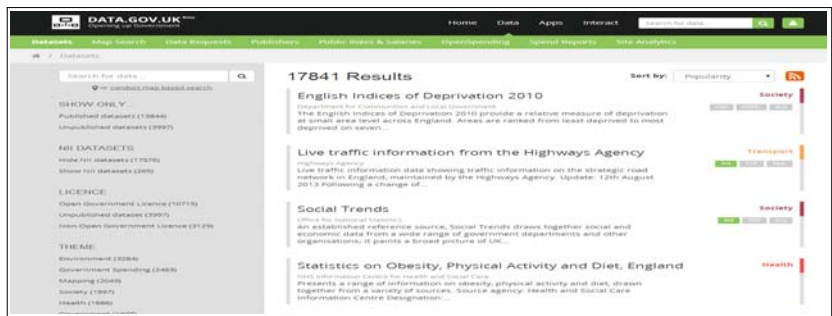
영국 오픈데이터 플랫폼(data.gob.uk) 포털의 보건분야 분류체계 화면과, 메인페이지 화면은 [그림 3-5]와 [그림 3-6]과 같다.

[그림 3-5] 영국 Open data 보건분야 분류체계



자료: 영국 공공데이터 포털(data.gov.uk) 참조(2014년 2월 21일 기준)

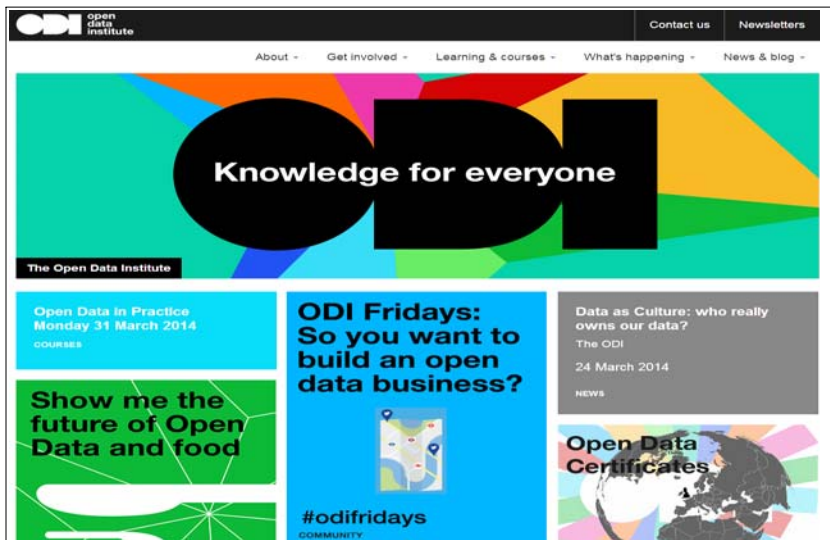
[그림 3-6] 영국 Open data 포털 화면(2014년 2월 21일 기준)



자료: 영국 공공데이터 포털(data.gov.uk) 참조(2014년 2월 21일 기준)

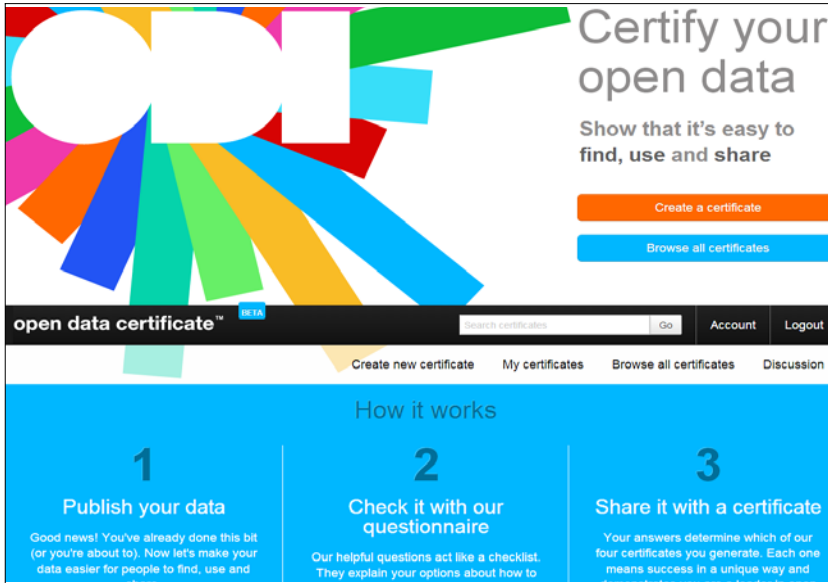
세계 최초 오픈 데이터 전문기관인 영국은 오픈 데이터 연구소(Open Data Institute)의 포털을 통해서 데이터를 공유(등록)하여 데이터 인증 후 서비스를 이용자에게 제공하고 있다. 데이터 공유자(제공자)는 데이터 등록 시 데이터에 대한 메타 정보를 입력, 데이터 주기성 체크(단기, 시계열 등 체크), 개인정보보호에 대한 권리 및 법률에 관한 동의, 서비스 이용자와의 피드백 소통 방법 등에 대한 정보를 입력을 통해 오픈데이터 라이선스를 발급받을 수 있다. 오픈데이터 인증을 받음으로써 데이터 이용자와 공유자(제공자)는 상호소통(피드백)을 주고받을 수 있다(그림 3-7, 3-8 참조).

[그림 3-7] 영국 Open data Institute 포털 화면(2014년 3월 27일 기준)



자료: 영국 오픈 데이터 전문 기관(theodi.org)

[그림 3-8] 영국 ODI 데이터 인증 페이지



자료: 영국 오픈 데이터 전문 기관(theodi.org)

〈표 3-9〉과 같이 교통부에서는 철도, 버스, 도로, 항공, 운전데이터를 제공하고 있으며, 국세청에서는 국세청 국가 공식 통계, 조달, 세금, 수입자 데이터를 제공 하고 있다. 보건부에서는 의사 업무 관련 일반 데이터, 암 치료 데이터, 국민 보건 의료 서비스 데이터를 공개하고 있다.

〈표 3-9〉 영국 정부 부처별 데이터 공개내용

부처	데이터구분	공개내용
교통부	철도 데이터	철도운영, 성능, 철도망 지리정보, 요금 등
	버스 데이터	버스 시간표, 버스 출도착 알림 등
	도로 데이터	(고속)도로 정보, 도로보수 정보, 도로상태 등
	항공 데이터	승객 경험데이터 수집을 통한 항공프로그램
	운전 데이터	운전시험 불합격 데이터, 운전면허 정보, 교통과 운전자 정보에 관한 요청사항
국세청	국세청 국가공식 통계	100가지 이상의 국가공식통계 데이터 (세금, 혜택, 거래, 데이터 등)
	조달 데이터	조달 거래 관련 데이터
	세금 데이터	납부해야 할 세금과 건강보험 계산 데이터
	수입자 데이터	EU로부터 영국으로 상품을 수입하는 13만 명 이상의 사업자 이름과 주소가 제공되는 수입자 세부 정보
문화 매체부	올림픽 정보	자원봉사자, 올림픽으로 인한 영향력 통계, 훈련캠프 통계 등
	복권	복권 당첨자 세부 데이터
	브로드 밴드	브로드밴드 펀드에 참가하는 프로젝트의 총 부가가치 및 고용데이터
	지속가능사업 모델	문화유산, 스포츠 등 지속가능한 사업 재정데이터
	정부예술품	예술작품 검색가능 이미지 데이터
기업 혁신 기술부	고등교육 데이터	학생들의 고등교육 진학 시 고려사항
	기업등록소 데이터	기업 명, 기업 수, 등 주소, 분류 상태 및 원문 코드(SIC)무로 다운로드
	국립지리원 데이터	국립지리원의 오픈데이터, 국립자연탐방로 데이터 세트
노동 연금부	워크 프로그램	고용유지 지원금, 일자리 창출 결과 등에 대한 통계
	건강 정보기록	근로자 건강정보
	일반 신용거래	근로자 신용관련 데이터
보건부	의사업무 관련 일반 데이터	의료진의 정보를 공개, 환자들이 병원간 치료 및 생존률 비교, 치료 관련 정보(치료시간, 대기시간, 상담자료 등)
	암 치료 데이터	1차 암 치료와 관련된 과정정보 및 결과정보
	국민 보건서비스	연도 별, 트렌드 별로 성과비교가 가능한 국민보건 서비스 데이터
외무부	영사관, 대사관 데이터	영국 국민에 대한 통계 (외국 여행, 외국에서 사망, 외국에서의 체포여부 등)

자료: 윤미영, 주요국의 빅데이터 추진전략 분석 및 시사점, 한국정보화진흥원 과학기술정책 23(3)

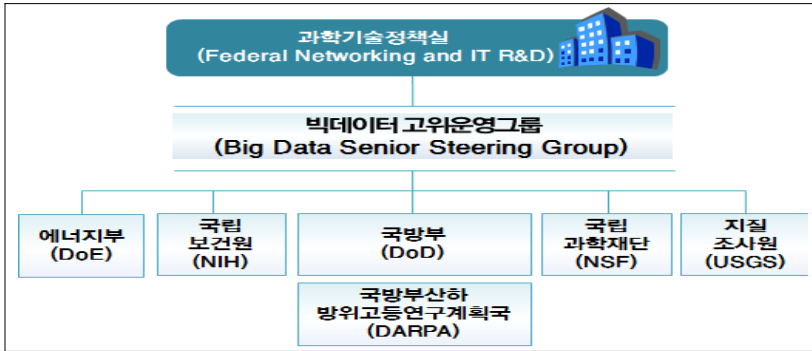
나. 미국

미국은 범정부 차원의 빅데이터 거버넌스 체계를 구축하여 현재 6개 연방정부부처 및 기관이 빅데이터 고위 운영그룹과 함께 프로젝트를 진행하고 있으며 지속적으로 확대를 할 예정에 있다(그림 3-9 참조).

[그림 3-9]은 미국 정부의 빅데이터 거버넌스 체계를 구축하기 위한 추진체계이다. 빅데이터 의사결정기관인 과학기술정책실(Federal Networking and IT R&D)은 빅데이터 연구 개발 조정, 이니셔티브 목표 확인 등을 위해 빅데이터 협의체인 빅데이터 고위 운영 그룹(Big Data Senior Steering Group)을 구성하였다. 과학기술정책실은 빅데이터 고위운영그룹을 중심으로 빅데이터 R&D 이니셔티브 대상 프로젝트를 선정 및 개발하고 투자규모를 결정하는 역할을 담당하고 있다. 데이터 수집에서 저장, 보존, 관리, 분석, 공유와 관련된 핵심 기술의 최신성을 유지하고, 빅데이터 관련 부처 간 연계 프로젝트로 얻을 수 있는 편익을 분석하고 실현 가능한 협업 프로젝트를 개발 하고 제안하고 있다.

빅데이터 추진기구인 빅데이터 고위 운영 그룹은 정부기관 및 소속기관으로 구성되어 있으며, 유전자 연구 및 의료, 교육, 지구과학 등 빅데이터 활용 효과가 뛰어난 분야의 기관들이 우선적으로 참여하여 주도적으로 추진을 하고 있으며, 참여 부처는 에너지부, 국립보건원, 국방부, 국방부 산하 방위고등연구계획국, 국립과학재단, 지질조사원이 참여하고 있다.

[그림 3-9] 미국의 빅데이터 추진전략 및 체계



자료: 윤미영(2013), 주요국의 빅데이터 추진전략 분석 및 시사점, 한국정보화진흥원 과학기술정책 23(3)

미국 공공데이터 포털(data.gov)에서는 농업, 소비자, 교육, 에너지, 건강, 공공안전 등의 분야별 공개데이터를 서비스 하고 있다. 포털을 통해 서비스 하고 있는 데이터 셋은 총 90,565건을 공개하고 있다(표 3-10 참조). 특히, 보건관련 분야의 오픈데이터는 별도의 사이트에서 서비스하고 있다. 보건분야의 경우 ‘미국 공공데이터(data.gov)’에서는 미국 사회 보건부(U.S. Department of Health and Human Service)에서 서비스하는 1,127개 데이터 셋을 구성하여 서비스 하고 있으며(표 3-11 참조), ‘보건관련 공공데이터 제공 사이트(healthdata.gov)’에서 1,450개 데이터 셋을 별도로 주제별로 분류하여 제공하고 있다(그림 3-10 참조).

〈표 3-10〉 미국 공공데이터 제공 사이트 Open DB 현황(2014년2월 24일 기준)

분류	DB 구축 현황
사회보건부	1,127 (100.0)
기관 단위 부서	1,124 (99.7)
Medicare 및 Medicaid 센터	1 (0.1)
미국국립보건원	1 (0.1)
약물남용 및 정신건강 서비스 관리	1 (0.1)

〈표 3-11〉 미국 보건관련 공공데이터 제공 사이트 DB 현황(2014년2월 24일 기준)

분류	DB 구축 현황
Medicare	510 (35.2)
보건(건강)	174 (12.0)
NNDSS	84 (5.8)
인구통계	74 (5.1)
공중보건	70 (4.8)
병원비교	55 (3.8)
관리	53 (3.7)
사회/보건의료	52 (3.6)
기타	45 (3.1)
보건 및 휴면 서비스	39 (2.7)
안전	36 (2.5)
의료인	33 (2.3)
어린이 건강	27 (1.9)
의료비용	25 (1.7)
Medicaid	25 (1.7)
의생명과학연구	20 (1.4)
역학	17 (1.2)
질관리	14 (1.0)
치료	13 (0.9)
투석설비비교	11 (0.8)
방문건강비교	8 (0.6)
건강통계	7 (0.5)
방문간호비교	6 (0.4)
보급	4 (0.3)
보건 및 휴면 서비스	4 (0.3)
흡연 및 니코틴의존	4 (0.3)
분류체계	3 (0.2)
메디케어 연rak관련	3 (0.2)
의사 비교	3 (0.2)
임신 및 백신	1 (0.1)
서비스	1 (0.1)
공급자	1 (0.1)
web metrics	1 (0.1)
미분류	27 (1.9)
계	1,450 (100.0)

자료: 미국 보건관련 공공데이터 제공 사이트(healthdata.gov)+

[그림 3-10] 미국 Open data 포털 화면(2014년2월 21일 기준)

미국 공공데이터 (data.gov)	보건관련 공공데이터 제공 사이트(healthdata.gov)
	

다. 일본

일본 정부의 빅데이터 추진전략은 2013년 6월 고도 정보통신 네트워크 사회추진전략본부(약칭 IT종합전략본부(본부장: 수상))가 발표한 ‘세계 최첨단 IT 국가창조선언’(이하 ‘창조적 선언’)의 공정표에 잘 나타나 있다. IT종합전략본부는 IT 정보자원을 활용하여 미래를 창조하는 국가 비전으로서 ‘창조적 선언’(2013년 6월 14일 내각 결정)을 수립했다. 이 선언에는 향후 2020년까지 세계 최고 수준의 IT 활용 사회 실현을 목표로 ① 혁신적인 신산업·신서비스의 창출과 전체 산업의 성장을 촉진하는 사회의 실현, ② 국민이 건강하고, 안심하고 쾌적하게 생활하는 세계에서 가장 안전하고 재해에 강한 사회의 실현, ③ 공공서비스를 누구나 언제 어디서나 원스톱으로 받을 수 있는 사회의 실현의 3개 항목에 대한 지향해야 할 사회 모습을 제시하였다. 이 공정표는 어느 부처가 언제까지 구체적으로 무엇을 실시하는지를 밝히고, 각 부처 간 연계가 필요한 시책

에 대해서는 개별 역할분담과 달성해야 할 사항을 명확히 하여 꾸준히 구체적인 성과로 연결시키는 것을 목적으로 책정한 것이다. 공정표에는 ‘창조적 선언’에 나타난 전략과 목표에 대해 단기·중기·장기로 나누어 각 부처가 실시할 시책이 명시되어 있다(표 3-12 참조). 오픈데이터·빅데이터 활용의 추진 시책은 상기의 첫 번째 목표인 ‘혁신적인 신산업·신서비스의 창출과 전체 산업의 성장을 촉진하는 사회의 실현’이 제시되었으며, 다음은 단기·중기·장기별로 나누어 추진할 계획이다(송태민 외, 2013).

〈표 3-12〉 일본의 빅데이터 장기 추진계획 (2019~2021년)

계획	세부계획
제도정비	IT중립전략본부 아래 설치된 새로운 검토 조직에서 정리된 제도 재검토의 방침에 따라 국제협력도 배려하면서 각 시책을 실시한다(내각관방, 관계부처).
활용촉진	각 분야(지역 활성화, 대중교통, 방재, 의료·건강, 에너지 등)의 실증 프로젝트 등의 성과를 실행, 새로운 서비스의 창출을 촉진한다(내각관방, 내각부, 총무성, 후생노동성, 농림수산성, 경제산업성, 국토교통성, 문부과학성).
인재육성	새로운 서비스, 새로운 사업전략 입안과 신기술 창출에 빅데이터를 활용할 수 있는 인재(데이터 사이언티스트 등) 육성에 착수한다(문부과학성).
기술개발	빅데이터의 활용을 촉진하기 위해 데이터 및 네트워크의 안전성·신뢰성 향상 및 상호 운용성 확보, 대규모 데이터의 축적·처리기술의 고도화 등 공통기술의 조기 확립을 도모함과 동시에 새로운 비즈니스, 새로운 서비스의 창출로 이어지는 새로운 데이터 활용기술의 연구개발 및 그 활용을 추진한다. 구체적으로는 데이터 활용기술(수집·전송, 처리, 활용·분석 등)에 대한 각 부처의 역할을 명확히 하고 나아가 각 부처가 연계하여 다른 목적으로 수집된 다양한 데이터에서 유용한 정보·지식을 실시간으로 추출할 수 있는 기술을 실용화하고, 정보를 유통·순환시켜 분야의 경계를 넘어 정보가 활용됨으로써 신사업, 신서비스의 창출을 촉진한다(총무성, 문부과학성, 경제산업성)

자료: 高度情報通信ネットワーク社会推進戦略本部, 世界最先端IT国家創造宣言工程表, 2013. 6. 14.

라. 뉴질랜드⁷⁾

뉴질랜드(data.govt.nz)에서는 농업, 산림 및 어업, 예술 문화유산, 건축구조 및 주택, 무역 및 산업, 교육, 근로자, 에너지, 환경 보전, 국가재

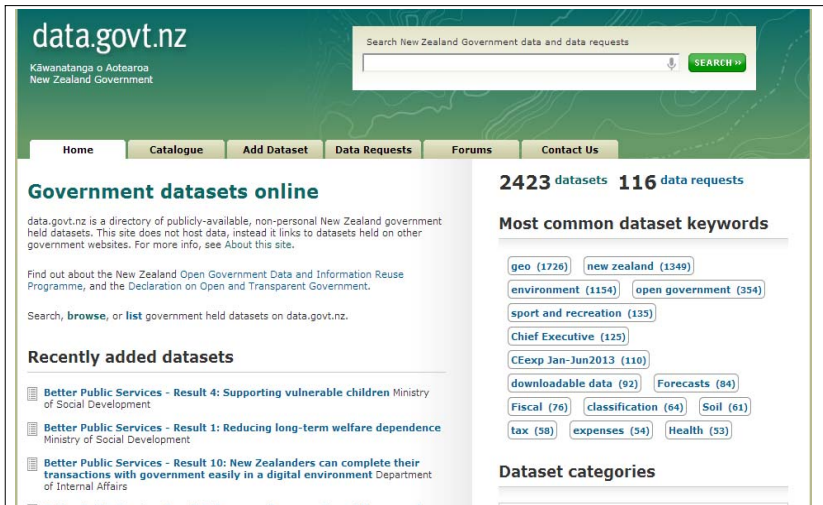
7) 뉴질랜드 공공데이터 포털(data.govt.nz), 2014.02.25. 검색

정·세금 및 경제, 보건, 사회 공공기반 시설, 인구 및 사회, 과학 및 연구 등으로 분류되어 있다. 뉴질랜드 오픈데이터 포털은 현재 2,423개의 데이터 셋을 구성하여 서비스 하고 있다.

〈표 3-13〉 뉴질랜드 Open data DB 현황

분류	DB 구축 현황
농업, 산림 및 어업	15 (0.62)
예술, 문화 유산	8 (0.33)
건축구조 및 주택공급	13 (0.54)
무역 및 산업	37 (1.53)
교육	25 (1.03)
근로자	17 (0.70)
에너지	17 (0.70)
환경보전	52 (2.15)
국가재정 및 세금	186 (7.68)
보건	46 (1.90)
사회 공공기반 시설	4 (0.17)
형평성	9 (0.37)
land	1737 (71.69)
지방 및 지역정부	9 (0.37)
Maori and pasifika	7 (0.29)
이민	2 (0.08)
인구 및 사회	38 (1.57)
과학 및 연구	13 (0.54)
국가 공무 성과	144 (5.94)
관광	13 (0.54)
운송	24 (0.99)
장관 내각 및 포트폴리오	2 (0.08)
미분류	5 (0.21)
계	2,423 (100.0)

[그림 3-11] 뉴질랜드 Open data 포털 화면(2014년 2월 25일 기준)



마. 캐나다⁸⁾

캐나다 정부는 2011년 3월 18일 세 가지 주요 공개정책(공개정보, 공개 데이터, 열린 대화)에 따라 공공데이터에 대한 이니셔티브를 발표했다. 지방정부별로 공공정보 제공 사이트를 별도로 운영하고 있으며, 정보 공개에 대한 접근성을 강화시키고 있다. 2011년 공개 데이터에 대한 이니셔티브에 따라 공공데이터 포털이 구축되어 운영되고 있으며, 19개 분류 체계에 의하여 244,617건의 데이터 셋을 구축하여 제공하고 있다.

공공데이터(data.gc.ca/eng) 포털에서는 기관별, 데이터 타입별, 제공하는 데이터의 파일 형태별(엑셀, PDF, XML 등), 키워드 별(Tag), 주제별, 라이선스 여부 등으로 분류하여 데이터를 서비스 하고 있다. <표

8) 캐나다 공공데이터:

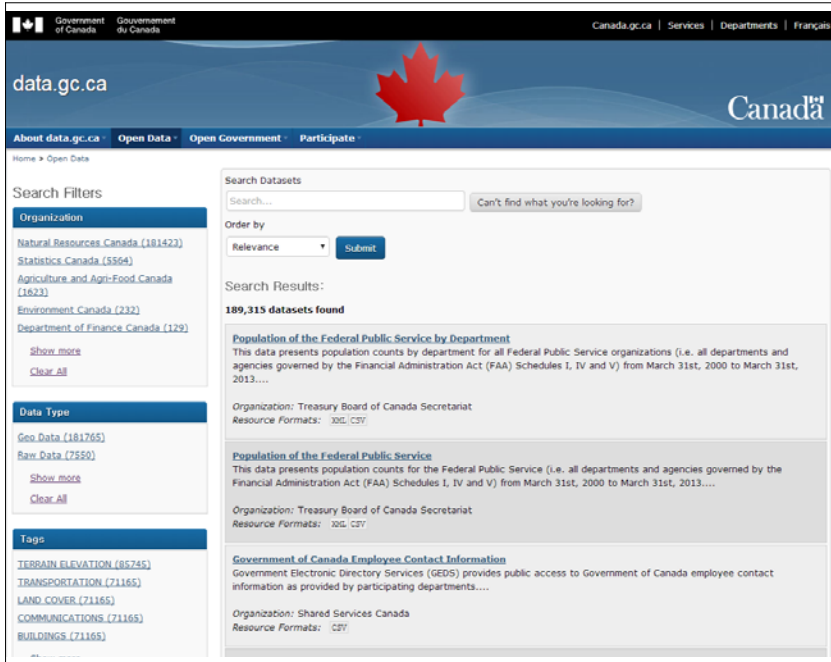
[http://data.gc.ca/eng/canadas-action-plan-open-government\(2014.03.27](http://data.gc.ca/eng/canadas-action-plan-open-government(2014.03.27) 검색)

3-14)는 캐나다 오픈데이터 포털에서 제공하는 주제별 분류에 따른 DB 현황이다. 기술, 자연 및 환경, 과학기술, 사회 및 문화, 건강 및 안전, 법 등의 분류하여 DB를 제공 하고 있으며, 주제별 분류체계에 따라 제공하고 있는 데이터 현황은 244,617건을 제공하고 있다.

〈표 3-14〉 캐나다 Open data 주제별 분류에 의한 DB 현황

분류	DB 구축 현황
기술	175,626 (71.80)
자연 및 환경	31,089 (12.71)
과학기술	27,177 (11.11)
경제 및 산업	3,895 (1.59)
사람	1,539 (0.63)
농업	1,311 (0.54)
사회 및 문화	1,198 (0.49)
건강 및 안전	630 (0.26)
정부 및 정치	626 (0.26)
교통수단	554 (0.23)
노동	395 (0.16)
교육 및 훈련	222 (0.09)
정보통신	198 (0.08)
법	101 (0.04)
미술, 음악 문학	23 (0.01)
역사 및 고고학	12 (0.00)
군	8 (0.00)
프로세스	7 (0.00)
언어 및 언어학	6 (0.00)
계	244,617 (100.0)

[그림 3-12] 캐나다 Open data 포털 화면(2014년 3월 27 기준)



바. 싱가포르⁹⁾

싱가포르 공공데이터(data.gov.sg) 포털에서는 제공 기관별, 주제별로 분류하여 데이터를 서비스 하고 있으며, 제공하는 파일 형태로는 XLS, CSV 형태로 지원하고 있다. 싱가포르 오픈데이터 포털에서는 전체 총 8,836건의 데이터 셋을 구성하여 제공하고 있다.

〈표 3-15〉와 같이 농업, 가축 생산 및 어업, 경영통계, 교육, 환경, 보건 등의 38가지 주제별로 분류하여 DB를 서비스 하고 있으며, 〈표

9) 싱가포르 공공데이터(data.gov.sg), 2014.02.25 검색

3-16>는 68개 기관에서 제공하고 있는 DB 제공 현황이다. 그 중 건강증진국 및 보건과학청에서 33건, 14건의 데이터 셋을 구성하여 서비스 하고 있다.

〈표 3-15〉 싱가포르 Open data 주제별 분류에 의한 DB 현황

분류	DB 구축 현황
농업, 가축 생산 및 어업	46 (0.5)
투자 및 국제 무역, 지불 균형	1750 (19.8)
금융, 보험 및 재정 통계	715 (8.1)
경영 통계	652 (7.4)
건축업	93 (1.1)
문화 및 휴양	30 (0.3)
국민 경제 계정	560 (6.3)
경제	2 (0.0)
교육	396 (4.5)
에너지	67 (0.8)
환경	139 (1.6)
농업	7 (0.1)
보건	324 (3.7)
가구 소득 및 가구 비용	20 (0.2)
주택 및 도시계획	249 (2.8)
Imagery Basemap Earth Cover	42 (0.5)
MDG 지표	1 (0.0)
정보사회	284 (3.2)
Justice, Crime and Crisis Management	69 (0.8)
노동	117 (1.3)
노동비용	20 (0.2)
(시설 등의)위치	21 (0.2)
거시경제 통계	61 (0.7)
제조업	974 (11.0)
Planning Cadastre	2 (0.0)
인구 및 가구 특성	534 (6.0)
물가	1047 (11.8)
공공재정	93 (1.1)
과학 기술 및 혁신	42 (0.5)
사회	21 (0.2)
공동체 및 시민사회	53 (0.6)

분류	DB 구축 현황
구조	8 (0.1)
관광 여행	9 (0.1)
수송	24 (0.3)
운송 및 입고	122 (1.4)
유틸리티 통신	2 (0.0)
기타	11 (0.1)
기타서비스	229 (2.7)
계	8,836 (100.0)

자료: 싱가포르 open data(data.gov.sg)

〈표 3-16〉 싱가포르 Open data 제공 기관별 분류에 의한 DB 현황

기관명	DB 제공 현황
Health Promotion Board	33
Health Sciences Authority	14
계	47

〔그림 3-13〕 싱가포르 Open data 포털 화면(2014년 2월 25일 기준)

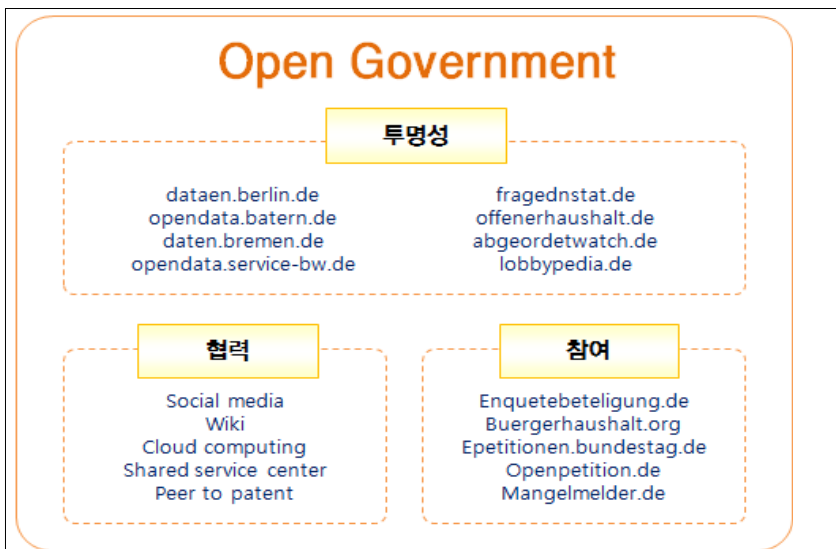
The screenshot displays the Singapore Government Open Data Catalogue interface. At the top, the header includes the 'data.gov.sg' logo, navigation links (Home, About Us, Data Sharing Principles, Data Catalogue, Application Showcase, For Developers, News & Events, FAQ, Useful Links), and a search bar. The main content area features a search bar with the text 'Search Data Catalogue' and a 'Search' button. Below the search bar, there are two filter sections: 'Filter by Category' and 'Filter by Agency'. The 'Filter by Category' section lists various categories such as Agriculture, Animal Production and Fisheries, Balance of Payments, Investments and International Trade, Banking, Insurance and Financial Statistics, Business Statistics, Construction, Culture and Recreation, Economic Accounts, National Accounts and Input-Output Tables, Economy, and Education. The 'Filter by Agency' section lists various agencies such as Accountant-General's Department, Accounting and Corporate Regulatory Authority, Agency For Science, Technology & Research, Agri-Food & Veterinary Authority of Singapore, Attorney-General's Chambers, Auditor-General's Office, Board of Architects, Building & Construction Authority, and Casino Regulatory Authority. The search results section shows '8836 record(s) found.' and a table of results. The first result is '3 Days Outlook (3 Days Weather Forecast)' with a download link, category 'Environment', agency 'National Environment Agency', and data type 'Textual'. The update frequency is listed as 'Daily'.

사. 독일

독일은 공공데이터 공개를 위하여 정보기술계획위원회(IT Planning Council)에서 국가 수준의 데이터와 지방 연방 수준의 데이터를 통합 게재하는 포털을 제안하여 2013년 1월부터 1년간 시범운영을 하였다.

독일은 오픈데이터 공개를 위하여 정부 스스로가 ‘정보 공개’라는 정치적 의미를 공개적으로 밝히는 것이 필요하며, 데이터 공개를 위하여 ‘열린 정부’ 전략을 펼 수 조건으로 내걸었다. 열린 정부 전략으로는 투명성, 협력, 참여를 내걸어 행정기관, 정치기관, 시민사회, 기업 등의 오픈데이터 네트워크를 형성해 나갈 수 있도록 환경을 조성하였다(그림 3-14, 3-15 참조).

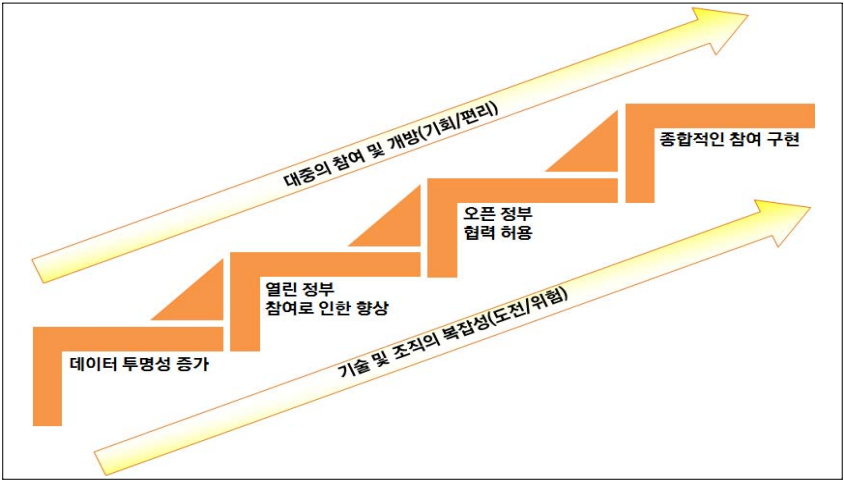
[그림 3-14] 독일 공공데이터 공개를 위한 ‘열린정부’ 전략



자료: 독일 공공데이터 포털 운영을 위한 프레임워크(2012.11.14.)

<http://www.bpb.de/gesellschaft/medien/opendata/149423/geruest-fuer-open-data-portal-steht>

[그림 3-15] 독일 공공데이터 공개를 위한 ‘열린정부’ 구현 모델



자료: Bundesministerium des Innern, Open Government Data Deutschland, 2012

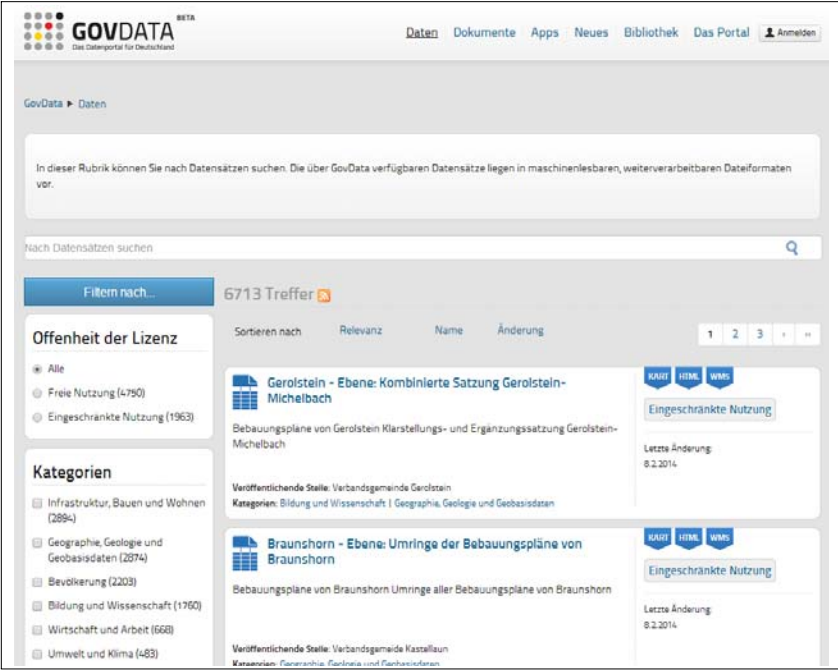
독일 공공데이터(govdata.de) 포털에서는 주제별, 주요 키워드 별, 자료 개방 분류별(무료/사용 제한), 제공 데이터 형식(HTML/WMS/지도/CSV/XLS 등)으로 분류하여 서비스를 제공하고 있다. <표 3-17>은 독일의 주제별 분류에 의한 DB 현황이다. 전체 데이터 12,298건 데이터 셋 중 건강 관련 DB는 104건 제공하고 있다.

<표 3-17> 독일 Open data 주제별 분류에 의한 DB 현황

분류	DB 구축 현황
인프라, 건설 및 주택	2,894 (23.5)
지리	2,874 (23.4)
인구	2,203 (17.9)
교육 및 과학	1,760 (14.3)
경제 및 고용	668 (5.4)
환경 및 기후	483 (3.9)
사회	481 (3.9)
소비자 보호	433 (3.5)

분류	DB 구축 현황
공공행정 및 예산/세금	218 (1.8)
건강	104 (0.8)
문화, 휴양, 스포츠, 관광	70 (0.6)
수송 및 교통	60 (0.5)
정치 및 선거	42 (0.3)
법	8 (0.1)
계	12,298 (100.0)

[그림 3-16] 독일 Open data 포털 화면(2014년 2월 26일 기준)



독일 Open data에서 제공하는 DB를 이용하여 오픈 플랫폼을 제공하는 현황은 <표 3-18>과 같다.

〈표 3-18〉 독일 Open 소스 기반의 어플리케이션 현황

분류	DB 구축 현황
환경 및 기후	8
소셜	4
교통수단	2
경제 및 노동	1
정치 및 선거	1
지리	1
교육 및 과학	1
인구	1
계	19

아. 호주

호주는 정부 2.0을 통한 정보 개방 및 투명한 문화로의 전환 및 시민과의 협력 강화를 목표로 공공데이터(data.gov.au)웹사이트를 개설하여 운영하고 있으며(한국정보화진흥원, 2010), 128개 정부기관에서 데이터 셋 정보를 제공받아 공개하고 있으며, 25개 그룹으로 분류하여 정보를 서비스 하고 있다. 또한 호주 공공데이터 포털은 지속적으로 정보공개 요구를 통해 데이터 접수를 받고 있다.

호주 공공데이터 포털(data.gov.au)에서는 사업지원 및 규정, 사회기반시설, 커뮤니케이션, 군사안보, 커뮤니티 서비스, 문화 업무, 환경, 교육 및 훈련, 고용, 거버넌스, 보건의료, 재무관리, 토착업무, 국제규정, 이민, 해상서비스, 천연자료, 사법 행정, 1차 산업, 보안, 과학, 스포츠, 관광, 무역, 운송 등으로 분류하여 데이터 셋을 제공하고 있다(표 3-19 참조).

〈표 3-19〉 호주 Open data 주제별 분류에 의한 DB 현황

분류	DB 구축 현황
사업지원 및 규정	69 (11.7)
사회기반시설	-
커뮤니케이션	15 (2.5)
군사안보	-
커뮤니티 서비스	133 (22.5)
문화 업무	19 (3.2)
환경	66 (11.2)
교육 및 훈련	13 (2.2)
고용	10 (1.7)
거버넌스	-
보건의료	37 (6.3)
재무관리	53 (9.0)
토착업무	3 (0.5)
국제 규정	-
이민	-
해상서비스	-
천연자원	-
사법행정	-
1차산업	-
보안	-
과학	92 (15.6)
스포츠 및 레크레이션	50 (8.5)
관광	6 (1.0)
무역	-
운송	24 (4.1)
계	590 (100.0)

[그림 3-17] 호주 Open data 포털 화면(2014년 2월 26일 기준)

The screenshot shows the data.gov.au website. At the top, there is a header with the Australian Government logo, the data.gov.au logo, and navigation links: Datasets, Organisations, About, Site Statistics, and a search bar. Below the header, there is a 'Search Your Data' section with a search bar containing 'eg. Gold Prices' and a magnifying glass icon. Below the search bar, there are 'Popular Tags' including geodesy, Geodetic Data, and Geodesy.

The main content area is divided into two columns. The left column contains text about data.gov.au, including a paragraph about finding, accessing, and reusing public datasets, and a paragraph about requesting data sets or new functionality through the data.gov.au roadmap. It also includes a list of links for Groups, Tags, and Search, and a paragraph about working across governments to publish more data. The right column contains several sections: 'data.gov.au statistics' showing 3.4k datasets, 128 organisations, and 25 groups; 'Latest data.gov.au News' with a link to 'Subscribe to the blog'; 'APS Better Practice Guide for Big Data' with a post date of 'Posted on Wed Apr 16 2014'; 'March 2014 data.gov.au report' with a post date of 'Posted on Mon Apr 14 2014'; 'Seeking proposals for joint private-public projects using big data' with a post date of 'Posted on Thu Mar 13 2014'; and 'February 2014 data.gov.au report' with a post date of 'Posted on Wed Mar 12 2014'.

2. 국내 공공데이터 현황

가. 공공데이터 분류별 현황

우리나라에서의 공공데이터 공개는 「공공데이터 포털(data.go.kr)」 제공하고 있으며, 2014년 6월 30일 현재 698개의 기관에서 9,378건의 데이터 셋을 제공하고 있다. 전체 데이터 중 보건/복지 분야 데이터는 15.9%로 보건분야 700건, 복지분야 771건으로 총 1,471건의 공공데이터를 제공하고 있다(표 3-20 참조). 제공데이터 중 지방자치단체에서 제공하고 있는 DB는 전체의 73.6%를 차지하고 있으며, 1,082건의 데이터 셋을 제공하고 있다. 공공기관에서는 224건, 국가행정기관에서는 150건, 정부 투자기관에서는 15건의 데이터 셋을 제공하고 있다.

〈표 3-21〉은 공공데이터를 제공하고 있는 서비스 유형으로 공공데이터 포털에서 다운로드 서비스가 857건으로 가장 높았으며, 링크 서비스도 507건으로 높게 나타났다.

〈표 3-20〉 제공주체별 공공데이터 자료 현황

구분		공공 기관 ¹⁾	국가 행정기관 ²⁾	정부 투자기관 ³⁾	지방 자치단체	계
보건	건강보험	43			2	45
	보건의료	41	52	1	355	449
	식품의약품안전	2	55		149	206
복지	공적연금	23				23
	기초생활보장		1		26	27
	노동	47	20	1	7	75
	노인·청소년	7	1	3	126	137
	사회복지보건의료	2	3	4	86	95
	보육·가족및여성	2	1		1	4
	보훈	13	1		29	43

구분	공공 기관 ¹⁾	국가 행정기관 ²⁾	정부 투자기관 ³⁾	지방 자치단체	계
사회복지일반	39	14	1	222	276
주택	1	1	5	36	43
취약계층지원	4	1		43	48
계	224 (15.2)	150 (10.2)	15 (1.0)	1,082 (73.6)	1,471 (100.0)

주: 1) 공공기관(총 37개 기관): (재)한국장애인개발원, 공무원연금공단, 국가보훈처, 국립암센터, 국립중앙의료원, 국민건강보험공단, 국민연금공단, 근로복지공단, 노사발전재단, 대한적십자사, 대한주택보증주식회사, 도로교통공단, 중소기업기술정보진흥원, 축산물품질평가원, 한국건강증진재단, 한국고용정보원, 한국과학기술정보연구원, 한국광해관리공단, 한국국제보건의료재단, 한국노인인력개발원, 한국보건복지정보개발원, 한국보건산업진흥원, 한국보건의료인국가시험원, 한국보훈복지의료공단, 한국사회복지협의회, 한국사회적기업진흥원, 한국산업인력공단, 한국양성평등교육진흥원, 한국장애인고용공단, 한국청소년정책연구원, 한국환경공단, 건강보험심사평가원, 국민건강보험공단, 국립암센터, 대한적십자사, 한국보훈복지의료공단, 한국산업안전보건공단

2) 국가행정기관(총 12개 기관): 고용노동부, 국가보훈처, 국세청, 국토교통부, 병무청, 보건복지부, 식품의약품안전처, 안전행정부, 여성가족부, 중소기업청, 질병관리본부, 특허청

3) 정부투자기관(총 7개 기관): (재)한국보육진흥원, 서울시복지재단, 울산광역시도시공사, 인천광역시의료원, 코레일네트웍스, 한국청소년활동진흥원, 한국토지주택공사

※ 공공데이터 현황파악 조사: 2014.06.30. 기준

〈표 3-21〉 공공데이터 서비스 제공 유형 현황

서비스 제공 현황	보건	복지	계
단수			
LINK ¹⁾	241	266	507
OPENAPI	22	16	38
다운로드	393	464	857
복수			
그리드, 다운로드, OPENAPI	25	16	41
다운로드, LINK ¹⁾	14	3	17
분류 없음	5	6	11
계	700	771	1,471

주: 1) 개방예정 자료까지 포함

※ 공공데이터 현황파악 조사: 2014.06.30. 기준

공공데이터 포털에서 제공하고 있는 데이터 셋의 확장자 유형으로는 오피스파일 유형의 공공데이터 셋이 596건, 한글/문서 파일 유형이 210건, 웹파일 유형이 99건 순으로 제공하고 있다(표 3-22 참조).

〈표 3-22〉 공공데이터 서비스 제공 데이터 확장자 유형

확장자 유형	보건	복지	계
오피스 파일	272	324	596
웹파일	31	68	99
이미지파일	2	6	8
텍스트파일	16	17	33
한글/문서파일	74	136	210
기타	5		5
계	400	551	951

주: 1) 오피스파일: CSV, Excel, PPT, Docx 유형의 파일

2) 웹파일: JSP, HTML 유형 파일

3) 이미지파일: JPG, PNG 유형 파일

4) 텍스트 파일: TEXT, TXT 유형 파일

5) 한글/문서 파일: HWP, PDF 유형 파일

6) 기타: DICOM, SHP, STL, Zip 유형 파일

※ 공공데이터 현황파악 조사: 2014.06.30. 기준, 홈페이지에 확장자 유형이라고 공개된 정보만을 취합함

나. Open API 현황

‘공공데이터의 제공 및 이용 활성화에 관한 법률’에는 공공기관은 활용 수요가 높은 공공데이터를 민간 이용자들이 개발에 용이한 방식(오픈 API)으로 제공함으로써 신규 비즈니스 창출 촉진하도록 규정하고 있다. 오픈API는 일반적인 상용소프트웨어(아래아한글, 엑셀)에서 데이터를 편집할 수 있는 형태나 모바일 앱 개발 등에서 바로 활용할 수 있는 표준화된 방식(실시간성 데이터)을 말한다. 2014년 6월 30일 현재 보건/복지분야 오픈API 자료 현황은 〈표 3-23〉와 같이 전체 1,471개 중 보건분야 76건, 사회복지분야 24건으로 총 100건을 제공하고 있다. Open API를 제공하는 기관은 공공기관에서 9건, 국가 행정기관에서 43건, 지방자치단체에서 2건, 기타(공공데이터활용지원센터)에서 25건의 API를 제공하고 있다.

〈표 3-23〉 「공공데이터 포털」에서 보건/복지 분야 Open API 자료 현황

대분류	중분류	제공DB
보건	보건의료	30
	식품의약품안전	46
사회복지	노동	4
	노인·청소년	11
	보육·가족·여성	2
	사회복지일반	5
	주택	2
계		100

1) 보건분야

보건분야에서 서비스되고 있는 오픈API 자료 30건 중 17건이 지역자치단체의 병/의원/약국 조회서비스를 제공하고 있으며, 위생관련정보 4건, 산후조리원 현황 3건, 기타 6건으로 분류되어 있다. 기타에는 미용실, 세탁실, 목욕장, 동물병원 현황 및 위치정보 DB 등을 제공하고 있다.

〈표 3-24〉 Open API제공자료 중 보건의료분야 서비스 분류

중분류	제공DB
병/의원 약국조회	17
산후조리원현황	3
위생	4
기타	6
계	30

식품의약품안전 Open API 자료 47건은 농림수산물식품부 홈페이지에서 제공하고 있는 정보로 DB가 31건, 모범음식점 정보 3건, 화학물질, 식품 정보, 의료기기정보, 의약품정보를 포털을 통해 서비스 하고 있다.

〈표 3-25〉 Open API제공자료 중 식품의약품안전 서비스 분류

중분류	제공DB
농림수산물식품부 홈페이지 메뉴정보	31
모범음식점 정보	3
화학물질	2
식품정보	4
의료기기정보	2
의약품	4
계	46

2) 복지분야

노인·청소년 Open API에서 서비스하고 있는 DB는 총 11건으로 노인 복지 관련 정보 8건(무료 경로식당, 마을회관 주소, 노인의료시설 현황 등), 청소년 자원봉사활동 관련 정보 2건을 서비스 하고 있다(표 3-26 참조). 〈표 3-23〉과 같이 보육/가족/여성 오픈API에서 서비스하고 있는 DB는 총 2건으로 지역아동센터와 여성복지시설 현황에 대한 정보를 서비스 하고 있으며, 사회복지일반 오픈API에서 서비스하고 있는 DB는 총 5건으로 복지회관정보 2건, 요양시설 및 장애인 시설 현황 2건, 물가정보 1건을 서비스 하고 있고, 주택 오픈API에서 서비스하고 있는 DB는 총 2건으로 토지통계 및 아파트 현황과 관련한 정보를 서비스 하고 있다.

〈표 3-26〉 Open API제공자료 중 노인·청소년 서비스 분류

중분류	제공DB
노인복지	8
청소년 봉사	2
기타	1
계	11

다. 공공데이터 활용사례

공공데이터 활용사례는 <표 3-27>과 같이 보건의료분야 20건, 복지분야 24건, 식품 및 의약품 분야 6건 총 50건이 활용사례로 등록되어 있으며, 이 중 국내 API를 활용하여 개발된 건수는 26건으로 웹페이지 및 어플리케이션 형태로 서비스 되고 있다.

<표 3-27> 「공공데이터 포털」에서 공공데이터 활용사례

분류	제공DB
보건의료	20
복지	24
식품, 의약품	6
계	50

1) 보건의료분야 활용사례

보건의료분야는 주로 건강관리, 공공보건 의료분야에 활용하고 있으며, 주로 홈페이지를 통하여 건강정보를 제공하거나 진료정보를 제공하는 인터넷 건강정보서비스가 대부분 인 것으로 나타났다.

<표 3-28> 보건의료분야 활용사례

구분	분류	건수	개발 어플리케이션
국내	건강관리	3	하이닥, 유투체, 중앙대학교병원
	공공보건 의료분야	7	강동경희대학교 홈페이지, 다음지도, 수의사 처방관리 시스템, 질병관리본부 예방접종도우미, 착한병원 착한의사, 아이체크, 병원 약국 찾기
	보건산업	1	메리라떼
국외	건강관리	5	23andme, UK Dentists, Nothing haminformation prescriptions, Obesity Data, NHS Dircetory

구분	분류	건수	개발 어플리케이션
	보건산업	8	Reference NHS, Prescription search, Patient opinion, Nursing Homes Directory, NRSA Action UK, HealthyApp, Find Pharmacies, Find Gps
계		24	

2) 복지분야 활용사례

복지분야는 주로 장애인복지, 아동·청소년복지, 노인복지 등 생애주기별 복지관련 서비스를 제공하기 위해 활용하고 있으며 장애인 자립정보, 나눔 서비스, 실버 잡 등 일자리 제공을 위한 서비스를 제공하는 것으로 나타났다.

〈표 3-29〉 복지분야 활용사례

구분	분류	건수	개발 어플리케이션
국내	장애인복지	1	서울장애인 자립생활 정보
	아동·청소년복지	8	실종아동 예방 NFC 솔루션 아이프티, 나누다(봉사활동 종합정보 서비스), 월드비전, 농어촌 청소년 장학재단, 오산시청 모바일 웹, 스마트 창원, 창원시 모바일 웹, 복지로 모바일 웹페이지 및 앱
	노인복지	2	실버잡, SK텔레콤 행복창업지원센터
국외	가족·다문화	7	StreeBank, Scottish Index of Multiple Deprivation, OpenCharities, Numberhood, Compare care Homes, City of San Francisco REcreation and Parks official Mobile, Carehome.co.uk
	장애인복지	1	CareHome Map
	노인복지	1	Best Care Home
계		20	

3) 식품 의약품 분야 활용사례

식품의약품 분야는 주로 화장품 위해정보, 유통정보, 식품안전 정보를 제공하기 위해 활용하고 있는 것으로 나타났다.

〈표 3-30〉 식품 의약품 분야 활용사례

구분	분류	건수	개발 어플리케이션
국내	화장품 위해정보	2	화해(화장품을 해석하다), 피부랑
	유통정보	2	팜투 테이블, 더 마켓 인 서울
국외	식품안전	2	Scores on the Doors, Ratemyplace
계		6	





제4장

보건복지 이슈 분석 및 통계제공 현황

제1절 보건복지 이슈 분석

제2절 보건복지 통계제공 현황

4

보건복지 이슈 분석 및 << 통계제공 현황

제1절 보건복지 이슈 분석¹⁰⁾

보건복지 영역의 이슈를 감지하고, 미래 전략을 수립하기 위한 기초자료로 활용하기 위해 보건복지분야의 주요 학술지에 게재된 키워드를 분석하였다.

1. 보건의료 이슈분석

국내 보건의료 분야에서 출판된 논문의 경우, 1980년대 6만 4천여 건에서 2000년대 11만3천여 건으로 그 증가율이 약 2배임이 보고된 바 있다(Jang, Kang, Tak, Lee, 2011). 이와 같이, 기하급수적으로 증가하고 있는 논문의 수를 감안할 때, 연구자들이 국내외 보건의료 분야의 연구 동향을 파악하고 이에 따라 후속 연구의 주제 및 방향을 설정하기는 결코 쉬운 일이 아니다. 보건의료 분야의 정책입안자 또한 효과적인 정책을 수립하기 위하여 해당 분야 연구의 복잡한 상호 연결성을 이해하는데 많은 어려움이 있다(He, 1999; Mane & Borne, 2004; Jung & Jung, 2008). 국내 보건의료 분야의 연구는 각 분야별 고유한 지식체(Body of Knowledge)로서의 괄목할 만한 성장을 했을 뿐 아니라 다양한 학문과의 교류를 통해 확장되어왔다. 특히, 지식체 확장을 위하여 그 동안 축적

10) 국내보건의료분야 학회지 논문키워드의 사회연결망 분석: 한국연구재단등재지 게재논문을 중심으로, 본 원고는 국내외 학술지에 투고하기 위하여 작업된 초안임을 밝힌다.

된 학문 활동의 현황을 분석하고 이를 바탕으로 방향성을 제시하고자 하는 노력이 지속되어 왔다. 그러나 그 동안 국내 보건의료 분야의 지식체를 분석한 연구들은 보건의료 분야 전체를 거시적으로 보기 보다는 하나의 학문분야를 분석하거나(Lee, Jeong, Kim, & Yom, 2011) 특정 단위의 연구주제 (Effken, Gephart, Bianchi, & Verran, 2010; Park & Park, 1995; Roh, 2008) 또는 연구방법의 종류 (Kim, Park, Park, Chung, & Ahn, 2007; Suh et al., 2007)등을 미시적으로 분석하였거나, 계량적 접근이 아닌 정성적인 접근을 통해 연구를 진행하는 전문가에 따라 그 결과 해석에 객관성을 확보하기에 어려움이 많았다.

이와 같은 한계를 극복하기 위하여 학계에서는 일찍이 문헌정보학의 한 분과학문인 계량서지학(Bibliometrics)을 통한 내용분석을 시도하였다(Price, 1965). 이는 주로 텍스트 기반 정보를 계량화하여 분석하는 것으로 이 중에서도 동시 출현 키워드 분석(Co-word analysis)과 동시 인용 분석(Co-citation analysis) 등을 적용하면 해당 학문분야의 선도연구주제(Research Front) 또는 시기별로 새롭게 출현하는 키워드를 확인할 수 있다(He, 1999; Sohn, 2011).

최근 국내 보건의료 분야에서도 지적구조를 파악하기 위해 보건의료분야 키워드를 분석단위로 하여 기존의 다른 통계적 분석기법에 비해 개체(node) 사이의 관계를 시각화하는 기능이 뛰어난 사회네트워크 분석(social network analysis; SNA)을 시도하고 있다. 문헌의 서지정보에서 키워드를 추출하여 의료정보학(Jung, Lee, & Kim, 2010), 역학(Jeong, 2010; Kang & Park, 2010), 간호학(Lee, Jeong, Kim, & Yom, 2011) 분야에서 핵심어 네트워크를 구성하거나, 국내 주요 의과대학을 단위로 하여 각 대학 간 연구주제 네트워크를(Kang & Park, 2010) 분석한 연구들을 예로 들 수 있다.

이와 같이 SNA를 활용한 국내 연구의 대부분은 학술활동의 결과물인 출판물에서 연구자가 직접 키워드를 추출하거나 용어 색인기를 사용하여 핵심 키워드를 추출하여 분석을 한 것이다. 해외의 경우처럼 MeSH (medical subject headings) 용어를 추출하여 SNA를 적용한 예로는 Influenza (Lee, 2010) 연구와 대장암 (Sohn, 2011) 연구, 그리고 미국 국립의학도서관(National Library of Medicine: NLM)의 PubMed 데이터베이스에서 소속기관(Affiliation)을 미국 보건의료 분야 연구의 대표적 기관이라 할 수 있는 미국국립보건연구원(National Institute of Health: NIH)으로 제한하여 이 분야의 연구동향을 분석한 연구(Jang, Lee, & An, 2012)가 대표적이다.

PubMed 데이터베이스에 등재된 논문은 Medline의 심사를 받아 통과된 경우에 한해서만 MeSH 키워드를 부여받는다(Huh, 2010). MeSH 키워드란 NLM에서 규정한 의학용어를 뜻하는 것으로 논문별로 문헌정보 전문가에 의해 통일된 형태로 부여된, 계층적으로 통제된 의학용어 분류이며(Courtial, 1994; Kim, 2008; Kwon & Chae, 2002), 의학문헌의 색인과 목록에 통일성과 일관성을 부여한다는 점에서 타 데이터베이스와 구별되는 독창성을 지니고 있으며(Kim, 2008; Kwon & Chae, 2002), 키워드를 색인하는 과정에서 저자의 주관성이 배제된 키워드를 선정할 수 있다는 장점이 있다. 국내 보건의료 분야의 연구동향을 파악하기 위하여 본 연구는 MeSH 키워드가 가지고 있는 이러한 장점에도 불구하고 다음의 이유에서 PubMed 데이터베이스를 통해 MeSH 키워드를 추출하지 않았다. 첫째, PubMed 데이터베이스를 활용할 경우, 국내에서 출판되는 대부분의 국문 논문을 포함할 수 없기 때문이다. 둘째, PubMed 데이터베이스는 일정 수준 이상의 영향력 지수를 가지는 저널들을 포함하기에 본 연구에 포함되는 논문의 범위가 사전에 이미 제한을 받기 때문이다.

관심 분야에 대한 동시 출현 키워드를 SNA를 적용하여 분석하면 같은 분야 또는 동일 연구주제라 하더라도 시간의 흐름에 따라 그 내부의 일정한 변화 양상을 파악해 낼 수 있다. 따라서 국내 보건의료 분야 학술활동의 키워드를 분석함으로써 연구주제 네트워크를 시계열적으로 확인하여 제시한다면 연구자들이 연구의 방향을 설정하는 데에 도움이 될 것이다 (Chen, 2004). 또한 정책적 차원의 연구개발(Research & Development; R&D)사업의 방향 및 국가적 차원의 투자 전략을 수립하기 위한 하나의 지표로도 활용할 수 있을 것이다.

가. 연구목적

본 연구는 국내 보건의료 분야 학술활동에 축적된 지식체의 구조를 사회연결망 분석 기법을 활용하여 파악함으로써 보건의료 분야 연구주제의 현황 및 향후 연구를 위한 방향성을 제시하는데 그 목적이 있다. 국내 보건의료 분야의 지식체를 분석하기 위하여 이와 같은 접근 방법을 시도한 것은 본 연구가 처음이다.

나. 연구방법

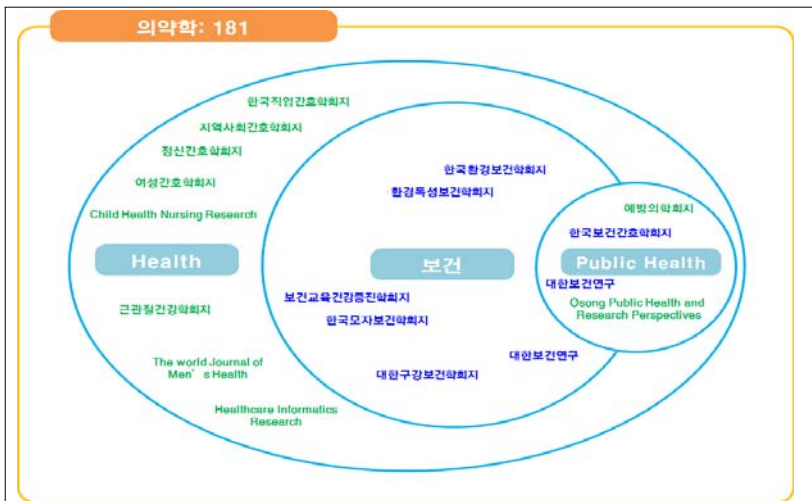
1) 연구설계

본 연구는 한국연구재단 등재학술지에 게재된 보건의료분야 논문의 저자 키워드(Author Keywords)를 사회네트워크분석(SNA) 방법을 활용하여 분석함으로써 이 분야의 연구주제의 동향을 파악하고자 한다.

2) 연구 자료 수집 대상

본 연구 대상은 2009년 1월부터 2014년 6월까지 한국연구재단의 등재학술지 목록에 포함된 학술지 중 의·약학으로 분류된 학술지 181개 중 국문으로 ‘보건’이라는 단어가 학술지 명에 포함된 등재논문을 1차 추출하였다. 그 결과, 181개의 학술지 중 8개가 검색이 검색되었으며, 영문으로 ‘Health’라는 단어가 포함된 학술지는 ‘보건’과 중복되는 학술지를 제외하고 총 10개가 검색되었다. 영문으로 ‘Public Health’라는 단어가 포함된 학술지는 총 4개였으며 ‘보건’과 ‘Health’가 중복되어 최종적으로 18개의 등재학술지에 수록된 논문의 저자 키워드를 분석 자료로 설정하였다(그림 4-1 참조).

[그림 4-1] 연구자료 수집대상



3) 자료 수집 방법 및 키워드 추출

본 연구 자료는 해당 학회지를 발행하고 있는 학회의 홈페이지 및 DBpia, RISS, KoreaMed, KISS를 통해 논문을 직접 검색하여 학회지명, 논문제목, 저자, 소속, 발행정보(Year, Volume, Page), 키워드(Key words) 및 초록(Abstract) 자료를 수집하였으며, 이 과정에서 영문 없이 국문으로만 표기된 키워드는 분석대상에서 제외하였다. 총 3,457편의 논문이 수집되었으며, 저자 키워드의 중복 및 유사 키워드를 확인한 결과, 최종적으로 총 13,465개의 키워드를 분석에 포함시켰다(표 4-1 참조).

수집된 논문 및 추출된 키워드의 수를 기준으로 할 때, ‘한국환경보건학회지’가 논문 291편(8.4%)과 1,291개(9.6%)의 키워드로 가장 높은 비중을 차지하였으며, 그 다음이 280편(8.1%)의 논문으로부터 1,111개(8.3%)의 키워드를 추출한 ‘대한구강보건학회지’, 270편(7.8%)의 논문으로부터 1,071개(8.0%)의 키워드를 추출한 ‘예방의학학회지’, 255편(7.4%)의 논문으로부터 1,017개(7.6%)의 키워드를 추출한 ‘보건교육건강증진학회지’, 252편(7.3%)의 논문으로부터 898개(6.7%)의 키워드를 추출한 ‘지역사회간호학회지’ 등의 순이었다.

〈표 4-1〉 연구자료 수집대상

번호	논문명	논문수	키워드
1	근관절건강학회지	122 (3.5)	467 (3.5)
2	대한구강보건학회지	280 (8.1)	1,111 (8.3)
3	대한보건연구	100 (2.9)	390 (2.9)
4	보건교육건강증진학회지	255 (7.4)	1,017 (7.6)
5	보건행정학회지	175 (5.1)	705 (5.2)
6	여성건강간호학회지	196 (5.7)	771 (5.7)
7	예방의학학회지	270 (7.8)	1,071 (8.0)

번호	논문명	논문수	키워드
8	정신간호학회지	214 (6.2)	821 (6.1)
9	지역사회간호학회지	252 (7.3)	898 (6.7)
10	한국모자보건학회지	117 (3.4)	457 (3.4)
11	한국보건간호학회지	171 (4.9)	637 (4.7)
12	한국직업간호학회지	170 (4.9)	595 (4.4)
13	한국환경보건학회지	291 (8.4)	1,291 (9.6)
14	환경독성보건학회지	110 (3.2)	381 (2.8)
15	Child Health Nursing Research	190 (5.5)	736 (5.5)
16	Healthcare Informatics Research	200 (5.8)	844 (6.3)
17	Osong Public Health and Research Perspectives	164 (4.7)	638 (4.7)
18	The World Journal of Men's Health	180 (5.2)	635 (4.7)
계		3,457 (100.0)	13,465 (100.0)

4) 자료 분석 방법

추출한 키워드는 용어 간 유사 키워드를 분류한 후 키워드의 출현빈도 및 가중치를 계산하였다. 유사 키워드 분류는 보건의료분야의 전문가 2인의 검토를 거쳤으며, 분석의 일관성을 위하여 키워드의 단/복수형을 복수로 통일하는 과정을 거쳤다.

선정한 키워드를 분석하기 위하여 Net Miner v. 4.0 (SYRAM, Korea)을 사용하였으며, 사회연결망을 구성하기 위하여 키워드를 노드(node)화 하여 이들 간의 연결망 구조를 분석하였다.

5) 연구결과

전체 키워드 13,465개를 대상으로 2006년부터 2014년 까지 6개 년도의 상위 50개 키워드를 네트워크로 구성한 결과 네트워크상 연결정도

중심성(Degree Centrality Value: DCV)이 높은 ($1.0 < DCV$) 키워드는 'Social Support(사회적 지지)', 'Self-esteem(자존감)', 'Quality of Life(삶의 질)', 'Health Status(건강 상태)', 'Attitude(태도)', 'Anxiety(불안)', 'Female(여성)', 'Health behavior(건강 행위)', 'Health(건강)', 'Middle aged(중년층)'으로 나타났다. 연결정도 중심성 상위 30개의 키워드를 정리하면 다음과 같다(표 4-2 참조).

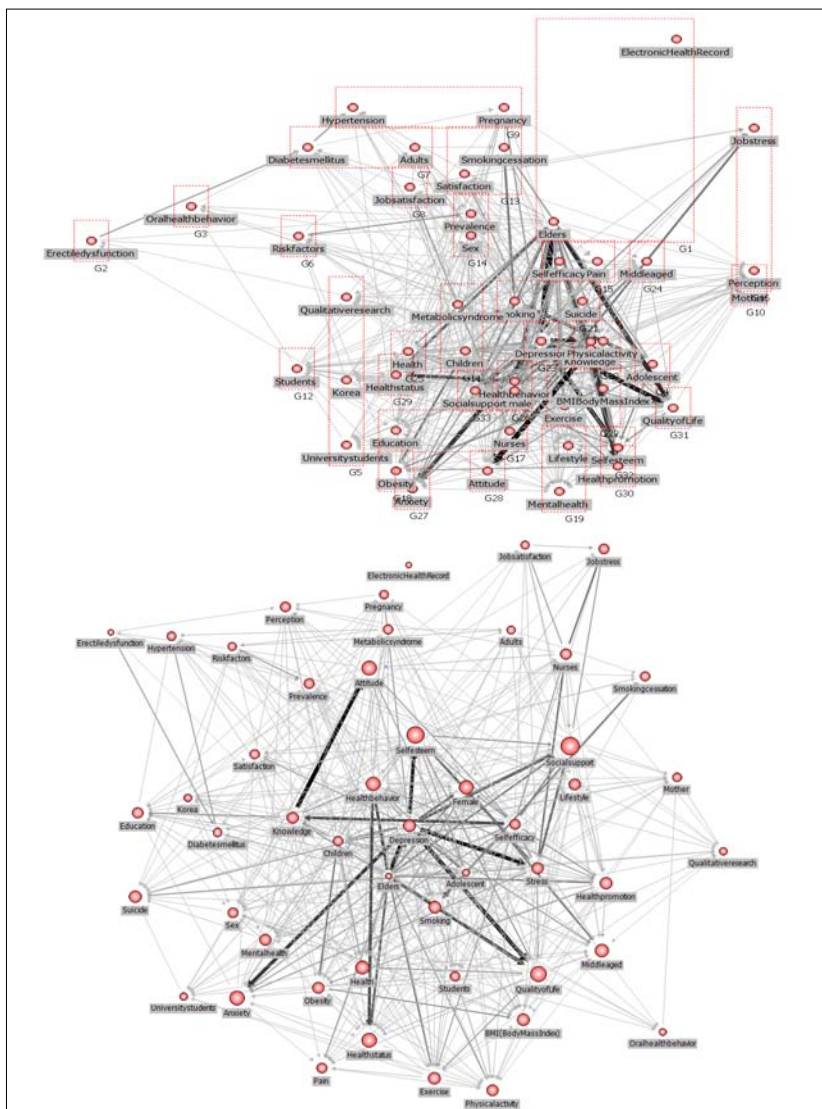
〈표 4-2〉 연결정도 중심성 상위 30위 키워드

순위	키워드	Degree Centrality Value
1	Social support	1.387755
2	Self esteem	1.285714
3	Quality of life	1.142857
4	Health status	1.020408
5	Attitude	1.020408
6	Anxiety	0.979592
7	Female	0.918367
8	Health behavior	0.918367
9	Health	0.897959
10	Middle aged	0.877551
11	BMI (Body Mass Index)	0.734694
12	Depression	0.734694
13	Exercise	0.714286
14	Health promotion	0.714286
15	Suicide	0.693878
16	Smoking	0.693878
17	Lifestyle	0.673469
18	Knowledge	0.673469
19	Stress	0.673469
20	Mental health	0.653061
21	Physical activity	0.632653
22	Obesity	0.632653

순위	키워드	Degree Centrality Value
23	Education	0.591837
24	Nurses	0.591837
25	Job stress	0.55102
26	Pain	0.530612
27	Self efficacy	0.530612
28	Perception	0.510204
29	Sex	0.510204
30	Prevalence	0.510204

전체 키워드간의 응집 구조 특성을 파악하기 위하여 분산되어 있는 키워드의 노드들의 군집(cluster)을 분석한 결과 31개 노드에 분포되어 있는 것을 알 수 있다. 키워드 간의 관계를 네트워크로 확인하면 ‘Social support(사회적 지지)’는 ‘BMI(체질량지수)’, ‘Job stress(직무 스트레스)’, ‘Satisfaction(만족도)’, ‘Health behavior(건강 행위)’, ‘Smoking(흡연)’, ‘Sex(성)’, ‘Mental health(정신 건강)’, ‘Middle aged(중년층)’등과 연결되어 연구되고 있는 것으로 분석되었으며, ‘Self-esteem(자존감)’은 ‘Knowledge(지식)’을 매개로 ‘Attitude(태도)’, ‘Self-efficacy(자기효능감)’과 강한 연결정도를 가지고 있는 것으로 나타났다. 뿐만 아니라 ‘Stress(스트레스)’, ‘Depression(우울)’이 매개가 되어 ‘Self-esteem(자존감)’과 연결되어 연구되고 있는 것으로 분석되었다. ‘Depression(우울)’은 ‘Self-esteem(자존감)’, ‘Anxiety(불안)’, ‘Stress(스트레스)’와 강한 연결정도가 있는 것으로 나타났으며, ‘Elders(노인)’을 매개로 ‘Quality of Life(삶의 질)’에 강한 연결정도가 있는 것을 볼 수 있다(그림 4-2 참조).

〔그림 4-2〕 보건의료분야의 키워드 네트워크 분석



2009년부터 2014년까지의 키워드 가운데 연결정도 중심성이 높은 순서대로 상위 30개 키워드의 변화를 분석한 결과, 2009년에는 ‘Parents

(부모)', 'Attitude(태도)', 'Quality of life(삶의 질)' 등의 순으로, 2010년에는 'Health behavior(건강 행위)', 'Suicide(자살)', 'Middle aged(중년층)' 등의 순으로, 2011년에는 'Mother(엄마)', 'Social support(사회적 지지)', 'Health(건강)' 등의 순으로, 2012년에는 'Self-esteem(자존감)', 'Job stress(직무 스트레스)', 'Self-efficacy(자기효능감)' 등의 순으로, 2013년에는 'Nurses(간호사)', 'Suicidal ideation(자살사고)', 'Quality of life(삶의 질)' 등의 순으로, 2014년에는 'Smoking(흡연)', 'Self-esteem(자존감)', 'Anxiety(불안)' 등의 순으로 나타났다(표 4-3 참조).

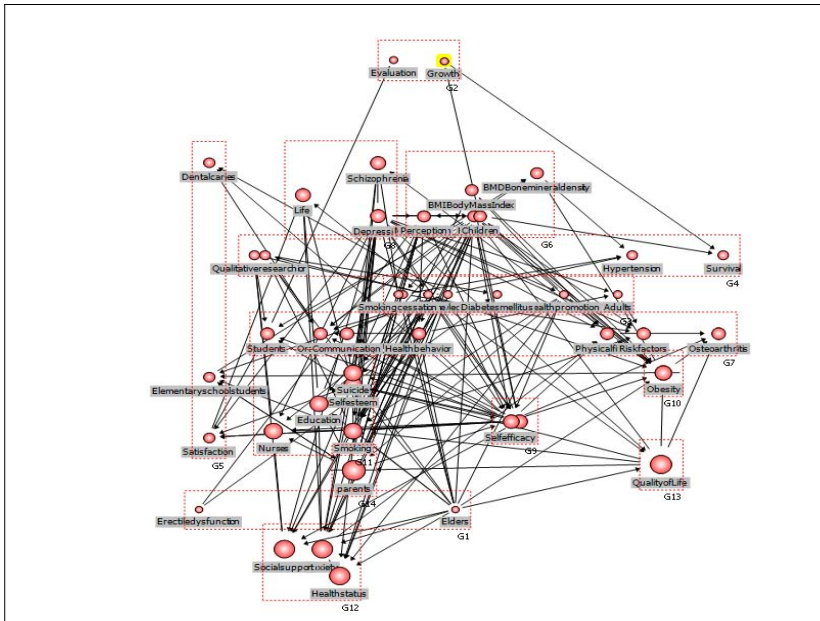
〈표 4-3〉 연도별 상위 키워드 변화

순위	2009	2010	2011	2012	2013	2014
1	Parents	Health behavior	Mother	Self esteem	Nurses	Smoking
2	Attitude	Suicide	Social support	Job stress	Suicidal ideation	Self esteem
3	Quality of Life	Middle aged	Health	Self efficacy	Quality of Life	Anxiety
4	Social support	Self esteem	Health status	Health status	Knowledge	Quality of life
5	Health status	Smoking	Pain	Immigrants	Self esteem	Lifestyle
6	Anxiety	Mental health	Female	Quality of life	Social support	Social support
7	Suicide	Activities of daily living	Sleep	Social support	Obesity	BMI
8	Education	Lifestyle	Activities of daily living	Lifestyle	Female	BMD
9	Nurses	Exercise	Physical activity	Attitude	Stress	Middle aged
10	Smoking	Anxiety	Self esteem	Health behavior	Children	Suicidal ideation
11	Health	Self efficacy	Health behavior	Empowerment	Physical activity	Nurses

순위	2009	2010	2011	2012	2013	2014
12	Self esteem	Health	Sex	Job satisfaction	Motor activity	Health
13	Obesity	Social support	Mental health	Stress	Health behavior	Satisfaction
14	Exercise	Health status	Quality of life	Depression	Stroke	Exercise
15	Self efficacy	Knowledge	Immigrants	Anxiety	Anxiety	Health promotion
16	Stress	Quality of life	Smoking cessation	Middle aged	Health promotion	Health behavior
17	Schizophrenia	Perception	Obesity	Metabolic syndrome	Immigrants	Stress
18	Life	BMI	Anxiety	Female	Middle aged	Education
19	Depression	Cancer	Education	Mortality	Health status	Data mining
20	Osteoarthritis	Attitude	Satisfaction	Long term care	Job stress	Mental health
21	Risk factors	Female	Attitude	Exercise	Self efficacy	Knowledge
22	Physical fitness	Smoking cessation	Exercise	Prevalence	parents	Breast cancer
23	Communication	Nurses	Self efficacy	Students	Metabolic syndrome	Qualitative research
24	Students	Health promotion	Stress	Health promotion	Students	Burn out
25	Oral health	Depression	Depression	Physical activity	Attitude	Environment
26	Health behavior	Osteoarthritis	ADHD	Fatigue	Elders	Chronic disease
27	BMD	Pregnancy	Pregnancy	Smoking	Intervention study	Meta analysis
28	BMI	Pain	Health promotion	Care givers	Hypertension	Job satisfaction
29	Perception	Case management	Students	Knowledge	Exercise	Male
30	Mental health	Education	Fatigue	Obesity	Smoking	Sex

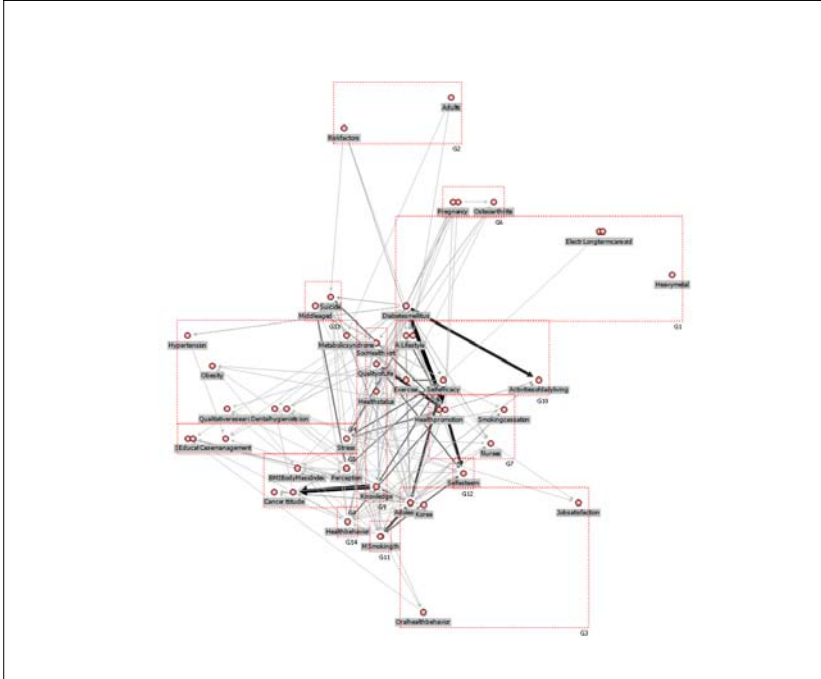
[그림 4-3]은 2009년의 키워드간의 응집 구조 특성을 파악하기 위하여 분산되어 있는 키워드의 노드들을 군집화한 결과, 14개 노드에 분포되어 있는 것을 알 수 있다. 2009년 키워드간의 네트워크 분포를 확인하면 ‘Parents(부모)’가 ‘Mental Health(정신 건강)’를 매개로 ‘Suicide(자살)’에 연결정도가 있는 것으로 확인되었다. ‘Attitude(태도)’는 ‘Parents(부모)’, ‘Suicide(자살)’, ‘Mental health(정신건강)’, ‘Self-efficacy(자기효능감)’, ‘Nurses(간호사)’, ‘Smoking(흡연)’, ‘Perception(인지)’에 연결정도가 있는 것으로 나타났으며, ‘Knowledge(지식)’가 ‘Attitude(태도)’와 강한 연결정도가 있는 것으로 분석되었다. ‘Depression(우울)’은 ‘Stress(스트레스)’, ‘Anxiety(불안)’, ‘Quality of Life(삶의 질)’, ‘Self-esteem(자존감)’과 연결되어 연구되고 있는 것으로 분석되었다.

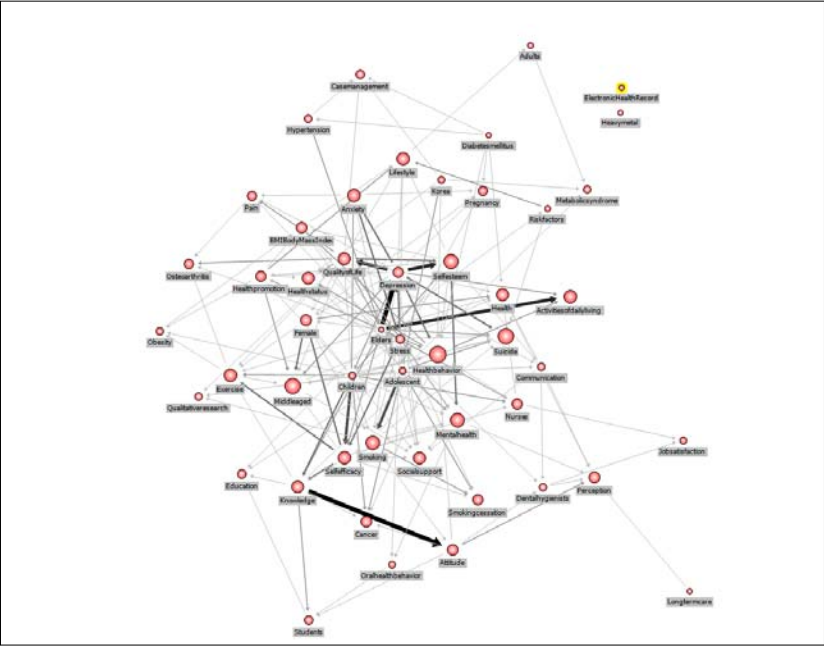
[그림 4-3] 2009년 보건의료분야 키워드 네트워크 분석



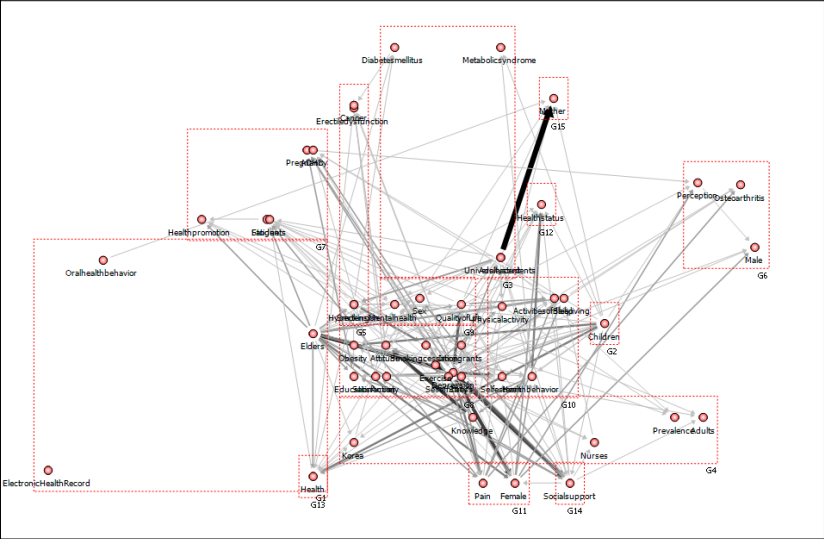
이며 총 15개 노드에 분포되어 있는 것을 확인 할 수 있다. 2011년도에 상위 키워드로 등장한 ‘Mother(엄마)’는 ‘Adolescent(청소년)’과 강한 연결 정도가 있는 것으로 나타났으며, ‘Oral health behavior(구강보건 행위)’, ‘Children(어린이)’, ‘Knowledge(지식)’이 연결 정도가 있는 것으로 분석되었으며, ‘Knowledge(지식)’을 매개로 ‘Attitude(태도)’가 ‘Mother(엄마)’와 강한 연결정도가 있는 것으로 나타났다. 이는 부모의 가치관이 자녀에게 건전한 사고방식 및 건강행위로 이어지는 전달 통로이기 때문임을 감안한다면 당연한 결과라고 할 수 있다. ‘Social support(사회적 지지)’는 ‘Health behavior(건강행위)’를 매개로 ‘Health status(건강상태)’로 강한 연결정도가 있는 것으로 나타났다.

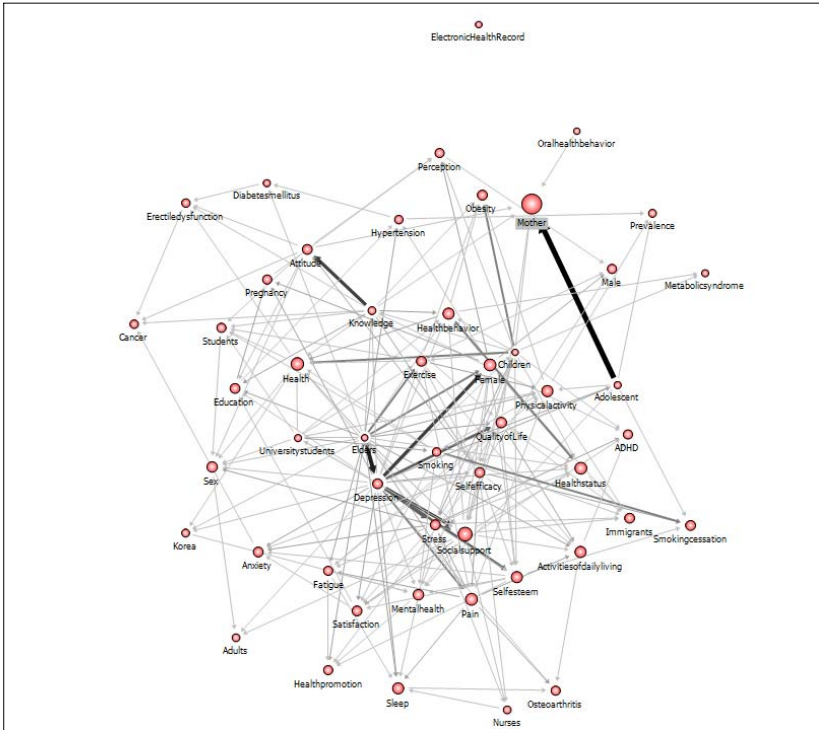
[그림 4-4] 2010년 보건의료분야 키워드 네트워크 분석





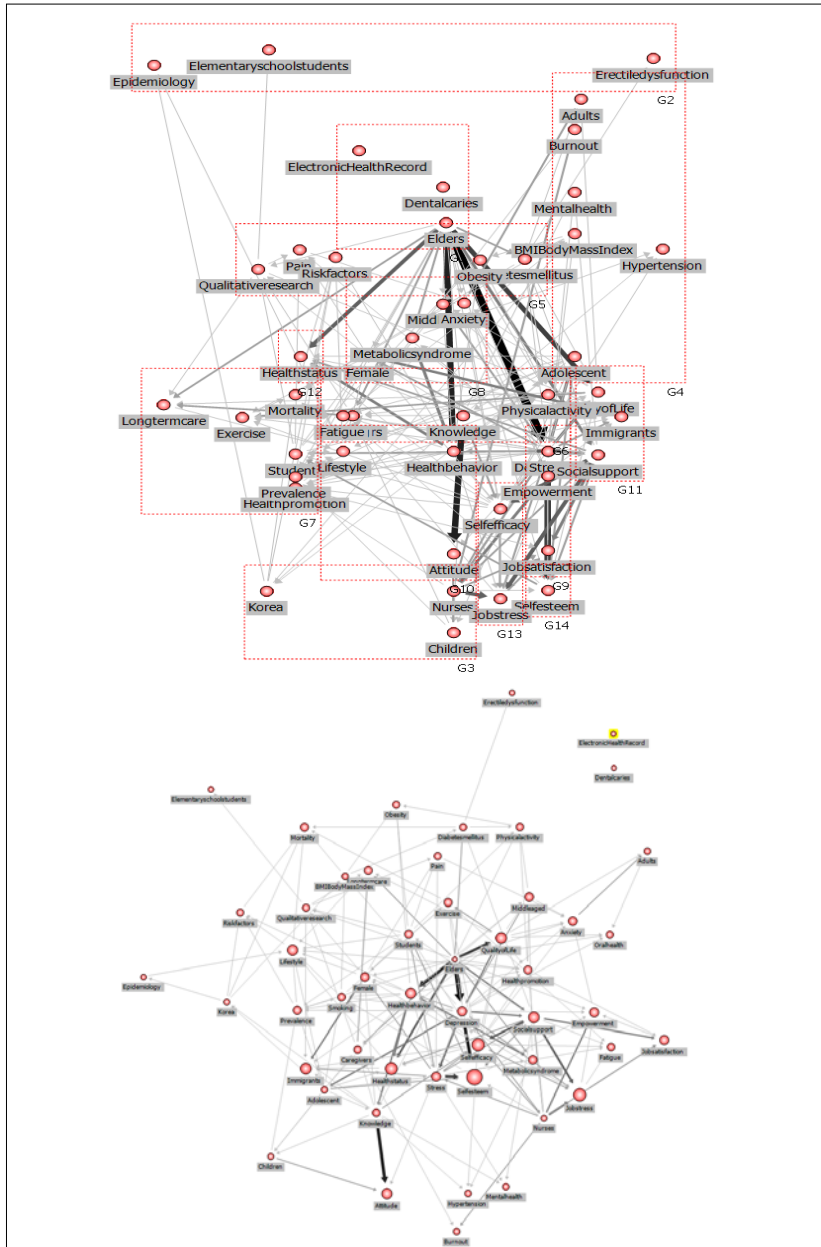
[그림 4-5] 2011년 보건의료분야 키워드 네트워크 분석





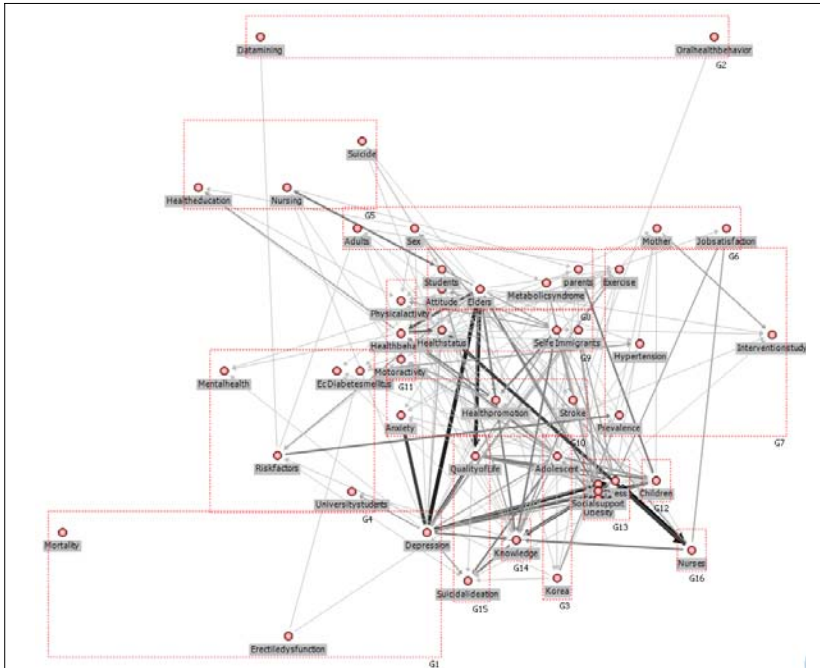
[그림 4-6]은 2012년의 키워드 간 응집 구조 특성을 분석한 결과 14개 노드에 분포되어 있는 것을 알 수 있다. 2012년 키워드 간의 네트워크 분포를 살펴보면, 'Self-esteem(자존감)'은 'Smoking(흡연)', 'Students(학생)', 'Empowerment(권한위임)'가 연결정도가 있는 것으로 나타났으며, 'Quality of Life(삶의 질)'이 'Stress(스트레스)'를 매개로 'Self-esteem(자존감)'과 매우 연결강도가 높은 것으로 나타났다.

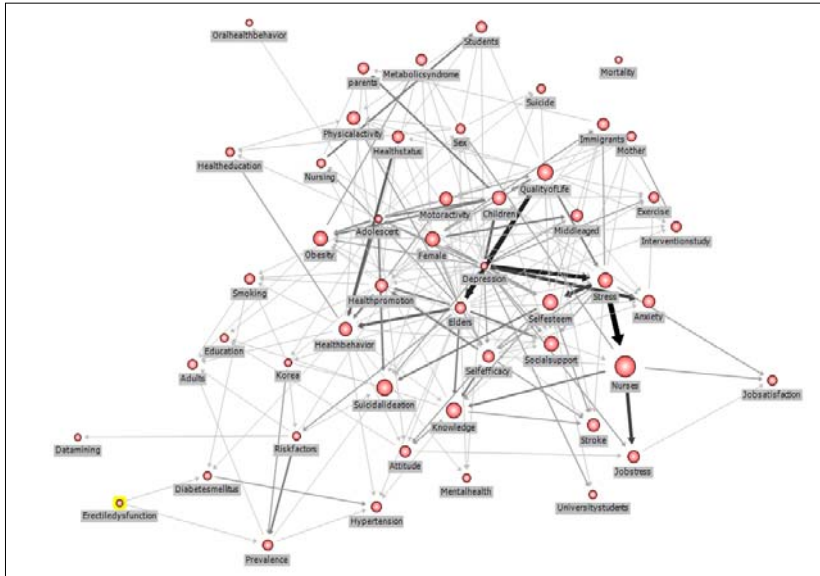
[그림 4-6] 2012년 보건의료분야 키워드 네트워크 분석



[그림 4-7]은 2013년 키워드 간 응집 구조 특성 결과 16개 노드에 분포되어 있었으며, 'Nurses(간호사)'를 매개로 'Sex(성)', 'Job satisfaction(직무 만족도)', 'Knowledge(지식)', 'Self-efficacy(자기효능감)', 'Job stress(직무 스트레스)'등과 연관되어 연구를 하고 있는 것으로 나타났으며, 'Job stress(직무 스트레스)', 'Stress(스트레스)'와는 강한 연결정도가 있는 것으로 확인되었다. 'Suicidal ideation(자살 사고)'은 'Stress(스트레스)', 'Adolescent'와 강한 연결 정도가 있는 것으로 나타났으며, 'Risk factors(위험요인)', 'Elders(노인)', 'Female(여성)', 'Children(어린이)' 등과 연관되어 연구가 활발히 진행되고 있다고 볼 수 있다.

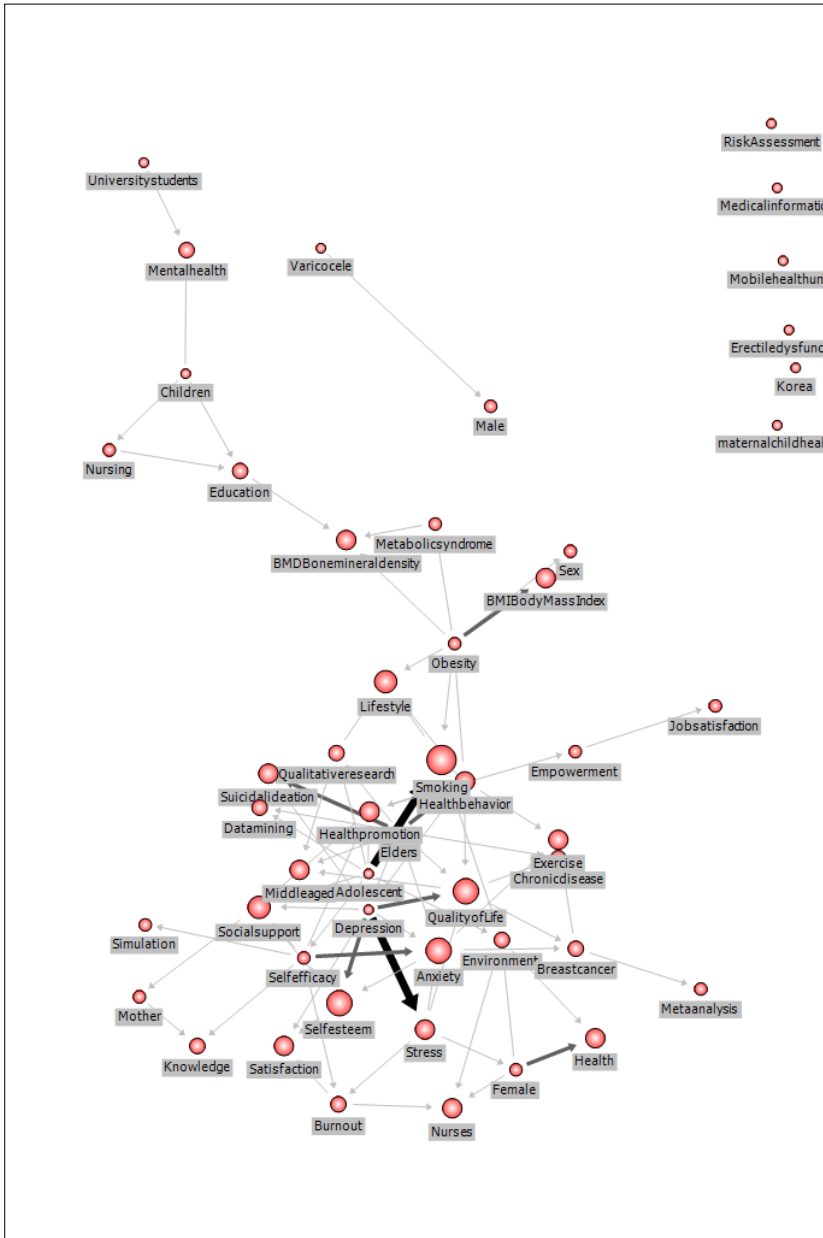
[그림 4-7] 2013년 보건의료분야 키워드 네트워크 분석

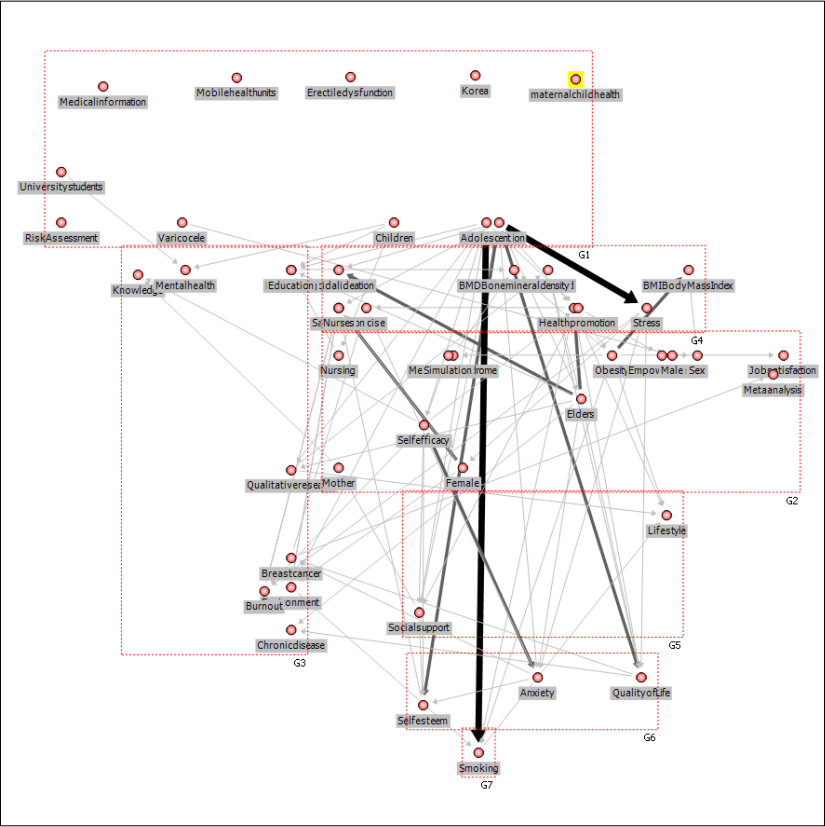




[그림 4-8]은 2014년 키워드 군집은 총 7개 노드에 분포되어 있는 것으로 나타났으며, 키워드 간 네트워크 분석 결과 ‘Adolescent(청소년)’와 ‘Smoking(흡연)’, 그리고 ‘Stress(스트레스)’와 강한 연결정도가 있는 것으로 나타났다. ‘Self-esteem(자존감)’은 ‘Social support(사회적 지지)’와 연관 되어 연구가 진행되고 있음을 예측할 수 있으며, ‘Self-efficacy(자기효능감)’와 ‘Anxiety(불안)’의 연결 관계로 ‘Self-esteem(자존감)’과 연결되고 있는 것으로 나타났다.

[그림 4-8] 2014년 보건의료분야 키워드 네트워크 분석





2. 사회보장분야 이슈분석¹¹⁾

저출산·고령화로 사회적 위험의 증가와 더불어 복지욕구가 다양해짐에 따라 사각지대 해소, 사회보장 전달체계의 합리화 등 사회보장부분의 현안해결 방법으로 공공부조, 사회보험, 사회서비스 등 다양한 사회보장 영역의 데이터 연계 및 활용방안이 강조되고 있다. 사회복지 전달체계는 중앙부처, 지자체, 공공기관, 민간 시설 및 단체, 개인 등 다양한 이해관계가 존재하고 있으며, 사회복지 대상자도 저소득층을 중심의 전통적 취약계층에서 일반국민으로 확대되는 추세에 있다. 사회보장의 범위는 출산, 양육, 실업, 노령, 질병, 빈곤 등 각종 사회적 위험으로부터 전 국민을 보호하고 삶의 질을 향상 시키는데 필요한 소득, 서비스를 보장하는 사회보험, 공공부조, 사회서비스를 포괄하고 있다. 사회보장 부분의 빅데이터 수요를 파악하기 위해서는 사전적으로 사회복지 빅데이터 적용가능 영역을 검토하는 것이 필요하다. ETRI(2014)의 연구에서 2012년 이후 주요 학술지에 게재된 논문들의 키워드를 분류하고 군집화하여 사회보장 부분의 학술문헌에 대한 이슈 분석 결과는 <표 4-4>와 같이 나타났다. 사회보장 일반 부문에서는 사회적자본, 사회적기업, 고용불안, 소득불평등, 사례관리, 네트워크 등이 주요 키워드로 나타났다. 사회보험 부문에서는 사회보험 명칭(국민연금, 건강보험, 산재보험, 고용보험), 의료비 부담, 형평성 등이 주요 키워드로 나타났다. 공공부조 부문에서는 소득, 빈곤선, 근로빈곤, 소득보장 등이 주요 키워드로 나타났고, 사회서비스 부문에서는 전달체계, 바우처, 돌봄, 보육, 서비스 질, 품질평가 등이 주요 키워드로 나타났다.

11) 본 내용은 'ETRI(2014.2). 사회보장부분 빅데이터 R&D 사업기획연구'를 분석 재처리함.

〈표 4-4〉 사회보장 부분 학술문헌 이슈분석

구 분		키 워 드
사회 보장	일반	네트워크, 비영리조직, 기부행동, 리더십, IT자원, 업무효율성, 삶의 질 향상, 조세체계, 소비세, 개인소득세, 법인소득세, 조세지출, 복지국가, 기부 노력, 행복감, 사회적 신뢰, 상호작용효과, 사례관리, 사회적 기업, 자원봉사, 자발성, 무보수성, 공익성, 권리, 감성자본, 지방정부 재정운용, 민관협력, 거버넌스, 지역사회복지협의체, 고용서비스, 시민조직, 사회구성주의, 양육스트레스, 문화적 역량, 일가족 양립, 결혼이주여성, 사회적지지, 장애인, 사회적 배제, 노동이주여성, 이주의 여성화, 젠더, 문제음주, 배우자 폭력, 피해자의 문제음주, 기부, 자원봉사, 소득불평등, 소득이동성, 소득불안정성, 체험휴, 사회보장기본법, 근로연계복지, 청년빈곤, 고용불안, 소득불평등, 새로운 사회적 위험구조, 소득보장지출, 사회투자지출, 사회지출구조 변화, 자선적 기부, 종교적 기부, 상호부조적 기부, 기부규모, 기부노력, 전일제, 시간제, 일가정 인식, 복지정책, 복지 축소, 사회복지지출, 다문화가족지원법, 협동조합, 임파워먼트, 디스임파워먼트, 결혼이주여성, 근로자지원제도, 근로자지원프로그램, EAP, 노동조합, 노조형 EAP
	사회보험	노인장기요양보험제도, 연금금 운용제도, 공무원연금, 보건의료비, 직업복귀기간, 연금개혁, 국민연금, 국민연금기금, 과부담 의료비 지출, 의료 접근성, 미충족 의료, 의료보장, 건강보험, 의료급여, 산재보험, 민영화, 운영주체, 건강불평등, 직업병, 산업복지, 노동정치, 국민연금부채, 부채벤치마크, 의료이용, 지역 간 차이, 형평성, 고용서비스, 실업급여, 고용센터, 적극적 노동시장정책, 보험료 체납자, 체납자 관리 방안, 과부담의료비, 의료비 부담, 건강불평등, 산재장애인, 직업복귀, 건강보험 지역가입자, 직장가입 확대, 직장피부양자, 임금근로자
	공공부조	가족부양자, 부양부담, 급여청구권, 복지조치청구권, 권리구제, 상대소득, 과거소득에 의해서 형성된 습관, 자영업, 근로소득, 불평등, 근로빈곤, 국민 기초생활보장제도, 자활사업, 사회복지예산, 소득보장, 지방재정조정제도, 이전재원, 빈곤선, 최저소득기준, 공공부조급여기준선, 최저생계비, 저임금 근로, 근로빈곤, 사회안전망, 부양비, 부양능력미약, 기본소득, 포괄적 수당
	사회 서비스	폭력, 아동보호전문기관, 노인복지관, 사회복지조직, 사회복지시설평가, 서비스 만족도, 장애영아, 조기개입 서비스, 장애아동복지지원법, 지역장애아동지원센터, 서비스 전달체계, 직무만족, 이직의도, 지역자활센터, 사회서비스 바우처, 사회서비스 산업, 조직혁신성, 조직성과, 서비스 질, 이용자만족도, 산모신생아, 방문요양, 서비스 관리, 서비스 구조, 서비스 과정, 서비스 결과, 미혼모자시설, 보수체계, 보수실태조사, 처우, 지역사회기반 서비스, 품질평가체계, 품질관리 전담기구, 지역사회복지관, 지역사회복지실천, 근린지역사회조직화모델, 자폐성 장애아동 어머니, ADHD 아동, 푸드뱅크, 기부식품 제공사업, 공공사회서비스 투자, 보육정책, 보육의 시장화, 서비스 제공자와 이용자, 돌봄 서비스 정책, 사회적 돌봄정책

자료: ETRI(2014.2). 사회보장부분 빅데이터 R&D 사업기획연구. p63-64.

제2절 보건복지 통계제공 현황

1. 보건복지부 통계

보건복지부 정책통계담당관에서 운영하는 통계 포털(stat.mw.go.kr)은 국가통계통합DB 관리시스템에서 통계를 작성하여 보건복지부 통계포털과 국가통계포털(kosis.kr)에서 링크 서비스를 지원하고 있으며, 제공하는 통계는 보건 및 복지 분야 통계를 엑셀파일 형태로 제공하고 있다. 보건복지부 통계포털에서 제공하는 지표는 보건/의료, 질병/건강, 보육/아동, 노인복지, 장애인 복지, 사회, 보험/연금으로 분류하여 승인된 조사 통계를 링크(link)서비스 형태로 제공하고 있다. 보건 분야 지표는 국민 보건의료 실태조사, 국민의료비 추계 및 국민 보건 계정, 병원경영 실태조사, 보건소 및 보건지소 운영현황, 의료기기 화장품 제도, 유통실태 조사, 의약품 소비량 및 판매액 통계, 퇴원 손상 심층조사, 한방의료 및 한약 소비 실태 조사, 환자조사 순으로 제공되고 있으며 총 518건의 통계표를 서비스 하고 있다. 질병/건강 분야 지표는 17종의 승인통계를 서비스 하고 있으며 827건의 통계표가 등록되어 있다. 보육/아동 복지 분야는 7종의 승인통계를 서비스 하고 있으며 총 414건의 통계를 서비스 하고 있다. 노인 분야는 6종, 155건의 통계표를 서비스 하고 있으며, 장애인 분야는 8종, 793건의 통계표를 서비스 하고 있다. 사회 분야는 12종, 3,511건의 통계표를 제공하고 있으며, 보험/연금 분야는 5종, 537건의 통계표를 서비스 하고 있다. 보건복지부 통계포털은 전체 6,755건의 통계표를 서비스 하고 있다(표 4-5 참조).

〈표 4-5〉 보건복지부 통계포털 제공통계 목록

주제별 분류	통계명	지표수	소계
보건	국민보건의료실태조사	112	518
	국민의료비 추계 및 국민 보건 계정	3	
	병원경영실태조사	294	
	보건소 및 보건지소 운영현황	5	
	의료기기, 화장품 제조, 유통실태	26	
	의약품 소비량 및 판매액 통계	2	
	퇴원 손상 심층조사	34	
	한방의료 이용 및 한약 소비 실태 조사	30	
	환자조사	12	
질병/건강	HIV/AIDS 신고 현황	16	827
	결핵현황	4	
	공중위생관계업소 실태보고	3	
	국민건강영양조사	97	
	국민구강건강실태조사	116	
	급성 심장정지 조사	-	
	법정 감염병 발생보고	6	
	사망원인 통계	12	
	수입 식품 현황	44	
	식품 수거 검사 실적	20	
	암등록 통계	11	
	인수공통 감염병 위험군의 감염실태조사	286	
	전국민장내기생충감염조사	-	
	정신질환실태조사	104	
	지역사회 건강조사	35	
	청소년 건강행태 온라인 조사	31	
	한국인 인체 치수 조사	42	
보육/아동 복지	가족위탁 국내 입양 소녀 소녀 가정	5	414
	보육실태조사	95	
	아동복지시설 보호 아동 및 종사자	1	
	아동종합실태조사	271	
	어린이집 이용자 통계	11	
	요보호 아동 현황 보고	1	
	학대 피해아동 보호 현황	30	
노인 복지	고령자 고용 현황	5	155
	국민노후보장 패널 조사	-	
	노인복지시설 현황	4	
	노인실태 조사	80	
	노인학대 현황	57	
	노후 준비 실태 조사	9	
장애인 복지	기업체 장애인 고용 실태 조사	324	793
	장애인 고용 패널 조사	197	

주제별 분류	통계명	지표수	소계
	장애인 구인 구직 및 취업동향	107	
	장애인 등록 현황	6	
	장애인 생활체육 실태 조사	143	
	장애인 실태조사	5	
	장애인 의무고용현황	6	
	장애인 편의시설 설치 현황 조사	5	
사회	가족실태조사	342	3,511
	국민기초생활수급자현황	6	
	국민노후보장패널조사	-	
	사회복지자원봉사현황	43	
	사회조사	1,572	
	여성가족패널조사	108	
	유망사회서비스 수요 실태 조사	86	
	전국 결혼 및 출산동향조사	75	
	전국출산력 및 가족보건복지 실태조사	213	
	최저생계비계측조사	17	
	한국복지패널조사	1,045	
	한국의 사회복지 지출	4	
보험/연금	건강보험 주요 수술 통계	29	537
	건강보험통계	265	
	국민연금통계	-	
	다층노후소득보장체계실태조사	-	
	지역별 의료이용 통계	243	
		6,755	

그리고 <표 4-6>과 같이 보건복지부 통계포털에서 보건복지분야 주요 지표를 선별하여 서비스 하고 있다. 보건복지분야 주요 지표는 건강상태 및 질병, 보건의료, 노인복지, 아동복지, 장애인 복지, 사회복지일반, 사회보험 및 사회보장, 국민계정 및 경제관련 사회통계로 분류하여 총 94건의 지표를 선별하여 서비스하고 있다. [그림 4-9]과 같이 제공 지표 중 보건복지부 제공 통계가 39종, 통계청 15종, 질병관리본부가 10종, OECD 7종, 국민건강보험공단 7종, 건강보험 심사평가원 3종, 국민연금공단 3종, 경찰청, 대검찰청, 대한적십자사, 식품의약품안전처, 전국경제인연합회, 한국보건사회연구원, 한국은행에서 각각 1종의 통계를 제공하고 있다.

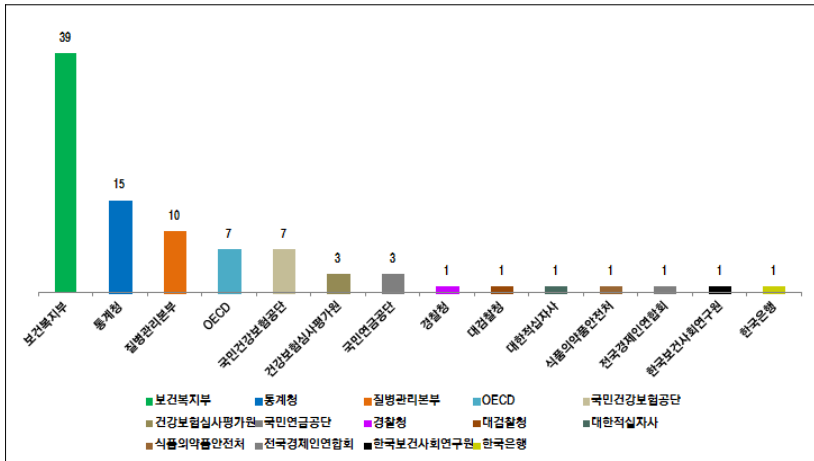
〈표 4-6〉 보건복지 주요지표

분류기준	해당지표	출처	제공기관
건강상태 및 질병	BCG 예방접종률	전국 예방접종률 조사	질병관리본부
	헌혈인구	혈액사업통계연보	대한적십자사
	신고결핵 신환자수	결핵환자신고현황연보	질병관리본부
	신규HIV/AIDS 신고현황	HIV/AIDS신고현황 연보	질병관리본부
	마약류 사범수	마약류 범죄백서	대검찰청
	자살률	사망원인통계	통계청
	OECD국가 자살률 비교	OECD Health Data	OECD
	주요 만성질환 사망률	사망원인통계	통계청
	만성질환 유병률	국민건강통계	질병관리본부
	월간 음주율	국민건강통계	질병관리본부
	OECD 국가 주유 소비량	OECD Health Data	OECD
	현재 흡연율	국민건강통계	질병관리본부
	OECD 국가 흡연인구 비율	OECD Health Data	OECD
	비만 유병률	국민건강통계	질병관리본부
	조사망률	사망통계	통계청
	합계출산율	출생통계	통계청
	우울증 환자수	건강보험 심사실적	건강보험심사평가원
	주요 암 5년 상대 생존율	국가암등록사업 연례보고서	보건복지부
	암발생률	국가암등록사업 연례보고서	보건복지부
	모성사망비	사망원인통계	통계청
	기대수명	생명표	통계청
	영아사망률	사망원인통계	통계청
보건의료	GDP 대비 제약산업 비중	식품의약품통계연보	식품의약품안전처
	GDP 대비 국민 의료비	OECD Health Data	OECD
	국민의료비 중 공공지출 비율	OECD Health Data	OECD
	항생제 처방률	통계정보공개부	건강보험심사평가원
	주사제 처방률	통계정보공개부	건강보험심사평가원
	뇌사장기 이식 건수	장기이식정보시스템	질병관리본부
	뇌사장기 기증자 수	장기이식정보시스템	질병관리본부
	국가 암 검진 수검률	국가암검진사업수검률	보건복지부
	응급의료기관 수	응급의료과	보건복지부
	시도 공공의료기관 현황	보건의료정보시스템	보건복지부
	건강검진 수진율	국민건강통계	질병관리본부

분류기준	해당지표	출처	제공기관
	치과의사 1인당 인구	보건복지부통계연보	보건복지부
	의사 1인당 인구	보건복지부통계연보	보건복지부
	OECD 국가 병상수 비교	OECD Health Data	OECD
	의료시설현황	보건의료정보시스템	보건복지부
	의료인 수	보건복지부 면허관리시스템	보건복지부
노인복지	65세 이상 1인 가구	장래인구가구추계	통계청
	치매노인 수	치매유병률조사	보건복지부
	화장률	장사문화 발전을 위한 국가전략계획 수립연구	한국보건사회연구원
	기초노령연금 수급자 수	기초노령연금과	보건복지부
	정부지원 노인 일자리 수	노인지원과	보건복지부
	노령화지수	장래인구추계	통계청
	노인 1인당 월 진료비	건강보험 주요통계	국민건강보험공단
	노인복지 시설 수	노인복지시설 현황	보건복지부
	OECD 국가의 노인 경제활동 참가율	OECD	OECD
	노년 부양비	장래인구추계	통계청
	노인인구비율 추계	장래인구추계	통계청
아동복지	실종아동 발생 수	실종아동찾기 센터 통계	경찰청
	학대피해아동 보호건수	학대피해아동보호현황	보건복지부
	입양아동 수	국내입양소년소녀가정현황	보건복지부
	국외 입양 현황	국내입양소년소녀가정현황	보건복지부
	요보호아동 발생인원	요보호아동 현황보고	보건복지부
	보육료지원 아동 수	보육통계	보건복지부
	취약(특수) 어린이집 현황	보육통합정보시스템	보건복지부
	아동급식 대상자 수	아동권리와	보건복지부
	어린이집 현황	보육통합정보시스템	보건복지부
장애인 복지	장애인 생활시설 수	장애인권익지원과	보건복지부
	장애인 편의시설 설치율	장애인편의시설 실태전수조사 보고서	보건복지부
	장애아동수당 수급자 수	장애인서비스과	보건복지부
	장애인 직업재활 시설수	장애인 직업재활시설 운영실적 보고서	보건복지부
	장애인 지역 사회 재활시설 수	장애인권익지원과	보건복지부

분류기준	해당지표	출처	제공기관
사회복지 일반	등록장애인 수	등록장애인현황	보건복지부
	장애수당 수급자 수	장애인서비스과	보건복지부
	긴급복지지원 실적	-	보건복지부
	자활사업 참여자 수	자립지원과	보건복지부
	의료급여 수급자 총 진료비	기초의료보장과	보건복지부
	기업의 사회공헌 규모	기업재단 사회공헌백서	전국경제인연합회
	사회복지시설 현황	사회복지시설 현황	보건복지부
	보건복지부 사회서비스 일자리 사업 예산	-	보건복지부
	보건복지관련사업 일자리 수	보건복지관련사업 일자리 조사	보건복지부
	사회복지자원봉사 참여율	사회복지자원봉사 통계연보	보건복지부
	최저생계비	최저생계비 계측조사	보건복지부
	사회복지전담 공무원	지역복지과	보건복지부
	GDP대비 사회복지지출 규모	사회복지지출추계	보건복지부
	기초생활보장 수급자 수	국민기초생활수급자현황	보건복지부
사회보험 및 사회보장	건강보험재정 정부지원금	건강보험 주요통계	국민건강보험공단
	1인당 평균 요양기관 입·내원 일수	건강보험 주요통계	국민건강보험공단
	지역보험료 징수율	건강보험 주요통계	국민건강보험공단
	건강보험 재정수지	건강보험 주요통계	국민건강보험공단
	건강보험료	건강보험 주요통계	국민건강보험공단
	건강보험 가입자 수	건강보험 주요통계	국민건강보험공단
	국민연금기금 운용 수익률	국민연금통계	국민연금공단
	국민연금 수급자 수	국민연금통계	국민연금공단
	국민연금 가입자 수	국민연금통계	국민연금공단
국민계정 및 경제관련 사회통계	GDP 성장률	국민계정	한국은행
	1인당 GNI	국민통계연감	통계청
	지니계수(가처분 소득기준)	가계동향조사	통계청
	조세부담률	한국의 사회지표	통계청
	장래인구추계	장래인구추계	통계청

[그림 4-9] 보건복지부 주요지표 제공기관 현황



2. 통계청 e-나라지표

통계청에서 운영하는 e-나라지표(index.go.kr)는 국가 공식 승인 통계뿐만 아니라 각종 행정자료들을 가공한 내용들을 포함하여 이용자들에게 도식화하여 제공하고 있다. <표 4-7>은 e-나라지표에서 제공하는 보건 분야 제공 목록이다. 보건 분야에서 제공하는 지표의 주제별 분류는 건강증진, 보건산업, 보건일반, 식품, 질병으로 분류되어 있다. 건강증진 분야 지표는 4종, 보건산업분야 2종, 보건일반 6종, 식품분야 2종, 질병 분야 6종으로 총 20종의 지표들을 서비스하고 있다. 건강증진 분야에서 제공하는 통계는 비만도, 운동, 평균수명 및 건강수명, 음주 및 흡연, 청소년의 흡연율에 관한 통계 등을 제공하고 있으며, 보건산업 분야에서 제공하는 통계는 의약품 및 의료기기의 생산실적 통계를 제공하고 있다. 보건일반 분야에서는 GDP대비 국민 의료비 추이, 항생제 및 주사제 처방율, 헌혈 인구 및 개인헌혈 비율, 뇌사자 장기 기증 현황, 영아 및 모성 사

망 통계, 의료인력 및 병상 수 추이에 대한 통계지료를 서비스하고 있다. 식품분야는 식품검사 부적합률, 식품 및 식품 첨가물 생산 실적을 지표로 제공하고 있으며, 질병 분야는 만성질환, 국가 암 검진, 정신질환, 식중독 신고 건수 및 환자 수 등에 대한 지표를 제공하고 있다.

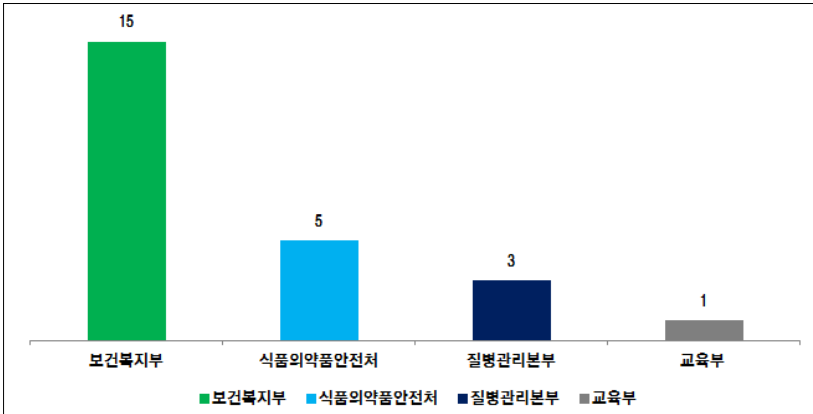
〈표 4-7〉 e-나라지표 보건지표 제공 목록

주제별 분류	통계명	출처	제공기관
건강 증진	비만도 및 규칙적 중증도 운동 실천율	국민건강통계	보건복지부 질병관리본부
	평균수명 및 건강수명	통계청, 생명표 WHO	보건복지부
	음주 및 흡연현황	국민건강통계	보건복지부 질병관리본부
	청소년 현재 흡연율, 청소년 현재 음주율	청소년건강행태 온라인조사	보건복지부 교육부
보건 산업	의약품 생산실적	식품의약품통계연보	식품의약품안전처
	의료기기 생산실적	식품의약품통계연보	식품의약품안전처
보건 일반	GDP대비 국민의료비 추이	국민의료비 및 국민보건계정, OECD Health Data	보건복지부
	항생제 및 주사제 처방률	건강보험심사평가원	보건복지부
	헌혈인구 및 개인헌혈 비율	혈액사업통계연보	보건복지부
	뇌사자 장기기증 현황	장기이식 및 이식통계연보	보건복지부
	영아/모성 사망	영아모성사망조사	보건복지부
	의료인력 및 병상수 추이	보건복지부통계연보	보건복지부
식품	식품검사 부적합률	식품의약품통계연보	식품의약품안전처
	식품 및 식품첨가물 생산실적	식약처 보고자료	식품의약품안전처
질병	만성질환 현황	국민건강통계	보건복지부 질병관리본부
	국가 암 조기검진사업 수검률	국가암검진사업 정보시스템	보건복지부

주제별 분류	통계명	출처	제공기관
	정신질환 유병률	정신질환실태조사	보건복지부
	법정 감염병 발생현황	감염병감시시스템 및 통계시스템	보건복지부
	집단 식중독 신고건수 및 환자수	집단식중독발생현황	식품의약품안전처
	암 발생 및 사망현황	암등록통계	보건복지부

e-나라지표 ‘보건분야’에 지표를 제공하는 기관은 총 4개 기관으로 [그림 4-10]과 같다. 보건복지부에서 15종, 식품의약품안전처에서 5종, 질병관리본부 3종, 교육부 1종을 제공하고 있다.

[그림 4-10] 통계청 e-나라지표 보건지표 제공기관



〈표 4-8〉은 e-나라지표에서 제공하는 복지분야 제공 목록이다. 복지 분야에서 제공하는 지표의 주제별 분류는 노인복지, 복지일반, 아동복지, 연금 및 건강보험, 장애인 복지, 저출산 고령화로 분류하고 있다. 노인복지 분야에서 제공하는 지표는 4종, 복지 일반에서 제공하는 지표는 11종, 아동복지 분야 5종, 연금 및 건강보험 분야에서 제공하는 지표는 5종, 장

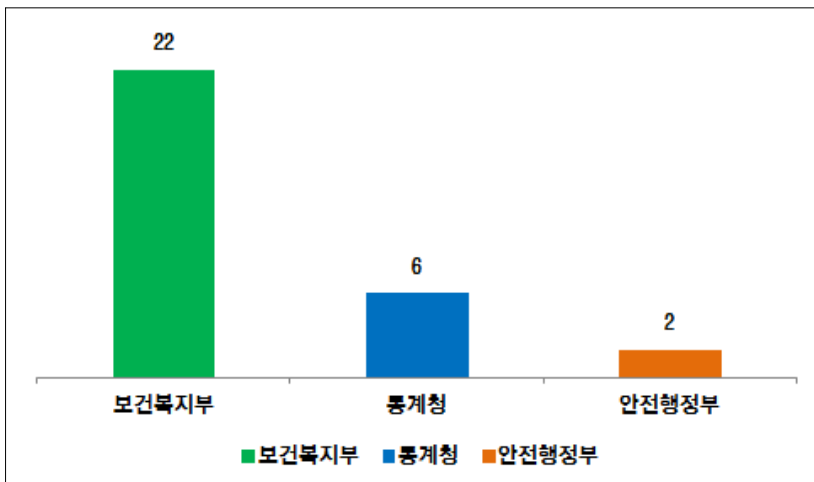
애인 복지 3종, 저출산 고령화 분야에서 제공하는 지표는 2종으로 총 30개의 복지관련 지표를 서비스하고 있다. [그림 4-11]은 e-나라지표 ‘복지분야’에 지표를 제공하는 기관별 지표종수 표이다. 보건복지부에서 22종, 통계청에서 6종, 안전행정부에서 2종을 서비스 하고 있다.

〈표 4-8〉 e-나라지표 복지지표 제공 목록

주제별 분류	통계명	출처	제공기관
노인 복지	노인 취업률 현황	노인실태조사	보건복지부
	노후 준비 방법	사회조사	통계청
	노인복지시설 현황	노인복지시설현황	보건복지부
	노인 일자리 및 노후생활 현황	사회조사	보건복지부
복지 일반	보건복지부 예산현황	연도별 예산자료	보건복지부
	의료급여 수급현황	의료급여통계	보건복지부
	소득분배	가계동향조사	통계청
	향후 늘려야 할 복지서비스	사회조사	통계청
	현금 기부 인구	사회조사	통계청
	내년 가구의 재정상태	사회조사	통계청
	사회복지 지출규모	한국의 사회복지지출	보건복지부
	국민기초생활보장 수급현황	내부자료	보건복지부
	최저생계비 및 빈곤율 추이	최저생계비 계측조사 연구	보건복지부
	국민영양 현황	국민건강영양조사	보건복지부
	자치단체 사회복지 비중	지방자치단체 예산개요	안전행정부
아동 복지	요보호아동현황	시도별 보고자료	보건복지부
	학대아동 보호 건수	전국아동학대 현황보고서	보건복지부
	아동급식지원 현황	내부자료	보건복지부
	국내입양아수 및 입양비율	내부자료	보건복지부
	아동 안전사고 현황	사망원인통계	보건복지부

주제별 분류	통계명	출처	제공기관
연금 건강 보험	세대당 건강보험료 부과액	-	보건복지부
	공적연금 수급률	고령자통계 국민연금통계연보 공무원연금통계 사학연금통계연보	보건복지부
	건강보험 재정 및 급여율	진료비 본인부담실태조사	보건복지부
	국민연금 재정현황		보건복지부
	공무원연금기금 증식 및 운용현황	공무원연금기금 결산서	안전행정부
장애인 복지	취업장애인 월평균 소득	장애인실태조사	보건복지부
	장애인 현황	장애인등록현황	보건복지부
	장애인 시설 및 사업현황	보건복지부 통계연보	보건복지부
저출산 고령화	합계출산율	출생통계	통계청
	노년 부양비	장래인구추계	보건복지부

[그림 4-11] 통계청 e-나라지표 복지 지표 제공기관



3. 통계청 보건·복지분야 통계지표

통계청에서 운영하는 국가통계포털(kosis.kr)은 국내·외 통계와 북한 통계를 이용자들이 원스톱으로 이용 할 수 있도록 서비스 하고 있으며 전체 120여개 기관에서 작성하는 국가승인통계를 지표화하여 제공하고 있다. 통계청에서 제공하는 대부분의 통계는 엑셀 파일 형태로 제공하고 있으며 부분적으로 원시데이터를 마이크로데이터 서비스 시스템(mdss.kostat.go.kr)에서 제공하고 있다.

〈표 4-9〉는 통계청 국가통계포털의 보건 분야 제공하는 승인통계와 제공 지표수이며, 〈표 4-10〉은 통계제공기관별 제공지표수 현황표이다. 보건 분야에서 제공하는 통계는 건강보험통계, 결핵현황, 공중위생 관계 업소 실태보고, 국민건강영양조사를 포함하여 총 30종의 국가승인통계를 서비스하고 있으며 2,596개의 지표를 제공하고 있다. 제공기관별로는 보건복지부가 938건, 국민건강보험공단에서 659건, 한국보건산업진흥원에서 353건, 문화체육관광부에서 315건, 건강보험심사평가원에서 116건, 식품의약품안전처에서 76건, 고용노동부에서 61건, 산업통상자원부에서 42건, 국립중앙의료원에서 20건, 통계청 승인통계 자료에서 12건, 교육부에서 4건 순으로 제공하고 있다(그림 4-12 참조).

〈표 4-9〉 통계청 보건분야 통계지표

조사명	지표수
건강보험	793
결핵현황	4
공중위생관계업소 실태보고	3
국민건강영양조사	99
국민구강건강실태조사	116
국민보건의료실태조사	117
국민의료비추계 및 국민보건계정	3

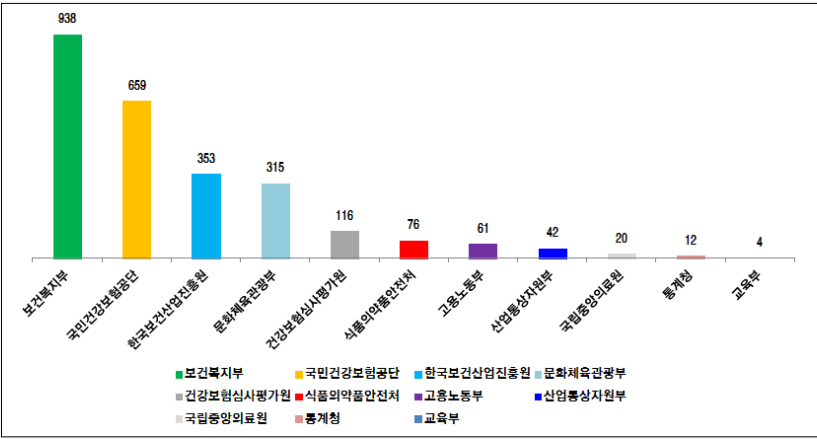
조사명	지표수
법정감염병 발생보고	6
병원경영실태조사	339
보건소 및 보건지소 운영현황	5
사망원인	12
수입식품현황	54
식품수거 검사실적	22
암등록 통계	11
완제의약품 유통정보 통계	34
응급의료현황 통계	20
의료기관별 급여성정성 평가현황	9
의료기기화장품 제조유통 실태조사	26
의약품·의료기기 연구개발 실태조사	14
의약품소비량 및 판매액 통계조사	2
인수공통전염병위험군의 감염실태조사	286
장애인생활체육 실태조사	83
정신질환실태조사	104
지역사회건강조사	35
청소년건강행태 온라인조사	31
체력 및 체격	278
퇴원손상 심층조사	33
한방의료이용 및 한약소비실태조사	30
환자조사	11
HIV/AIDS 신고현황	16
계	2,596

〈표 4-10〉 통계청 보건분야 통계제공기관별 제공지표수 현황

통계기관	제공지표수
건강보험심사평가원	116
고용노동부	61
교육부	4
국립중앙의료원	20
국민건강보험공단	659
문화체육관광부	315
보건복지부	938

통계기관	제공지표수
산업통상자원부	42
식품의약품안전처	76
통계청	12
한국보건산업진흥원	353
계	2,596

[그림 4-12] 통계청 보건분야 통계제공기관별 제공지표 현황



〈표 4-11〉는 통계청 국가통계포털의 복지 분야에서 제공하는 승인통계와 제공 지표수이다. 복지 분야에서 제공하는 통계지표는 가정위탁 국내 입양 소년 소녀 가정현황, 가정폭력실태조사, 가족 친화지수 측정 및 분석, 고령화 연구패널조사, 국민기초생활보장수급자 현황 통계를 포함하여 30종의 승인통계를 제공하고 있으며, 2,842건의 통계지표를 서비스하고 있다. 복지 분야 통계제공기관별 제공지표수 현황은 〈표 4-12〉과 같다. 한국보건사회연구원에서 1,276건, 여성가족부에서 880건, 보건복지부에서 505건, 통계청에서 166건, 고용노동부 8건, 국민연금공단 4건, 국가보훈처 2건, 한국고용정보원 1건을 제공하고 있다(그림 4-13 참조).

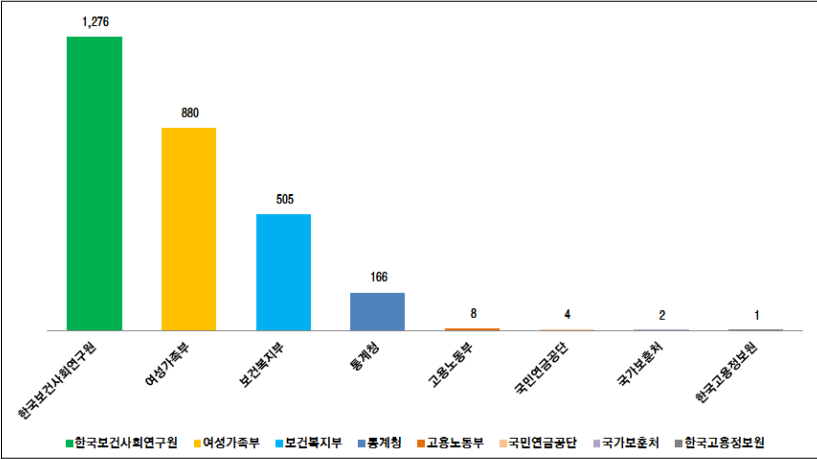
〈표 4-11〉 통계청 복지분야 통계지표

조사명	지표수
가정위탁국내입양소년소녀가정현황	4
가정폭력실태조사	324
가족친화지수측정및분석	190
고령화연구패널조사	1
국민기초생활보장수급자현황	6
국민연금통계	4
노인복지시설현황	2
노인실태조사	80
노인학대현황	57
노후준비실태조사	8
농림어업인복지실태조사	166
보육실태조사	96
보훈보상금지급현황	2
사회복지자원봉사현황	44
산재보험통계	8
성폭력실태조사	107
아동복지시설보호아동및종사자현황보고	1
어린이집및이용자통계	10
여성정책수요조사	88
요보호아동현황보고	-
유망사회서비스수요실태조사	85
장애인실태조사	57
장애인편의시설설치현황조사	5
장애인현황	5
전국출산력및가족보건복지실태조사	213
최저생계비계측조사	16
학대피해아동보호현황	30
한국복지패널조사	1,063
한국의사회복지지출	3
한부모가족실태조사	167
계	2,842

〈표 4-12〉 통계청 복지분야 통계제공기관별 제공지표수 현황

통계기관	제공지표수
고용노동부	8
국가보훈처	2
국민연금공단	4
보건복지부	505
여성가족부	880
통계청	166
한국고용정보원	1
한국보건사회연구원	1,276
계	2,842

〔그림 4-13〕 통계청 복지분야 통계제공기관별 제공지표 현황



〈표 4-13〉는 통계청 국가통계포털의 사회 분야에서 제공하는 승인통계와 제공 지표수이다. 사회분야에서 제공하는 통계지표로는 가족실태조사, 국민 노후 보장 패널조사, 근무 환경 조사, 범죄, 사고관련 통계, 사회조사, 생활시간조사, 신문·잡지 산업 실태조사, 아동종합실태조사를 포함하여 20종의 승인통계를 서비스 하고 있으며 전체 8,121건의 승인통계

를 제공하는 제공기관 현황은 <표 4-14>과 같다. 여성가족부에서 제공하는 통계는 3,395건, 통계청 1,874건, 한국산업안전보건공단 963건, 검찰청 528건, 경찰청 358건, 보건복지부 346건, 한국청소년정책연구원 272건, 한국여성정책연구원 108건, 고용노동부 94건, 한국언론진흥재단 46건, 소방방재청 41건, 국토교통부 23건, 국민연금공단 22건, 한국형사정책연구원 22건, 산림청 19건, 해양경찰청 7건, 기상청에서 3건 순으로 서비스를 제공하고 있다(그림 4-14 참조).

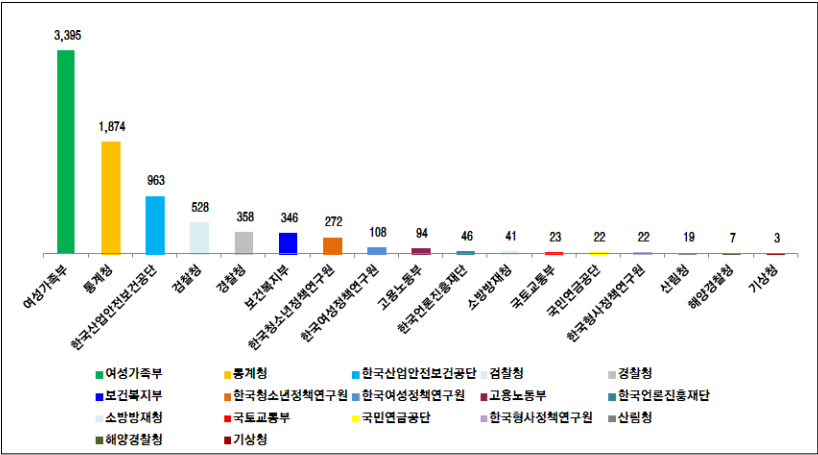
<표 4-13> 통계청 사회분야 통계지표

조사명	지표수
가족실태조사	342
국민노후보장패널조사	22
근로환경조사	222
범죄	893
사고	69
사회조사	1,681
생활시간조사	193
신문·잡지산업실태조사	46
아동종합실태조사	271
아동청소년인권실태조사	272
여성가족패널조사	108
위험물통계	2
재해	872
전국결혼및출산동향조사	75
전국다문화가족실태조사	767
청소년가치관조사	651
청소년매체이용실태조사	84
청소년상담지원현황	4
청소년유해환경접촉종합실태조사	1,442
청소년종합실태조사	105
계	8,121

〈표 4-14〉 통계청 사회분야 통계제공기관별 제공지표수 현황

통계기관	제공지표수
검찰청	528
경찰청	358
고용노동부	94
국민연금공단	22
국토교통부	23
기상청	3
보건복지부	346
산림청	19
소방방재청	41
여성가족부	3,395
통계청	1,874
한국산업안전보건공단	963
한국언론진흥재단	46
한국여성정책연구원	108
한국청소년정책연구원	272
한국형사정책연구원	22
해양경찰청	7
계	8,121

〔그림 4-14〕 통계청 사회분야 통계제공기관별 제공지표 현황



4. 국민건강보험공단 건강보험통계

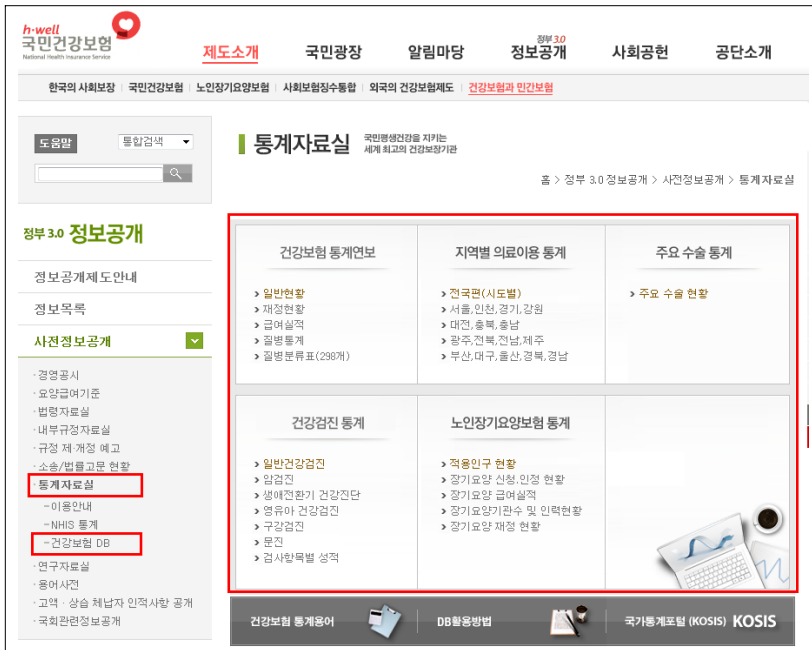
국민건강보험공단(nhis.or.kr)은 국민보건 향상과 사회보장 증진을 위하여 국민의 평생 건강을 지키는 건강보장기관으로서 국민이 질병으로부터 고통 받지 않고 건강한 삶을 영위 할 수 있는 최상의 서비스를 제공하는 기관이다. 국민건강보험공단에서 서비스 하고 있는 건강보험 통계DB 현황은 <표 4-15>과 같다. 건강보험공단에서 국민에게 공개하고 있는 건강보험 통계는 건강보험통계연보, 지역별 의료이용 통계, 주요수술 통계, 건강검진통계, 노인장기요양통계로 총 448개의 통계지표를 서비스하고 있다. 건강보험공단에서 제공하고 있는 통계지표는 국가통계통합DB관리 시스템에서 통계를 작성하여 국민건강보험공단과 국가통계포털에 링크하여 서비스를 지원하고 있다(그림 4-15참조).

<표 4-15> 국민건강보험공단 건강보험 통계DB 현황

통계명	세부분류	지표수
건강보험통계연보	일반현황	213
	재정현황	
	급여실적	
	질병통계	
	질병분류표	
지역별 의료이용 통계		39
주요수술통계		35
건강검진통계	일반건강검진	130
	암검진	
	생애전환기 건강진단	
	영유아 건강검진	
	구강검진	
	문진	
	검사항목별 성적	

통계명	세부분류	지표수
노인장기요양보험통계	적용인구현황	31
	장기요양신청, 인정현황	
	장기요양 급여실적	
	장기요양기관 수 및 인력현황	
	장기요양 재정현황	
계		448

[그림 4-15] 국민건강보험공단 건강보험통계DB 제공화면



5. 건강보험심사평가원 통계

건강보험심사평가원(hira.or.kr) 국민이 의료기관에서 진료를 받았을 경우 의료기관은 총 진료비 중 일부를 환자에게 받고, 나머지는 건강

보험심사평가원에 청구를 하며, 건강보험심사평가원은 청구된 진료비에 대한 심사와 진료가 적정하게 이루어 졌는지에 대한 평가를 하는 기관이다. 건강보험심사평가원에서 제공하고 있는 통계 지표는 국가통계포털에 등록된 통계지표를 링크서비스로 이용자들에게 제공하고 있으며, 제공하고 있는 통계목록으로는 영양기관현황, 건강보험심사실적, 의료기관별 급여 적정성 평가 현황, 완제품 의약품 유통정보 통계로 총 4종을 서비스하고 있으며 116개의 지표를 공개하고 있다(표 4-16 참조).

〈표 4-16〉 건강보험심사평가원 건강보험 통계DB 현황

통계명	지표수
영양기관현황	20
건강보험심사실적	53
의료기관별 급여 적정성평가 현황	9
완제품 의약품 유통정보 통계	34
계	116

6. 식품의약품안전처 통계

식품의약품안전처(mfds.go.kr)는 국민들이 안전하고 건강한 삶을 구현할 수 있도록 식품과 의약품을 비롯한 의료기기, 건강기능식품, 화장품, 생약제제, 바이오 의약품 분야 등에서 식의약 안전관리 망을 확충하고 건강한 환경을 조성하는데 목적을 두고 있다. 식품의약품안전처에서 제공하는 통계DB 현황은 〈표 4-17〉과 같다.

식품, 식품영양, 농축산수산물, 의약품, 바이오, 의료기기, 소비자 위해 정보, 임상통계로 분류하여 각 주제별로 업체현황, 산업현황, 수입현황, 감시현황, 수거 및 검사현황, 기타 식품 현황으로 분류하여 통계지표를 서비스 하고 있으며, 총 153개의 지표를 제공하고 있다.

〈표 4-17〉 식품의약품안전처 통계DB 현황

주제별 분류	세부 분류	지표수
식품	업체현황	5
	산업현황	11
	수입현황	10
	감시현황	6
	수거 및 검사현황	3
	기타 식품분야	6
식품영양	업체현황	1
	산업현황	1
	수입현황	2
	감시현황	4
	어린이 식생활 안전	1
농축수산물	업체현황	1
	산업현황	1
	수출입현황	2
	수거검사현황	4
	감시현황	5
	기타	3
의약품	업체현황	3
	산업현황	6
	수출입현황	6
	허가등현황	6
	시판 후 조사현황	6
	수거검사실적	2
	약시감시 등	3
	검사기관 지정현황	2
	마약류 현황	6
	실험동물 관련현황	3
바이오	업체현황	1
	산업현황	1
	수출입현황	5
	감시현황	4
	허가 및 신고현황	6
	수거검사실적	3

주제별 분류	세부 분류	지표수
	시한 후 조사현황	2
의료기기	업체현황	1
	산업현황	8
	수출입현황	-
	감시현황	5
	허가 및 신고현황	8
소비자위해정보	국내외위해정보수집	-
임상통계	임상통계	-
계		153

7. 국가암정보센터

국가암정보센터(cancer.go.kr)에서 제공하고 있는 통계지표는 보건복지부에서 중앙암등록본부 및 지역암등록본부를 운영하여 우리나라에서 진행되는 모든 암등록 사업 자료를 통합하여 국가 암발생 데이터베이스를 구축하였으며, 등록받은 암환자 자료를 기반으로 국민건강보험공단의 암 진료 환자, 중증환자등록자료 및 통계청의 암사망자료 중 등록되지 않는 암 사망자를 포함하여 암 발생 의심자에 대한 의무조사를 수행하여 통계를 포털에서 제공하고 있으며, 암에 관한 발생률, 사망률, 생존율, 유병율로 구분하여 총 13개 통계지표를 제공하고 있으며 통계DB 현황은 <표 4-18>과 같다.

<표 4-18> 국가암 정보센터 통계DB 현황

주제별 분류	지표명	지표수
발생률	암발생률	5
	암종별 발생현황	
	연령군 별 암발생률	
	암 발생률 추세분석	
	암 발생 국제비교	

주제별 분류	지표명	지표수
사망률	주요암 사망분율	2
	성별 주요암 사망분율	
생존율	암 생존율	4
	주요암 5년 생존율	
	요약병기별 생존율	
	암 생존율 국제비교	
유병률	암 유병률	2
	암종별 암유병률	
계		13

국가암정보센터에서 제공하고 있는 통계표는 국가통계포털(kosta.go.kr)또는 통계청 홈페이지에 접속하여 사망원인통계 및 사망원인통계연보를 통하여 확인할 수 있다. [그림 4-16]은 국가암정보센터 통계지표 제공화면이다.

[그림 4-16] 국가암정보센터 통계지표 제공화면



8. 한국보건사회연구원 보건복지 통계

한국보건사회연구원 보건복지통계포털(hawelsis.kihasa.re.kr)은 유용성 높은 보건복지 통계를 수집, 생산, 제공을 통하여 국가 보건정책의 수립에 기여하고 있다. 한국보건사회연구원 보건복지통계포털에서는 국가 보건 복지 통계뿐만 아니라 국외 보건복지 및 북한 보건복지 통계도 수집하여 제공하고 있다. 보건복지통계포털에서는 주제별, 생애주기별, 명칭 별로 구분하여 일반 이용자에게 서비스하고 있다. 생애주기별 분류로는 영유아부터 노인에 이르기까지 관련 통계를 분류하여 제공한다.

〈표 4-19〉는 보건 분야 통계분류와 서비스 하고 있는 지표수 현황이다. 보건 분야 분류로는 총 5가지로 분류하고 있으며 국민건강, 보건자원, 보건산업, 보건일반, 보건의료비용으로 분류하고 있다. 국민건강에는 국민영양, 질병, 보건의식행태로 분류하고 있으며, 보건자원은 의료인력, 의료시설, 의료장비, 보건산업은 의약품 산업, 의료기기산업, 화장품산업, 식품사업으로 분류하여 관련 지표를 수집, 생산하여 공개하고 있다. 보건분야 관련 등록 지표는 총 176개의 지표를 서비스 하고 있다.

〈표 4-19〉 한국보건사회연구원 보건복지통계 보건분야 통계지표

주제별 분류	세부 분류	지표수
국민건강	국민영양	-
	질병	42
	보건의식행태	12
보건자원	의료인력	7
	의료시설	1
	의료장비	3
보건산업	의약품산업	84
	의료기기산업	
	화장품산업	15
	식품산업	1

주제별 분류	세부 분류	지표수
보건일반		-
보건의료비용		11
계		176

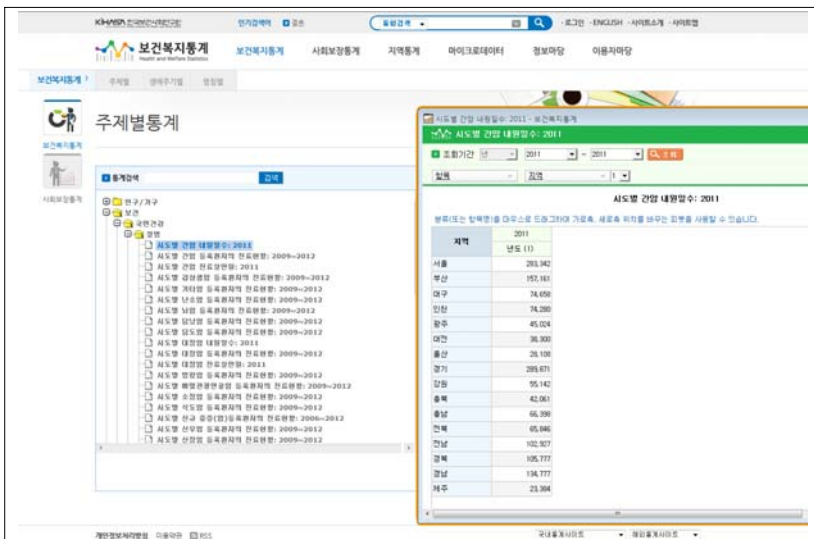
〈표 4-20〉은 복지 분야 통계분류와 서비스 하고 있는 지표수 현황이다. 복지 분야 통계 분류는 복지자원, 보육 및 아동, 사회보험, 복지일반, 청소년, 노인, 여성가족, 장애인, 기초생활보장(공공부조), 복지일반으로 구분하여 제공하고 있다. 복지자원에는 복지시설, 복지인력으로 카테고리화 하여 지표를 공개하고 있으며, 사회보험에는 건강보험, 국민연금, 공공연금, 기초노령연금, 노인장기요양보험, 산재보험, 고용보험으로 구분하여 지표를 서비스 하고 있다. 기초생활보장(공공부조)는 기초생활보장, 의료급여, 자활로 구분하여 지표를 제공하고 있다.

〈표 4-20〉 한국보건사회연구원 보건복지통계 복지분야 통계지표

주제별 분류	세부 분류	지표수
복지자원	복지시설	9
	복지인력	3
보육/아동		15
사회보험	건강보험	28
	국민연금	-
	공공연금	-
	기초노령연금	-
	노인장기요양보험	-
	산재보험	-
	고용보험	-
복지일반		-
청소년		10
노인		-
여성/가족		548
장애인		7

주제별 분류	세부 분류	지표수
기초생활보장 (공공부조)	기초생활보장	45
	의료급여	42
	자활	105
복지일반		188
계		1,000

[그림 4-17] 한국보건사회연구원 통계지표 제공화면





제5장

보건복지 빅데이터 수요공급 예측

제1절 빅데이터 수용요인 분석

제2절 공공과 보건의료 빅데이터 수요공급 예측

제3절 보건복지 정책 수요예측

제1절 빅데이터 수용요인 분석¹²⁾

1. 서론

빅데이터 시대에는 다양한 스마트 디바이스를 통해 생산되는 이질적인 데이터들을 처리하기 위한 혁신적인 기술들이 개발, 적용되고 있으며 이러한 데이터와 기술은 사회적 패러다임을 변화시키는 것이 가능하게 되었다. 국내외를 막론하고 이러한 빅데이터에 대하여 학계, 산업계, 공공의 관심이 크지만 아직까지 빅데이터 기술수용에 대하여 체계적으로 연구된 사례는 찾기 어렵다. 따라서, 본 연구는 국내 초기 시장의 빅데이터 기술수용 연구를 위해 기술수용모델(TAM)을 중심으로 하여 혁신확산이론(Innovation Diffusion Theory) 및 과업기술적합성(Task Technology Fit)이론을 통합적으로 활용하였다. 그리고 현재 빅데이터 기술의 수용자가 기업이 주체가 됨을 반영하여 기업 입장에서 빅데이터 기술 수용의 목적성을 기술수용모델의 조절변인으로 확장하였다. 그동안 대부분의 선행연구의 조절변인은 인구통계학적이거나 사회적 관점에서 적용되었지만 본 연구에서는 기술수용목적에 따른 조절변인을 반영하여 빅데이터 기술수용의 특성을 이해하고자 하였다. 이와 같은 접근이 필요한 이유는 빅데이터 기술 시장이 가지는 복합적인 성격에 기인한다.

12) 본 절의 내용은 본 연구의 일환으로 수행된 '김정선-송태민(2014). 빅데이터 기술수용 초기특성 연구-기술이용자 및 기술활용자 측면의 조절효과를 중심으로. 한국콘텐츠학회지, 14(9), 538-555'에서 발췌한 것임을 밝힌다.

초기 빅데이터 시장에서 기술의 수용자는 단순 기술 이용자이기도 하지만 빅데이터 기술을 기반으로 2차적 제품과 서비스를 개발, 제공하려는 기술활용자적 면모를 가질 수 있기 때문이다. 본 연구의 사용 용어인 ‘기술이용자’란 주어진 업무의 처리 및 효율화 목적으로 빅데이터 기술을 도입하는 수용자를 의미하며 ‘기술활용자’란 빅데이터 기반으로 새로운 제품이나 서비스를 기획, 개발하기 위한 수용자를 의미한다. 만약 기업 내의 ‘기술활용자’가 신제품 및 서비스 개발에 성공하여 시장에 활발하게 빅데이터 기술 기반의 상품을 제공하는 경우에는 빅데이터라는 혁신기술의 확산이 성공적으로 이루어질 것이다. 그러므로 ‘기술활용자’의 빅데이터 기술수용도는 혁신기술 확산의 선행지표로 활용 가능할 것이다.

기술에 대한 수용의 주체가 개인이던 기업이던지 간에 새로운 기술수용 과정에서 중요한 것은 이용자의 선택을 좌우하게 되는 변인이며, 이용자의 선택에 따라 기술의 확산여부가 결정되므로 기술 수용자들의 요구를 정확히 진단할 필요가 있다. 이에 본 연구에서는 빅데이터를 대상으로 수용자의 빅데이터 기술 선택과정을 설명하기 위하여 기술수용모델을 중심으로 기본적 틀을 구성하고, 혁신확산이론 내에서의 주요변인인 조직의 혁신성과 기업이 기술을 채택 시 중요 역할을 수행하는 과업기술적 합성이론을 활용하여 연구모형을 구성하였다. 설정한 외생 변인이 기술수용모델의 주요 변인들을 거쳐 실제 이용의도에 어떠한 영향을 미치는지 분석하고자 하였으며 기업의 기술수용 목적에 따른 조절변인 반영과 기술에 대한 지속적 사용의도를 종속변인으로 하여 기존 연구의 확장을 시도하였다.

또한 도출된 연구결과를 통해 기업에서 빅데이터 기술을 수용하는 과정에서 발생하게 되는 목적에 따른 차이점을 분석하고자 하였으며 이에 따른 기업의 빅데이터 기술 도입 시 고려해야 할 사항을 제안하고자 하였다.

2. 이론적 배경

가. 기술수용모델

기술수용모델(TAM)은 혁신기술의 하나인 컴퓨터 수용에 대한 사용자들의 행동을 설명하기 위해 Davis가 1986년 처음 개발하였고 1989년 이를 공식화한 모델이다. 이 모델은 Ajzen·Fishbein이 정립한 합리적 행동이론(Theory of Reasoned Action : TRA)에 근거한 것으로 TRA가 인간의 일반적인 행동을 설명하고자 한 것이라면 TAM은 정보기술 즉 컴퓨터와 같은 혁신기술의 수용행동을 설명하려는 것이다. 정보기술수용의 대표적인 연구모형 가운데 하나인 기술수용모델에서는 이성적 행동이론의 기본 틀에 따라 해당 정보시스템이 제공할 수 있는 혜택과 사용방식에 대한 이성적 판단인 유용성(utility)과 사용용이성(ease of use)을 근거로 수용 여부를 판단하는 것을 가정하고 있다. 합리적 행동이론이 기술수용모델의 기반이 되는 이유는 새로운 기술을 수용하는 행동에 대한 결정을 합리적인 기대가치(expectancy-value) 체계에 의존한다는 점 때문이다(Shah·Higgins, 1997). 인간이 자신의 행동에 대한 결과를 생각하고 그에 대한 합리적인 판단을 가지고 행동한다는 합리적 행동이론을 확장한 것이 계획행동이론이며, 합리적 행동이론을 새로운 기술수용 상황 하에 적용하여 수정한 것이 기술수용모델(TAM)이다(김광재, 2007).

기술수용모델(TAM)은 모델 자체가 가진 복잡성과 연계 용이성으로 기술수용 시 적용 가능한 다양한 영향요인들에 대해 통합적 접근이 용이한 장점을 지닌다. 물론 최근에는 지나친 변형과 확장이 이루어진다는 비판적 평가도 없지 않으나 기술의 특성과 사회적 환경이 급변하고 있는 현실

세계를 반영할 때 기본적인 TAM의 구조 안에서 다양한 이론을 활용하는 접근법에 대한 필요성은 크다. 최근에는 기술수용모델의 구성변인인 인지된 유용성과 용이성만으로 새로운 기술의 채택과정을 설명하는 데는 한계가 있다는 비판이 제기되었다. TAM의 기본적인 구성변인이 적용 가능한 정보시스템의 다양한 환경을 완전하게 반영하고 있지 못하다는 한계점이 존재하므로 이러한 문제를 극복하기 위해 TAM을 변형하고 다양한 형태로 확장함으로써 보다 정교화된 TAM모형이 제시되고 있다. 따라서 계획행동이론을 구성하는 변인인 주관적 규범을 추가시킨 확장된 기술수용모델인 TAM2, TAM3가 제안되기도 하였다. TAM3는 Venkatesh·Bala(2008)에 의해서 기존 연구들의 귀납적 통합을 위해 고안되었다. 기술수용모델(TAM)이 가지는 한계점으로 지적되는 것은 TAM은 개인이 정보기술을 수용하는 과정에서 조직이나 주변 환경 및 이해 관계자가 가질 수 있는 영향요인을 간과하고 있다는 점이다. 앞에서도 살펴본 인지된 유용성과 용이성만으로 새로운 기술의 채택과정을 설명하기에는 한계가 있다는 지적은 최근처럼 다양한 신기술 기반의 제품과 서비스가 시장에 출시되고 있는 상황에서 적합한 지적이다. 특히 빅데이터 기술과 같이 초기 시장의 수용자가 일반적 개인이 아니며, 이용자 측면에서 가시성(visibility)이 확보되기 어려우며, 그 목적성에 있어서도 다양한 가치로 전달되는 경우에는 적합한 외생변인을 설정하여 연구가 진행되어야 할 필요가 있다. 기존 연구들에서도 새로운 기술에 TAM을 적용할 경우 기술의 특성을 반영하여 모형을 설계해야 한다는 의견은 많은 학자들에 의해 지지되고 있다(백상용 2009; 유재현·박철, 2010). TAM에서 사용된 인지된 유용성과 인지된 용이성이 정보기술 수용의도에 미치는 영향연구는 대부분의 선행연구에서 정(+)의 관계가 검증되었다(Davis, 1989; Igbaria et al., 1997; Venkatesh·Davis, 2000). 그러

나 TAM을 기반으로 하는 많은 연구들이 개인의 동기적 요인 외에도 사회적 동기 혹은 다양한 형태의 기술수용에 영향을 미치는 보다 동태적인 요인들에 대한 고려를 통해 확장되고 있는 만큼 새롭게 등장하는 정보기술 환경을 제대로 반영하고 그 유형에 따라 TAM을 적용하기 위한 통합적인 모델 도출 노력이 필요하다. 기술수용모델이 적용된 신기술 수용 연구는 <표 5-1>과 같다.

<표 5-1> 기술수용모델 국내 선행연구

연구자	연구내용
김민규 외 (2007)	잠재적 사용자의 혁신성이 DMB 수용의도에 영향을 미치는 구조방정식을 활용하여 검증
강소라 외 (2007)	과업기술적합도(TTF)가 그룹지원시스템(GSS)의 사용에 미치는 성과와 영향에 대하여 전유방식 동의정도 및 전유 충실도 측면의 조절효과를 검증
김광재 (2007)	DMB 수용의도에 영향을 미치는 이용 용이성과 유용성에 대해 유의한 관계를 구조방정식을 활용하여 검증
김규동 (2008)	와이브로 서비스 채택요인에 대해 정보기술수용모형 및 개혁확산이론을 통해 통합적 모형을 구축하고 자기 효능감 및 와이브로 채택 태도, 인지된 비용 등이 유의한 관계를 가짐을 검증
이현미 (2008)	웨어러블 컴퓨터에 기술수용모델을 적용하여 유의한 결과 도출을 통해 모델의 적용가능성 확인
임형주 (2008)	EMR시스템에 대한 환자들의 수용의도를 인지된 보안, 인지된 편의성이 높을수록 영향관계를 가짐을 검증
오태동 (2009)	U-서비스 이용자들을 대상으로 기술수용모델을 적용, 인지된 가치, 인지된 위험, 인지된 신뢰, 자율성을 영향으로 형성된 태도가 부(-)적 관계임을 검증
김미선 (2010)	IPTV에 대한 지각된 용이성, 유용성 외 IPTV만이 가능한 정보성(다양한 콘텐츠 제공) 관련 이용자들에게 정보 품질로 지각됨을 검증
변숙은 (2010)	하이테크 제품으로서의 지문인식 ATM기기 수용 시 그 혜택과 위험요인에 따른 수용자를 군집분석을 통해 혜택 비공감형, 위험 민감형, 혁신 추구형, 3개 그룹으로 세분화하고 사후검정을 통해 그 영향요인 차이를 검증
최만수 (2011)	스마트폰 수용의도에 영향을 미치는 요인을 개인의 혁신성, 사회적 영향력, 사용자 인터페이스 요인으로 나누어 검증
김성일 (2011)	스마트폰 미사용자들을 대상으로 개인의 혁신성에 따라 다르게 나타나는 수용의도의 차이 분석
송태민 외 (2012)	U-health시스템에 대한 수용에 영향을 주는 서비스 품질 및 수용태도에 대한 척도가 수요자와 공급자에게 공통적으로 적용될 수 있는 척도인지를 교차타당성 분석을 통해 검증

연구자	연구내용
서정환 (2013)	기업용 클라우드 서비스 도입에 대한 결정요인 연구를 통해 인지된 혜택과 위험이 클라우드 서비스 도입에 미치는 영향을 연구함
김은영, 이정훈, 서동욱 (2013)	빅데이터 기술수용의도에 영향을 미치는 요인을 수용조직의 환경요인 및 빅데이터 시스템 특성요인으로 구분 후 조직환경 요인으로 조직의 혁신성, 조직의 여유자원, 정보시스템 성숙도를, 빅데이터 기술의 사용 잇점을 빅데이터 기술특성으로 적용하여 검증

선행연구들은 대부분 개인으로서의 기술수용의도를 연구하였다. 이는 대부분 완성형 제품 및 서비스로 제공되는 하이테크 신제품 또는 기업의 정보시스템 측면에서 연구주제가 다루어지고 있다. 그런데 이러한 연구들은 기업의 새로운 IT기술 도입과정에서의 기술수용의 목적에 따른 연구모형 적용 측면을 고려하지 못하고 있다. 기업의 ERP 시스템이나 모바일 오피스의 기술수용과는 다르게 빅데이터 기술은 H/W를 제외하고는 완제품이나 서비스의 형태로 제공되는 경우가 드물고 업무 및 적용단위의 수요 자체가 개별적이기 때문에 그 수용목적의 차별성이 있다. 최근 김은영 등(2013)의 빅데이터 기술수용 연구에서는 단지 조직 환경적 요인과 빅데이터 특성 요인을 인지된 잇점으로 구분하여 모델을 구성하였으나 여기에서도 기술을 수용하는 주체에 대한 이해 및 기술 수용 목적성에 대한 부분이 반영되지 못하였다. 따라서 본 연구에서는 기업의 빅데이터 기술도입에 따른 기업의 개별 목적성을 반영하여 단순수용의 목적을 가진 ‘기술이용자’ 측면과 기술을 이용하여 새로운 제품(서비스) 창출을 목적으로 하는 ‘기술활용자’ 측면의 이원화된 관점으로 TAM기반 연구적용을 확장하고자 한다.

나. 조직의 혁신성과 과업기술적합성

1) 조직혁신성

혁신적인 기술을 수용하기 위해서는 창의적이고 혁신적인 개인과 조직 문화가 필수적이다. 조직의 혁신능력은 조직의 자원과 신기술을 성공적으로 활용하게 하는 선행요소이다(박상준·유수천, 2009). 피터 드러커(Drucker P. F, 2007)는 혁신을 기업의 사회적, 경제적 잠재력에 변화를 가져오기 위하여 분명한 목표를 설정하고 집중적인 노력을 펴는 것으로 정의하고 있다. 혁신기술 관련 선행연구를 통해 살펴본 주요 영향요인은 조직의 혁신역량이었다. Lawson·Samson(2001)은 혁신역량을 높은 수준의 통합능력이라고 정의하고, 이런 혁신역량을 보유한 조직은 주요 역량과 자원을 통합할 수 있다고 하였다. Guan·MA(2006)는 혁신역량을 학습역량, R&D역량, 생산역량, 자원탐색역량, 마케팅역량, 조직역량, 전략역량으로 설명하였다. 학습역량은 지식을 발견해내어 흡수하고 이용하는 능력이고, R&D역량은 새로운 제품을 개발하는 역량이고, 생산역량은 연구개발 성과를 시장의 니즈에 맞게 제품으로 만들어내는 역량을, 자원탐색역량은 기술, 인적자원, 재무자원을 획득하는 역량을 의미한다. 마케팅역량은 소비자의 현재, 미래의 니즈를 이해하여 제품을 제공하고 홍보하는 역량이며, 조직역량은 조직문화를 육성하고 훌륭한 관리기술을 습득하여 조직체제와 조화를 이루어내는 역량을 의미한다. 전략역량은 환경변화 속에서 경쟁력을 유지하기 위해 서로 다른 전략에 적응하는 전략을 의미한다. Tornatzky et al.(1990)과 Zhu et al.(2006)은 혁신확산에는 기술특성(technological context), 조직특성(organizational context) 및 환경특성(enviromental context)이 영향을 준다고 하였

으며 Rubenstein et al.(1976)은 최고 경영층의 관심과 지원, 담당자의 열의 및 추진력, 조직 및 구성원의 변화에 대한 태도가 기술혁신의 성패 요인으로 작용한다고 주장했다. 한편 조직 문화적 연구관점에서는 조직의 혁신성향은 구성원의 혁신성, 리더의 혁신성, 조직적 지원인식 등으로 구성되며 개인이 기존 업무방식이나 절차에 대해 스스로 개선하려는 의지와 구체적인 노력의 정도를 의미한다(Scott·Bruce, 1994).

조직은 환경과 끊임없이 상호작용을 하는 개방 시스템이며, 따라서 혁신확산 과정을 규명하기 위해서는 이에 영향을 미치는 환경요인을 충분히 검토해야 한다. 혁신적인 조직은 외부환경으로부터 혁신에 대한 아이디어나 시장 및 고객정보 등을 수집하여 조직 내에서 적절한 반응을 하는 조직이라고 할 수 있으며 기존의 여러 연구 결과에서 조직의 상 황요인이 혁신의 확산 과정에 영향을 준다고 보고하고 있다. Rogers(1983)는 조직의 혁신 수용과정을 설명하는 변수로서 개인, 조직구조, 조직외부의 특성을 제시하고 있으며, Kwon·Zmud(1987)는 개인, 과업, 혁신, 조직구조, 환경 요인을 제시하고 있다. 일반적으로 혁신 수용자 개인의 변화에 대한 긍정적 태도, 조직의 여유자원, 조직의 규모, 조직시스템의 개방성 등이 혁신 수용과정에서 긍정적인 영향을 주며, 조직구조의 집권화, 공식화 등은 부정적인 영향을 주는 것으로 알려져 있다(Grover·Goslar, 1993; Kwon·Zmud; 1987). 기업의 IT혁신은 조직의 경영 프로세스와 생산제품에 직·간접적으로 영향을 미침으로써 조직의 의사결정에 큰 영향력을 미친다(장활식 외, 2008). 기술혁신도가 높은 기업에서는 그렇지 않은 기업보다 강력한 혁신성향 및 시장지배력의 증대로 기술혁신의 중요성이 점차 커지고 있다(Stratopoulos·Lim, 2010). 그러므로 빅데이터 기술을 수용하는 조직은 반드시 ‘조직의 혁신성’이 전제되어야 할 것이다.

2) 과업기술적합성

기술수용모델과 과업기술적합성이론은 기술수용과 이용을 다룬다는 측면에서 중복되는 부분이 많다. 그러나 기술수용모델은 기술이용에 대한 의도에 보다 초점이 맞춰지며 과업기술적합성이론은 그 결과에 보다 초점이 맞춰진다. 두 이론 모두 이용자들의 선택과 평가를 기반으로 개발되었는데 이것이 통합될 때에는 기술수용에 있어서 보다 강력한 모델이 될 것이다. Goodhue·Tompson(1995)은 새로운 기술이 얼마나 특정 과업의 필요조건에 적합한 지에 따라 과업에 적절한 적합성이 존재한다면 그 기술은 적용될 수 있을 것이라고 했다. 그들이 제시한 결합모형에 따르면 과업과 기술 사이의 적합성을 측정하는 것을 과업기술적합성이라고 부른다. 이러한 과업기술적합성은 사용과 생산성의 선행 변수에 영향을 준다.

IT기술은 개인의 차원을 넘어 조직의 구성원들이 업무를 처리하기 위해 사용하는 수단으로 발전되었으며 지속적으로 IT기술과 인프라 및 비즈니스 지원의 관계가 더욱 강화되어 가고 있다. 과업기술적합성모델은 이러한 조직 내의 과업(Task)을 지원하는 정보기술 역량을 설명하는 이론으로 과업기술적합도(Task Technology Fit)란 개인이 과업을 수행할 때 필요한 기능을 IT가 얼마나 잘 지원해 주는가의 정도로 정의할 수 있다(Goodhue·Tompson, 1995). 과업기술적합도에서 기술이란 목표지향적인 개인의 과업 실행을 지원하는 수단으로 간주된다. 과업과 기술이 적합하다는 것은 과업 수행에 있어 기술이 과업 수행에 적절한 정보와 기능을 제공해 줌으로써 개인의 과업수행에 최적의 의사결정이 이루어질 수 있도록 지원하고 성과를 내도록 하는 것이다.

Goodhue(1995)는 과업에 필요한 정보와 이를 지원해주는 기술과의

적합성을 시스템 가치사슬이라고 정의하고 적합성에 대한 사용자의 태도가 성과에 영향을 미친다고 했다. 다양성, 어려움, 상호의존성의 관점에서 과업특성은 기술을 사용하는 면에서 개인의 신뢰성과도 연관될 수 있는데, 이것은 기술이 업무를 잘 실행하도록 돕는다는 것을 안다면 그 기술을 유용하고 중요한 것으로 받아들일 것이라는 것이다. 또한 과업기술 적합성을 강제적이고 자발적인 환경에서 사용하는 것이 적절하다는 의견을 밝히기도 했다(Goodhue·March, 2000). Dishow·Strong(1999)은 기술수용모델(TAM)과 과업기술적합성(TTF)의 통합모델을 고안했는데 이들의 연구결과에서도 다양한 회사의 경영정보 시스템에 대한 TAM과 TTF의 개별적 모델 적용보다 두 개 이론을 통합한 모델이 보다 설명력이 높은 결과를 보였다. 기술초기 시장은 혁신기술 수용자들이 기술을 수용한다. 개인으로서의 혁신기술 수용자의 성향은 위험을 감수하고 혁신적이며 혁신기술 그 자체를 선호할 뿐만 아니라 자신의 판단에 따라 작은 가능성에도 도전하는 이들이다(김상훈, 2008). 그러나 기업은 조직으로서 개인으로서의 기술 수용과는 다른 의사결정 체계를 가진다(이장우·장수덕, 2010; 조항정 외, 2009). 기술의 채택과정은 조직의 성과와 효율성에 기여하기 위한 의도된 활동이므로 조직이 혁신을 수용하는 목적과 태도가 중요하다. 그러므로 기술의 수용 주체가 기업인 경우 혁신을 장려하는 조직문화가 선행되어야 하며 이러한 조직문화는 조직원들의 태도와 행동을 결정하는 중요한 가이드라인을 제공한다(이대기, 2012).

창의성과 혁신행동 간의 선형관계를 규명한 연구결과들(Amabile, 1988; 이문선·강영순, 2003)은 기업이 혁신 기술기반으로 신제품 및 서비스를 생산하고 기획하는 것은 기업의 경쟁우위 획득 및 유지를 위하여 매우 중요하며 이를 위한 선행조건인 조직혁신성이 중요하다는 것을 밝히고 있으므로 빅데이터 기술 초기 시장 내에서도 기업이 혁신기술 확산을

지속적으로 이어나가기 위해서는 조직의 혁신성이 매우 중요할 것이다.

한편 Goodhue·Thompson(1995)은 기술의 과업기술적합성이 높을 때에만 많은 이용을 통해 기업의 보다 나은 생산성을 가져온다고 주장한다. 어떠한 종류의 기술이라도 과제의 요건이나 개인의 능력에 따라 생산성에 다른 결과를 가져오며(Goodhue et al., 2000) 과업과 기술의 적합도는 생산성과 기술 이용도에 영향을 주게 된다(Larsen et al., 2009). 또한 업무수행을 위한 특정 기술은 어느 부서에는 필요하지만 또 다른 부서에는 전혀 필요하지 않은 것일 수 있다(Goodhue et al., 2000). 선행연구의 결과에서 사용자의 기술에 대한 경험을 통한 과업기술적합성은 조직의 생산성을 예측할 수 있게 하므로(Goodhue·Thompson, 1995; Becerra-Fernandez·Sabherwal, 2001) 빅데이터 기술 역시 해당 기술의 초기 수용시장 내에서 어떠한 과업영역의 목적성을 가지고 활용되고 있는지 살펴보는 것도 필요하다. 앞에서 살펴본 다양한 선행연구를 통해 보고된 것처럼 혁신기술의 기업 내 수용과정에서는 ‘조직의 혁신성’과 ‘과업기술적합성’이 전제되어야 할 것으로 판단된다. 그러므로 본 연구에서는 빅데이터 기술의 특성을 고려한 기업의 기술수용 목적성의 차별점과 특성을 고려하여 기술수용모델의 외생변인으로 ‘조직의 혁신성’과 ‘과업기술적합성’을 적용하였다.

3. 연구의 방법 및 설계

연구모델 개발에 사용된 도구로는 <표 5-2>와 같이 문헌연구를 통해 검토한 기술수용모델의 이론적인 틀과 일반적인 IT 적용기술과 관련된 평가에 대한 기존 연구결과들을 토대로 적용하였다. 이를 통해 기술수용모델의 외생변인은 조직의 혁신성, 주관적 규범, 과업기술적합성으로 설정하고 TAM의 기본 모형인 인지된 용이성, 인지된 유용성, 태도, 수용의

도를 내생변인으로 적용한 후 최종 종속변인으로는 빅데이터 기술에 대한 지속적 수용의도를 설정하였다.

〈표 5-2〉 연구모델의 개발도구

변수	요인	항목	기존연구
TAM과 관련된 독립변수	주관적 규범 (Subjective norm)	4	Davis(1989), Taylor & Todd(1995), Venkatesh & Davis(2000), An J. Y. et al.(2007)
과업기술적합성 (TTF) 이론 독립변수	과업기술적합성 (Task-Technology Fit)	4	Goodhue(1995), Folz-Murphy et al.(1998), Zigurs(2000), Dishaw & Strong(1999), DeLone & McLean(1992)
혁신확산이론 (IDF) 독립변수	조직의 혁신성 (innovativeness)	8	Moore & Benbasat(1991), Adams et al. (1992), Grover & Goslar(1993), Scott & Bruce(1994), Agarwal & Prasad(1997), 오창규(2002), Rogers (2004), 박은아(2007), 김해룡, 양필석(2008), 김성개(2009)
TAM 변수	인지된 용이성 (Perceived ease of use)	5	Adams & Todd(1992), Chen et al.(2002), Davis(1986, 1989), Henderickson & Cronan (1993), Mathieson(1991), Moon & Kim(2001), Moore & Benbasat(1991), Straub & Brenner (1997), Szajna(1994), Moon & Kim(2001), Kim & Lee(2002)
	인지된 유용성 (Perceived usefulness)	5	Adams & Todd(1992), Davis(1986, 1989), Henderickson & Cronan(1993), Mathieson(1991), Moon & Kim(2001), Moore & Benbasat (1991), Straub & Brenner (1997)
	사용에 대한 태도 (Attitude to use)	5	Henderickson & Cronan(1993), Straub & Brenner(1995), Adams & Todd(1992), Taylor and Todd(1995)
	사용에 대한 행동의도 (Behavioral intention to use)	4	Davis.(1989), Venkatesh & Davis(2000), Jackson & Leitch(1997), Subramanian(1994), Mathieson(1991), Venkatesh & Morris(2000), Moon & Kim(2001), Szajna(1994), Taylor & Todd (1995), Moon & Kim(2001), Kim & Lee(2002)
종속변수 (variable)	빅데이터 지속수용의도 (Big Data loyalty)	2	Kim & Lee(2002) , Yoo & Donthu(2001)

가. 주요 개념의 정의와 측정방법

1) 조직의 혁신성향

Rubenstein et al.(1976)은 최고 경영층의 관심과 지원, 담당자의 열의 및 추진력, 조직 및 구성원의 변화에 대한 태도가 기술혁신의 성패요인으로 작용한다고 주장하였다. 한편 조직 문화적 연구관점에서는 조직의 혁신성향은 구성원의 혁신성, 리더의 혁신성, 조직적 지원인식 등으로 구성되며 개인이 기존 업무방식이나 절차에 대해 스스로 개선하려는 의지와 구체적인 노력의 정도를 의미한다고 하였다(Scott·Bruce, 1994).

본 연구에서는 Lim·Yun(1999)의 연구에서 사용된 혁신성향 측정도구를 활용하여 ‘개인 및 리더 그리고 조직차원에서 새로운 것을 받아들이고 이를 통해 기존 업무를 개선하려는 의욕과 노력수준’으로 측정하였다. 조직의 혁신성향을 측정하는 도구로 개인과 리더, 집단수준의 혁신성향을 측정하여 점수가 높을수록 조직의 혁신성향이 높은 것으로 보았다. 문항은 총 7개로 구성하고, 7점 척도로 측정하였으며, Cronbach's alpha .909로 나타났다.

2) 주관적 규범

주관적 규범은 계획행동이론에 근거한 변인으로 확장된 기술수용모델에 추가로 적용되었다. Fishbein·Ajzen(1975)은 주관적 규범을 ‘의문에 대한 행동 수행여부를 생각하는 사람의 지각’으로 정의하였다. 본 연구에서의 주관적 규범은 빅데이터 기술을 도입하여 이용하는 것이 이용하지 않는 사람보다 사회적으로 상대적인 우월감을 가지는 정도로 정의하고

‘빅데이터 기술을 이용해야 하는 것에 대한 개인적인 필요의 지각 정도’를 측정하였다. 문항은 총 4개로 구성하였으며, 7점 척도로 측정하였으며, Cronbach’s alpha .864로 나타났다.

3) 과업기술적합성

과업기술적합성(TTF)이란 기술의 기능과 과업의 특성, 그리고 개인의 능력이 얼마나 잘 결합되었는가를 의미하는데, Goodhue(1998)는 TTF가 높을 때 성과가 높다고 하였다. 빅데이터 기술이 과업 수행을 얼마나 잘 지원해 주는가의 여부가 조직 및 개인의 성과와 밀접한 관계가 있기 때문에 과업기술적합성을 기술수용모델의 외생변인으로 채택하였다. 이는 빅데이터 기술이 반영된 정보기술시스템이 사용자의 과업에 대한 필요성을 얼마나 충족시키는가에 대한 만족도를 측정하는 것으로 DeLone·McLean(2003)의 연구에서 사용되었다. 본 연구에서의 과업기술적합성은 ‘빅데이터 기술에 대한 만족 정도’로 측정하였다. 문항은 총 4개로 구성하였으며, 7점 척도로 측정하였으며, Cronbach’s alpha .888로 나타났다.

4) 인지된 유용성

인지된 유용성(utility)은 혁신기술이나 제품이 이용자에게 전달해 주는 가치가 기존의 것보다 우수하다고 이용자가 지각하는 정도를 말한다. Davis(1989)는 ‘인지된 유용성이란 정보시스템이 조직 내 개인이 수행하는 업무과정과 목표에 기여하는 정도’로 정의한 바 있다. Davis(1989)와 Chin·Todd(1995)의 연구에서 의미하는 인지된 유용성

을 바탕으로 본 연구에서의 인지된 유용성은 ‘빅데이터 기술을 적용 시 개인의 성과를 향상시킬 수 있다는 믿음의 정도’로 정의하였다. 관련 문항은 총 5개로 구성하였으며, 7점 척도로 측정하였으며, Cronbach’s alpha .894로 나타났다.

5) 인지된 용이성

인지된 용이성(easy to use)은 기술 이용자가 많은 노력 없이도 새롭게 도입된 기술을 사용할 수 있을 것으로 기대하는 정도를 말한다. 선행 연구들에 의하면 이용에 편이성을 느끼는 기술 및 제품은 그렇지 않은 경우보다 긍정적 태도 및 수용의도가 형성된다는 연구결과가 다수 제시되었다(Davis, 1989; Ajzen, 1991; 김성개, 2009; 김태문·한진수, 2009; 김성일, 2012). 본 연구에서 인지된 용이성은 빅데이터 기술이 ‘사용하기 편리하고 편의적 기능을 제공하며 과업을 수행함에 있어 신체적, 정신적 수고를 덜게 하는 정도’로 측정하였다. 관련 문항은 총 5개로 구성하였으며, 7점 척도로 측정하였으며, Cronbach’s alpha .908로 나타났다.

6) 사용태도

사용태도는 ‘빅데이터 기술 이용 이전에는 발생하지 않았지만 기술을 접하게 되었을 경우 느끼는 정서적 자극의 정도’이다. Davis(1989)와 Adams·Todd(1992)의 연구에 사용된 도구를 바탕으로 본 연구의 빅데이터 기술에 대한 사용태도는 ‘기술에 대한 정서적 호감도로 긍정성과 경제적 호감의 정도’를 측정하였다. 관련 문항은 5개이며, 7점 척도로 측정하였으며, Cronbach’s alpha .908로 나타났다.

7) 수용의도

수용의도는 행위를 수행하려는 개인적 의지의 강도라고 정의되었다 (Fishbein·Ajzen, 1975). 기술수용모델(TAM)에서의 수용의도는 태도에 의해 결정된다. 특히 선행연구 결과 인지된 유용성이 인지된 용이성보다 수용의도에 대해 보다 더 강하게 영향을 가지는 요인으로 검증되었다 (Venkatesh·Davis, 2000). 본 연구에서의 수용의도는 '이용자의 빅데이터 기술에 대한 직접적인 이용의도'로 정의하여 측정하였다. 관련 문항은 4개이며, 7점 척도로 측정하였으며, Cronbach's alpha .941로 나타났다.

8) 지속적 수용의도

실제 이용과 마찬가지로 지속적으로 기술을 이용할 것이라는 기술에 대한 충성도는 혁신기술을 수용하는 의사결정 과정에 매우 중요한 영향을 미치며(Agrawal·Prasad, 1997) 향후 혁신기술의 확산에 기반이 된다. 본 연구에서는 기본적으로 기술수용모델에서 제시하는 수용의도가 아닌 '빅데이터 기술의 실제 사용 이후 지속적인 수용의도'를 최종 종속변수로 활용하여 보다 정확한 혁신기술의 확산을 살펴보고자 하였다. 관련 문항은 2개로 구성하였으며, 7점 척도로 측정하였으며, Cronbach's alpha .838로 나타났다.

나. 측정절차 및 분석방법

설문조사는 2014년 4월 1일부터 4월 30일까지 온라인 설문조사 Tool

인 iweb survey3.0을 활용한 이메일 설문조사 및 오프라인 설문조사를 병행하여 빅데이터 유관 업무경험을 가진 현업 종사자들을 대상으로 진행하였다. 설문조사에 응답한 응답자들은 393명으로 대기업 혹은 중소기업에 근무하면서 빅데이터 기술 및 서비스를 직접 구매하여 이용하는 ‘기술이용자’로서의 응답자와 해당 기술을 활용하거나 빅데이터 기술기반으로 2차적 가공을 통하여 그 산출물 혹은 서비스를 제공하는 ‘기술활용자’로서의 응답자로 구성되었다. 특히 대기업의 경우에는 조직의 특성 및 담당 업무에 따라 동일 기업이지만 한 부서는 수용자로서 또 다른 부서는 공급자로서 응답이 가능하다. 설문에 응답한 응답자는 빅데이터 기술 및 관련 분야 기업 또는 부서에 근무하는 사람들로서 적절한 응답결과를 얻었으며 성실하게 응답하지 않은 샘플은 제외 후 분석에 사용하였다. 조사를 통해 측정된 설문항목은 조직의 혁신성, 주관적 규범, 과업기술적합성 및 기술에 대한 유용성과 용이성, 사용태도, 수용의도, 실제도입 및 지속 수용의도에 대한 질문으로 정하여 리커트(Likert) 7점 척도를 활용하여 측정하였다.

4. 분석방법 및 연구모형

가. 통계분석

구조모형의 적합도를 평가하기 위하여 χ^2 검정과 적합도 지수를 고려하였다. 여러 적합도 지수 중 에서 본 연구는 증분적합지수(Incremental Fit Index)인 CFI(Comparative Fit Index), NFI(Normed Fit Index), TLI(Tucker-Lewis Index)와 절대적합지수(Absolute Fit Index)인 GFI(Goodness Fit Index)와 SRMR(Standardized Root

Mean Squared Residual)을 사용하였다. 일반적으로 CFI를 비롯한 증분적합지수들은 0.9보다 크면 모형 적합도가 양호하다고 해석한다(Hu et al., 1999). GFI는 0~1사이의 값을 가지며 보통 0.9이상이면 양호한 수준이며, SRMR은 모델 간 비교 시 사용되는 지표로 0.05이하이면 양호한 수준이라고 할 수 있다. 본 연구의 간접효과에 대한 유의성 검증은 모든 자료가 정규성 분포를 따른다는 가정 하에 유의성을 검증하는 Sobel test(Preacher·Hayes, 2004)를 실시하였으며, 매개효과 검증은 Hair et al.(2006)가 제시한 검증방법을 사용하였다.

나. 연구모형 및 가설

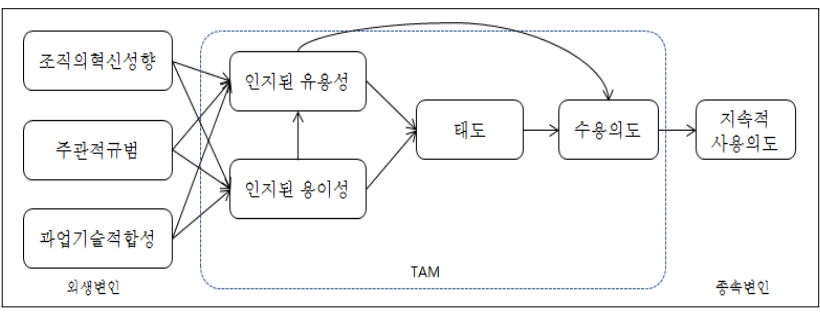
본 연구는 TAM과 TTF의 통합모형 및 혁신적인 조직문화를 가진 기업의 기술수용도가 높다는 선행연구 결과를 바탕으로 연구모형을 구성하였다. 연구의 실험데이터를 적용한 연구모형의 검증결과 일부 적합도 지수에서 낮게 나타나, 과업적합성이 수용의도에 직접 영향을 미친다는 Dishow·Strong(1999) 연구의 이론적 배경 하에 두 변인간의 경로를 추가로 설정하여 수정모형을 개발하였다. <표 5-3>과 같이 수정모형의 대부분의 적합도 평가지수가 높게 나타나 빅데이터 기술 수용모형은 수정 연구모델이 적합한 모형으로 판단할 수 있다.

〈표 5-3〉 모형비교

모형	χ^2	df	SRMR	GFI	NFI	TLI	CFI
연구모형	106.281	13	.077*	.941**	.930**	.865*	.937**
수정연구모형	64.355	12	.049**	.961**	.958**	.918**	.965**

주: **: 양호, *: 보통

〔그림 5-1〕 연구모형



본 연구는 초기 시장 내의 빅데이터 기술 수용의 특성을 연구하기 위하여 앞에서 제시된 수정 연구모형을 이용하여 세가지 연구문제를 제시하였다.

[연구문제 1] 조직의 혁신성, 주관적규범, 과업기술적합성 등의 외생변인과 기술수용모델의 변인이 지속사용 의도에 미치는 경로별 영향도는 어떠한가?

[연구문제 2] 조직의 혁신성, 주관적규범, 과업기술적합성 등의 외생변인과 기술수용모델의 변인이 지속사용 의도에 미치는 경로별에서 ‘기술이용자’와 ‘기술활용자’ 간에 차이는 있는가?

[연구문제 3] 조직의 혁신성, 주관적규범, 과업기술적합성 등의 외생변인은 TAM변인의 매개효과가 있는가?

5. 연구결과

가. 주요변인들의 기술통계

주요변수의 정규성을 검증하기 위해 기술분석을 실시한 결과, <표 5-4>와 같이 모든 변수에서 왜도(Skewness)는 절대값 3 미만, 첨도(Kurtosis)는 절대값 10 미만으로 나타나 정규성 가정을 충족(Kline, 2010)하는 것으로 나타났다.

<표 5-4> 주요 변인들의 기술통계

구분	전체			'기술이용자' 집단			'기술활용자' 집단		
	M±SD	왜도	첨도	M±SD	왜도	첨도	M±SD	왜도	첨도
조직혁신성	1.37±1.01	-1.21	2.97	5.38±.96	-.93	2.32	5.34±1.09	-1.55	3.56
주관적규범	4.67±1.09	-.49	.60	4.61±1.07	-.27	.03	4.78±1.12	-.88	1.80
과업기술적합성	4.65±1.05	.03	.13	4.50±1.09	.20	.24	4.93±.92	-.11	.13
용의성	3.84±1.22	.08	.28	3.71±1.19	.15	-.12	4.10±1.25	-.09	-.38
유용성	5.08±.95	-.67	1.50	5.06±.91	-.29	.44	5.13±1.03	-1.22	2.95
사용태도	5.03±.96	-.39	.86	4.00±.98	-.33	.86	5.09±.92	-.53	.97
수용의도	5.25±1.02	-.43	.50	5.16±1.09	-.41	.40	5.41±.85	-.16	-.29
지속사용	4.80±1.17	-.46	.27	4.69±1.24	-.51	.15	4.99±1.00	-.35	.19

나. 연구모형 분석

1) 연구모형의 적합도 검증

수정연구모형에서 제시한 외생변인(조직 혁신성, 주관적 규범, 과업기술적합성)이 TAM변인을 매개하여 지속사용에 어떠한 구조적인 관계를

가지는지 수정연구모형 전체, ‘기술이용자’, ‘기술활용자’를 대상으로 구조방정식모형을 통해 검증하였다. 빅데이터 수용자 전체와 ‘기술이용자’, ‘기술활용자’ 모두 연구모형과 실제 자료와의 적합도가 χ^2 를 제외하고는 모든 적합도에서 높게 나타났다(표 5-5 참조).

〈표 5-5〉 다중집단 구조모형의 적합도 비교

구분	χ^2	df	SRMR	GFI	NFI	TLI	CFI
전체	64.355	12	.049**	.961**	.958**	.918**	.965**
기술이용자	62.813	12	.054*	.945**	.945**	.894*	.955**
기술활용자	33.917	12	.049**	.941**	.925**	.879*	.948**

주: *: 보통, **: 양호

2) 전체 모형의 경로분석

연구문제 1을 검증하기 위하여 〈표 5-6〉과 같이 빅데이터 전체 수용자에 대해 외생변인들과 TAM변인들이 빅데이터 지속사용에 미치는 영향을 살펴보았다. 외생변인인 주관적규범($b=.173$, $p<.01$), 과업기술적합성($b=.313$, $p<.01$)은 용이성에 정적(+)인 유의한 영향을 주는 것으로 나타났다, 조직의 혁신성은 용이성에 부적(-)($b=-.138$, $p<.01$)인 영향을 미치는 것으로 나타났다. 외생변인인 조직혁신성($b=.118$, $p<.01$), 주관적규범($b=.298$, $p<.01$), 과업기술적합성($b=.353$, $p<.01$)은 유용성에 정적(+)인 유의한 영향을 주는 것으로 나타났다. 용이성은 유용성에 정적(+) 영향($b=.179$, $p<.01$)을 미치는 것으로 나타났다. 유용성($b=.551$, $p<.01$)과 용이성($b=.181$, $p<.01$)은 사용태도에 정적(+)인 영향을 미치는 것으로 나타났다. 사용태도($b=.573$, $p<.01$)은 수용의도에 정적(+)인 영향을 미치는 것으로 나타났다. 외생변수인 과업기술적합성은 수용의도

에 정적 영향($b=.287$, $p<.01$)을 미치는 것으로 나타났다. 수용의도는 지속사용에 정적(+) 영향($b=.805$, $p<.01$)을 미치는 것으로 나타났다.

3) 기술이용자와 기술활용자의 다중집단 구조모형 분석

다중집단 구조모형 분석은 집단 간 경로계수를 가지고 서로 통계적인 차이 여부를 검증하는 것이다. 다중집단 구조모형 분석은 측정동일성 제약이 끝난 후, 집단 간 등가제약 과정을 거쳐 경로 간 유의미한 차이를 검증할 수 있다. 본 연구에서의 구조모형은 경로모형으로, 각 요인에 대한 측정동일성을 검증할 필요가 없기 때문에 요인사이의 경로계수로 집단 간(사용목적 수용자인 ‘기술이용자’, 활용목적 수용자인 ‘기술활용자’) 차이를 검증하였다. 연구문제 2의 빅데이터 수용요인에 대한 ‘기술이용자’와 ‘기술활용자’의 차이를 살펴보기 위하여 다중집단 구조모형 분석을 실시하였다. 전체 구조모형이 ‘기술이용자’와 ‘기술활용자’ 집단 간에 차이가 있는지 확인한 다음, 두 집단의 각각의 경로에서 차이가 있는지를 검증한 결과 ‘기술이용자’ 집단과 ‘기술활용자’ 집단 간에 유의미한 차이가 있었다($\chi^2(24)=96.732$, $p<.000$). 외생변수가 용이성에 미치는 영향은 두 집단 간 유의한 차이가 있는 것으로 나타나, ‘기술이용자’($b=.281$, $p<.01$)가 ‘기술활용자’($b=.056$, $p<.01$) 보다 주관적 규범에서 용이성으로 가는 경로가 유의하게 큰 것으로 나타났다. 반면, ‘기술활용자’($b=.414$, $p<.01$)가 ‘기술수용자’($b=.171$, $p<.01$) 보다 과업기술적합성에서 용이성으로 가는 경로가 유의하게 큰 것으로 나타났다. 과업기술적합성이 유용성에 미치는 영향은 두 집단 간 유의한 차이가 있는 것으로 나타나($C.R.=3.490$, $p<.01$), ‘기술활용자’($b=.515$, $p<.01$)가 ‘기술이용자’($b=.306$, $p<.01$) 보다 과업기술적합성에서 유용성으로 가는 경로가

유의하게 큰 것으로 나타났다. 유용성이 사용태도에 미치는 영향은 두 집단 간 유의한 차이가 있는 것으로 나타나(C.R.=-3.652, $p<.01$), ‘기술이용자’(b=.645, $p<.01$)가 ‘기술활용자’(b=.412, $p<.01$) 보다 유용성에서 사용태도로 가는 경로가 유의하게 큰 것으로 나타났다. 유용성에서 수용의도에 미치는 영향은 두 집단 간 유의한 차이가 있는 것으로 나타나(C.R.=-2.326, $p<.05$), ‘기술이용자’는 유용성이 수용의도에 정적인 영향(b=113, $p<.1$)을 미치는 것으로 나타났으며, ‘기술활용자’는 유용성이 수용의도에 부적(-)인 영향(b=-.139, $p>.1$)을 미치는 것으로 나타났으나 통계적으로 유의하지 않는 것으로 나타났다.

〈표 5-6〉 사용목적 수용자와 활용목적 수용자 집단 간의 구조모형 분석결과

Path	전체수용자		기술이용자		기술활용자		C.R.
	β	C.R.	β	C.R.	β	C.R.	
조직혁신성→용이성	-.138	-2.65***	.000	.01	-.293	-3.34***	-2.608***
주관적규범→용이성	.173	3.23***	.281	4.18***	.056	.64	-2.044**
과업기술적합성→용이성	.313	5.51***	.172	2.30**	.414	4.64***	2.589***
조직혁신성→유용성	.118	2.92***	.050	.99	-.177	2.70***	1.521
주관적규범→유용성	.298	7.14***	.342	6.37***	.206	3.25***	-1.423
과업기술적합성→유용성	.353	7.77***	.306	5.29***	.515	7.42***	3.490***
용이성→유용성	.179	4.59***	.223	4.63***	.110	1.77*	-1.290
유용성→사용태도	.551	13.13***	.645	13.25***	.412	5.33***	-3.652***
용이성→사용태도	.182	4.34***	.144	2.95***	.199	2.58***	.426
사용태도→수용의도	.573	13.16***	.504	8.64***	.648	9.40***	.404
유용성→수용의도	.002	0.04	.113	1.75*	-.139	-1.54	-2.326**
과업기술적합성→수용의도	.287	6.67***	.269	5.33***	.269	3.15***	-1.188
수용의도→지속사용	.805	26.89***	.837	24.47***	.700	11.39***	-1.611

주: *** $P<.01$, ** $p<.05$, * $p<.1$

4) 외부변인의 매개효과 검증

연구문제 3을 검증하기 위해 외생변인과 유용성의 경로에 용이성의 매개효과(mediator effect)와 과업기술적합성과 수용의도 경로에 유용성의 매개효과를 살펴보았다. <표 5-7>에서 용이성에 대한 간접효과의 유의성 검증을 위한 Sobel test 결과 ‘기술이용자’의 간접효과가 모든 외생변인에서 유의한 것($p < .05$, $p < .01$)으로 나타났으나, ‘기술활용자’의 간접효과는 유의하지 않는 것(조직혁신성: $sb=1.241$, $p=.21$, 주관적규범: $sb=1.25$, $p=.21$, 과업적합성: $sb=.91$, $p=.37$)으로 나타났다. Hair 외(2006)이 제시한 매개효과를 검증한 결과 ‘기술이용자’와 ‘기술활용자’ 집단 모두 외생변인에서 유용성으로 경로에서 용이성이 부분매개(partial mediation) 하는 것으로 나타났다. <표 5-7>에서 과업적합성과 수용의도의 경로에 유용성 대한 간접효과의 유의성은 기술이용자는 유의한 것($p < .01$)으로 나타났으나, ‘기술활용자’는 유의하지 않는 것($sb=1.203$, $p=.229$)으로 나타났다. 그리고, ‘기술이용자’와 ‘기술활용자’ 집단 모두 과업기술적합성에서 수용의도 경로에서 유용성이 부분매개(partial mediation) 하는 것으로 나타났다.

<표 5-7> 외생변인과 TAM변인의 매개효과 검증

Path	전체			기술이용자			기술활용자		
	Total Effect	Direct Effect	Indirect Effect ¹⁾	Total Effect	Direct Effect	Indirect Effect ¹⁾	Total Effect	Direct Effect	Indirect Effect ¹⁾
조직혁신성→유용성	.366	.348	.018	.318	.254	.064*	.439	.476	-.037
DP→MP ²⁾		.366*→.348*			.318*→.254*			.439*→.476*	
주관적규범→유용성	.554	.483	.071**	.598	.487	.100**	.483	.456	.026
DP→MP ²⁾		.554*→.483*			.598*→.487*			.483*→.456*	
과업적합성→유용성	.609	.540	.070**	.580	.486	.096**	.700	.681	.091
DP→MP ²⁾		.609*→.540*			.580*→.486*			.700*→.681*	
과업기술적합성→수용의도	.529	.327	.202*	.559	.297	.262*	.406	.313	.093
DP→MP ²⁾		.529*→.327*			.559*→.397*			.406*→.313*	

주: 1) Sobel Test: ** $p < 0.01$, * $p < 0.05$.

2) Mediator Effect: DP(Direct Path coefficient), MP(Mediator Path coefficient)

6. 논의

본 연구는 태동기인 국내 빅데이터 기술의 수용특성을 연구하기 위해 기술수용 과정의 연구모델로서 보편성과 적합성을 확보한 기술수용모델(TAM)을 기본 프레임으로 혁신확산이론(IDT)과 과업기술적합성이론(TTF)의 통합적인 접근을 시도하였다. 선행연구를 바탕으로 연구모형을 설계하고 시장에서 기업이 빅데이터 기술을 수용하는 경우 그 목적과 수용과 관련한 성격이 상이한 부분에 주안점을 두어 이를 조절변인화하여 구조방정식의 다중집단 분석을 적용한 결과 연구문제에 대한 분석결과 및 각각의 함의는 다음과 같다.

연구문제 1을 분석하기 위해 연구모형으로 설정한 ‘조직의 혁신성’, ‘주관적 규범’, ‘과업기술적합성’과 TAM 변인 간 경로분석을 수행한 결과, 가장 큰 영향을 미치는 요인은 ‘주관적 규범’과 ‘과업기술적합성’이었다. 설정한 세 개의 외생변인인 ‘조직의 혁신성’과 ‘주관적 규범’, ‘과업기술적합성’은 모두 ‘유용성’ 경로에 정적(+)인 영향을 가지는 것으로 나타났다. ‘조직의 혁신성’에서 ‘용이성’ 경로는 부적(-)인 영향을 나타내는 것으로 확인되었다.

또한 TAM 변인 간 경로 또한 유의하게 정적(+)인 효과를 가지는 것으로 확인되었으나 선행연구들과는 다르게 ‘유용성’에서 ‘수용의도’로 직접 이어지는 경로는 유의하지 않은 것으로 나타났다. 반면 ‘과업기술적합성’에서 ‘수용의도’로 직접 이어지는 경로는 유의하게 나타나 빅데이터 기술 수용에 있어서는 기술의 ‘유용성’보다 ‘과업기술적합도’가 보다 유의한 영향력을 갖는 것으로 확인되었다. ‘주관적 규범’이 ‘조직의 혁신성’보다 유의한 효과를 가짐이 증명되었는데, 이는 혁신기술 초기 시장에서의 빅데이터 기술수용과 관련하여 현재 국내 시장에서는 빅데이터 기술

자체에 대한 선택이 사회적인 환경 안에서 수용자를 보다 우월하게 만들 것이라는 기대사항이 반영된 것으로 해석할 수 있다. 이는 초기 국내 빅데이터 기술수용의 특성이라 할 수 있으며, 합목적적이고 합리적인 기술수용 현상보다는 기술수용에 대한 사회적 영향도가 높게 작용하여 빅데이터 기술이 수용되고 있음이 확인된 결과이다. 조직의 학습단계 상 현재 상황이 기업 내 소수의 혁신적 의사결정을 하는 그룹에 의해 기술이 흡수(Assimilation)되는 단계인 것으로 유추할 수 있는데 설문에 응답한 결과 중 직급과 연차가 높을수록 빅데이터 기술수용 의도가 높게 나타난 점을 고려할 때 현재 기업 내의 의사결정권자들의 빅데이터 기술에 대한 기대수준이 높음을 알 수 있었다. 연구문제 2에 대한 분석결과 빅데이터 기술을 단순히 이용하려는 목적인 ‘기술이용자’ 집단과 빅데이터 기술기반으로 새로운 서비스나 상품을 만들하고자하는 ‘기술활용자’ 집단의 연구 모델별 경로는 유의한 차이를 가지는 것이 확인되었다. 특히 ‘기술이용자’ 집단의 경우 ‘주관적 규범’에서 ‘용이성’으로 가는 경로가 ‘기술활용자’ 집단보다 유의하게 큰 것으로 나타났는데 ‘기술이용자’ 그룹의 경우 ‘주관적 규범’과 ‘과업기술적합성’의 영향 후 기술의 ‘용이성’과 ‘유용성’을 거쳐 ‘지속적 사용의도’에 이르는 모든 경로가 정(+)의 방향으로 유의함을 나타냈다.

‘조직의 혁신성’에서 ‘용이성’과 ‘유용성’으로 이어지는 경로는 ‘기술이용자’ 집단에게는 모두 유의하지 않은 것으로 나타났으나 ‘기술활용자’ 집단에는 모두 유의하게 나타났다. 다만 두 집단 간의 차이에 대한 유의성 검증을 한 결과에서는 조직의 혁신성에서 용이성으로 이어지는 경로에서 ‘기술활용자’ 집단에 보다 유의한 차이를 갖는 것으로 확인되었다. 이는 분석 대상이 되는 표본이 커질 경우 통계적으로 유의한 차이를 가질 수 있는 부분이므로 해석에 주의할 필요가 있다. 다시 말하면 기술기반의

새로운 서비스나 상품을 기획하고 개발하고자 하는 목적의 ‘기술활용자’ 집단에게 ‘조직의 혁신성향’은 기술의 수용의도에 영향을 미치는 유의한 외생변인이나 단순히 기술을 이용하고자 하는 ‘기술이용자’ 집단에게는 오히려 ‘주관적 규범’이 영향력을 가진 것으로 나타났다. 유용성에서 수용의도로 직접 이어지는 경로에서도 ‘기술이용자’ 집단은 유의한 정의효과를 나타내었으나 ‘기술활용자’ 집단은 유의하지 않았으며 오히려 부정(-)결과를 보였다. 이것은 기업의 빅데이터 기술수용 과정에 있어 그 목적성에 따라 조금 더 쉬운 의사결정 과정이 가능하거나 그 반대일 수도 있다는 것을 반증하는 결과이다. 대체적으로 새로운 기술이나 서비스를 도입하는 경우 기업의 의사결정과정은 해당 기술이나 서비스를 검토하는 주관부서가 주축이 되어 의사결정을 하게 되지만 새로운 서비스나 상품을 만들기 위한 기술은 투자에 대한 의사결정 과정에서 많은 이해관계 부서가 참여하게 된다. 그러므로 기술의 유용성이 직접적인 수용의도도 나타나기 어려울 것이다. 연구문제 3의 분석결과와 TAM 변인에 대한 매개효과에서는 연구문제 2의 분석결과를 통해 각 집단별 유의한 차이가 검증된 경로에 대하여 용이성 변인의 매개효과 여부를 검증하였다. 분석결과 ‘기술이용자’ 집단은 ‘용이성’의 매개효과가 검증되었고 ‘기술활용자’ 집단은 유의하지 않았다. 이는 빅데이터 기술을 도입함에 있어 단순 이용 목적의 수용자 집단에게 기술의 ‘용이성’이 ‘유용성’에 영향력을 가짐이 확인된 것이다. ‘과업기술적합성’에서 ‘수용의도’로 이어지는 직접 경로에서도 ‘기술이용자’ 집단에만 TAM변인의 매개효과가 유의하게 검증되었다. 요약하자면 기술수용과정에서 ‘기술이용자’ 집단은 TAM에서 제시하는 단계별 의사결정의 경로가 모두 유의하게 나타난 반면 빅데이터 기술을 활용하여 새로운 제품과 서비스를 개발하려는 ‘기술활용자’ 집단의 경우에는 TAM의 매개효과가 유의하지 않음이 확인되었다.

제2절 공공과 보건의료 빅데이터 수요공급 예측¹³⁾

본 연구는 국내에서 빅데이터가 처음으로 논의된 2011년부터 2013년 12월까지의 우리나라 온라인 뉴스사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 바탕으로 국내 빅데이터의 동향과 수요공급 예측모형을 제시하고자 한다. 본 연구의 목적은 소셜 빅데이터를 활용하여 데이터마이닝의 의사결정나무 분석을 통해 국내의 빅데이터 수요공급 예측모형을 제시하는 것으로 구체적인 목적은 다음과 같다. 첫째, 국내의 빅데이터 수요공급에 영향을 미치는 요인을 파악한다. 둘째, 국내의 빅데이터 수요공급을 예측할 수 있는 의사결정나무를 개발한다.

1. 연구 방법

가. 연구대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅 데이터를 대상으로 하였다. 본 분석에서는 214개의 온라인 뉴스사이트, 4개의 블로그(네이버, 네이트, 다음, 티스토리), 3개의 카페(네이버, 다음, 뽐뿌), 2개의 SNS(트위터, 미투데이), 4개의 게시판(네이버지식인, 네이트지식, 네이트톡, 네이트판) 등 총 227개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서(버즈)를 소셜 빅데이터로 정의하였다. 빅데이터 관련 토픽의 수집은 2011. 1. 1~2013. 12. 31까지 해당 채널에서 요일별, 주말, 휴일을

13) 본 절의 일부 내용은 본 과제의 일환으로 작성된 '김정선·권은주·송태민(2014). 분석지의 확장을 위한 소셜 빅데이터 활용연구-국내 '빅데이터' 수요공급 예측-. 지식경영연구 15(3). 173-192.'에서 발췌한 내용임을 밝힌다.

고려하지 않고 매 시간단위로 수집하였으며, 수집된 총 70,485건(2011년: 2,159건, 2012년: 25,023건, 2013년: 43,303건)의 텍스트(Text) 문서를 본 연구의 분석에 포함시켰다. 본 연구를 위한 소셜 빅 데이터의 수집은 크롤러(Crawler)를 사용하였고, 토픽의 분류는 주제분석(text mining) 기법을 사용하였다. 크롤러는 인터넷 링크를 따라다니며 방문한 사이트의 모든 페이지의 복사본을 생성함으로써 문서를 수집한다. 주제 분석은 자연어 처리기술을 이용하여 유용한 정보를 추출하거나 연계성을 파악하고, 분류 혹은 군집화함으로써 소셜 빅데이터의 의미있는 정보를 발견하는 것이다. 빅데이터 토픽은 모든 관련 문서를수집하기 위해 ‘빅데이터’와 ‘Big Data’를 사용하였으며, 수집기간에 빅데이터와 관련 없는 용어인 ‘컨퍼런스, 전시회’의 2개의 불용어를 지정하여 수집하였다.

나. 연구도구

빅데이터와 관련하여 수집된 버즈는 주제분석(text mining)과 요인분석(factor analysis)의 과정을 거쳐 다음과 같이 정형화 데이터로 코드화하여 사용하였다.

1) 빅데이터 관련 수요자 공급자 정의

본 연구의 종속변수인 수요자와 공급자의 정의는 주제분석에서 분류된 수요자의 태도(결정, 계획, 고려, 도입, 선정, 수용, 이용, 필요, 활용)는 수요자, 공급자의 태도(공급, 구축, 구현, 사례, 소개, 운영, 전망, 출시, 형성, 준비, 진행, 참여, 육성)는 공급자로 정의하였다.

2) 빅데이터 관련 기술

빅데이터 관련기술은 선행연구에서 분석된 분석기술, 처리기술, 저장 기술, 수집기술, 활용기술, 기타기술로 정의하였다. 분석기술로는 ‘NDAP, 기계학습, 네트워크, 마이닝, 시맨틱, 온톨로지, 패턴인식, EDW, 텍스트마이닝, 음성인식, 영상인식, AR, AI, mining, Mahout, semantic, ETL, 알고리즘, 알고리즘, Algorithm, Pregel, 데이터마이닝’으로 정의하였다. 처리기술은 ‘MapReduce, 가상화, 맵리듀스, 스케일아웃, 어플라이언스, 하둡, Inmemory, Hadoop, Hbase, MR, Appliance, Dremel, Percolator, 검색, 인텍싱, Storm, 스톱, HDFS, 인메모리’로 정의하였다. 저장기술은 ‘데이터베이스, 데이터웨어하우스, 클라우드, 클라우드컴퓨팅, Cloud, 디스크, NoSQL, DW, 데이터웨어하우스, san’로 정의하였다. 수집기술은 ‘크롤러, Crawler, 수집기, X86, Cassandra’로 정의하였다. 활용기술은 ‘BI, 인포그래픽스, SaaS, Paas, BaaS, Daas’로 정의하였다. 기타기술은 ‘apache, 아파치, 자바, Java, html5’로 정의하였다. 정의된 모든 기술은 해당기술이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

3) 빅데이터 관련 산업

빅데이터 관련 산업은 ‘의료헬스, 패션의류, 정부공공기관, 게임, 금융보험, 선거정치, 영화, 자동차교통, 통신, 제조, IT벤처’의 11개로 해당 산업이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

4) 빅데이터 관련 기반

빅데이터 관련 기반은 ‘프라이버시(개인정보, 보안, 프라이버시), 분석 전문가(인력, 전문가), 정책전략(관리, 정책, 전략), 기반구축(서버, 스토리지, 시스템, 인프라, 컴퓨팅, 플랫폼, 기술), 비용, 서비스, 콘텐츠, 품질’의 8개로 해당 기반이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

5) 빅데이터 관련 업무

빅데이터 관련 업무는 ‘고객관리마케팅(고객관리, 마케팅, 영업), 생산 관리(생산, 공정관리, 물류관리), 인사재무(인사, 재무, 전략기획, 회계, 전산), 연구개발, 의사결정, 컨설팅, 통계’의 7개로 해당 기반이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

6) 빅데이터 관련 국가

빅데이터 관련 국가는 ‘한국, 글로벌, 유럽, 미국, 일본, 중국, 싱가포르’의 7개국으로 해당 국가가 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

7) 빅데이터 관련 감정

빅데이터 관련 감정은 주제분석과 요인분석의 과정을 거쳐 ‘성장(성장, 강화, 경쟁력, 수요, 주요), 관심(유용, 도움, 관심, 이해, 고민, 중요, 가

능, 발전), 효율(절감, 효과, 효율, 부담, 기대), 혁신(혁신, 성공, 최고, 위기), 경제적(경제적, 불가능) 우려(논란, 우려), 신속(신속, 정확), 신뢰(용이, 신뢰)의 8개로 해당 감정이 있는 경우는 '1', 없는 경우는 '0'으로 코딩화 하였다.

다. 분석 방법

본 연구에서는 한국의 빅데이터 수요공급을 설명하는 가장 효율적인 예측모형을 구축하기 위해 특별한 통계적 가정이 필요하지 않은 데이터마이닝의 의사결정나무 분석방법을 사용하였다. 데이터마이닝의 의사결정나무 분석은 방대한 자료 속에서 종속변인을 가장 잘 설명하는 예측모형을 자동적으로 산출해 줌으로써 각기 다른 속성은 가진 빅데이터 수요공급에 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 분리기준으로 카이제곱(χ^2 -검정)을 사용하며, 모든 가능한 조합을 탐색하여 최적분리를 찾는다. 정지규칙(stopping rule)으로 관찰치가 충분하여 상위노드(부모마디)의 최소케이스 수는 100으로 하위노드(자식마디)의 최소 케이스 수는 50으로 설정하였고, 나무깊이는 3수준으로 정하였다. 본 연구는 기술분석, 다중응답분석, 로지스틱 회귀분석, 의사결정나무 분석 방법을 위해 SPSS 22.0을 사용하였다.

2. 연구 결과

가. 빅데이터 관련 분야 버즈 현황

1) 전체 분야 빅데이터 버즈 현황

〈표 5-8〉과 같이 한국과 관련한 기술분야 버즈는 저장기술, 처리기술, 분석기술, 활용기술 등의 순으로 나타났다. 기반분야 버즈는 기반구축, 서비스, 정책전략, 분석전문가, 프라이버시, 비용 등의 순으로 나타났다. 산업분야 버즈는 정부공공기관, 자동차교통, 통신, 보험금융, 선거정치, 게임, 의료헬스, 제조 등의 순으로 나타났다. 글로벌과 관련한 기술분야 버즈는 저장기술, 처리기술, 분석기술, 활용기술 등의 순으로 나타났다. 기반분야 버즈는 기반구축, 정책전략, 서비스, 분석전문가, 비용, 프라이버시 등의 순으로 나타났다. 산업분야 버즈는 정부공공기관, 자동차교통, 통신, 보험금융, 영화, 게임, 선거정치, 제조, 의료헬스 등의 순으로 나타났다. 〈표 5-9〉와 같이 의료헬스 산업과 관련한 업무분야 버즈는 고객마케팅, 생산관리, 통계, 인사재무, 컨설팅, 의사결정, 연구개발 순으로 나타났다. 기술분야 버즈는 저장기술, 분석기술, 처리기술, 기타기술, 활용기술, 수집기술 순으로 나타났다. 기반분야 버즈는 기반구축, 서비스, 정책전략, 분석전문가, 프라이버시, 비용, 콘텐츠, 품질 순으로 나타났다. 정부공공기관 산업과 관련한 업무분야 버즈는 고객마케팅, 생산관리, 컨설팅, 통계, 의사결정, 연구개발, 인사재무 순으로 나타났다. 기술분야 버즈는 저장기술, 분석기술, 처리기술, 활용기술, 기타기술, 수집기술 순으로 나타났다. 기반분야 버즈는 기반구축, 정책전략, 서비스, 분석전문가, 프라이버시, 비용, 콘텐츠, 품질 순으로 나타났다.

〈표 5-8〉 국가별 빅데이터 관련 분야 버즈 현황

N(%)									
분야	속성	한국	글로벌	유럽	미국	일본	중국	싱가포르	합계
기술	분석기술	1940 (23.8)	1923 (25.9)	429 (27.2)	1049 (26.2)	396 (26.0)	373 (28.4)	58 (27.0)	6168 (25.5)
	처리기술	2318 (28.5)	1967 (26.4)	426 (27.0)	1181 (29.5)	449 (29.4)	333 (25.4)	64 (29.8)	6738 (27.8)
	저장기술	3215 (39.5)	2939 (39.5)	634 (40.2)	1451 (36.3)	604 (39.6)	550 (41.9)	82 (38.1)	9475 (39.1)
	수집기술	18 (.2)	33 (.4)	5 (.3)	13 (.3)	4 (.3)	4 (.3)	0 (0.0)	77 (.3)
	활용기술	436 (5.4)	394 (5.3)	59 (3.7)	181 (4.5)	55 (3.6)	37 (2.8)	8 (3.7)	1170 (4.8)
	기타기술	220 (2.7)	182 (2.4)	25 (1.6)	122 (3.1)	17 (1.1)	15 (1.1)	3 (1.4)	584 (2.4)
	합계	8147	7438	1578	3997	1525	1312	215	24212
기반	프라이버시	1294 (6.9)	1198 (6.8)	395 (8.8)	763 (7.8)	280 (6.7)	250 (7.0)	40 (7.7)	4220 (7.2)
	분석전문가	1463 (7.8)	1355 (7.7)	351 (7.8)	873 (8.9)	320 (7.6)	272 (7.7)	35 (6.8)	4669 (7.9)
	정책전략	3849 (20.5)	3628 (20.6)	959 (21.3)	2061 (21.0)	883 (21.0)	760 (21.4)	114 (22.0)	12254 (20.8)
	기반구축	5724 (30.4)	5412 (30.8)	1282 (28.5)	2865 (29.2)	1217 (29.0)	1082 (30.4)	156 (30.1)	17738 (30.1)
	비용	1253 (6.7)	1259 (7.2)	300 (6.7)	751 (7.7)	309 (7.4)	253 (7.1)	40 (7.7)	4165 (7.1)
	서비스	3919 (20.8)	3423 (19.5)	894 (19.9)	1805 (18.4)	875 (20.8)	681 (19.2)	98 (18.9)	11695 (19.8)
	콘텐츠	936 (5.0)	932 (5.3)	221 (4.9)	474 (4.8)	194 (4.6)	168 (4.7)	24 (4.6)	2949 (5.0)
	품질	371 (2.0)	384 (2.2)	96 (2.1)	207 (2.1)	122 (2.9)	88 (2.5)	11 (2.1)	1279 (2.2)
	합계	18809	17591	4498	9799	4200	3554	518	58969
산업	의료헬스	490 (6.7)	444 (6.4)	163 (6.5)	327 (7.0)	153 (6.8)	101 (4.9)	38 (13.1)	1716 (6.6)
	패션의류	208 (2.8)	191 (2.7)	87 (3.5)	143 (3.0)	69 (3.1)	59 (2.9)	8 (2.8)	765 (2.9)
	정부공공기관	2375 (32.4)	2053 (29.5)	795 (31.9)	1377 (29.3)	652 (29.2)	621 (30.2)	101 (34.8)	7974 (30.6)
	게임	514 (7.0)	535 (7.7)	149 (6.0)	314 (6.7)	170 (7.6)	170 (8.3)	10 (3.4)	1862 (7.1)
	영화	441 (6.0)	539 (7.7)	153 (6.1)	393 (8.4)	143 (6.4)	155 (7.5)	17 (5.9)	1841 (7.1)
	통신	708 (9.7)	733 (10.5)	221 (8.9)	356 (7.6)	187 (8.4)	155 (7.5)	22 (7.6)	2382 (9.1)

분야	속성	한국	글로벌	유럽	미국	일본	중국	싱가포르	합계
	제조	463 (6.3)	434 (6.2)	127 (5.1)	191 (4.1)	127 (5.7)	129 (6.3)	20 (6.9)	1491 (5.7)
	보험금융	777 (10.6)	724 (10.4)	253 (10.2)	415 (8.8)	200 (9.0)	217 (10.6)	27 (9.3)	2613 (10.0)
	선거정치	519 (7.1)	455 (6.5)	221 (8.9)	582 (12.4)	173 (7.7)	173 (8.4)	14 (4.8)	2137 (8.2)
	자동차 교통	826 (11.3)	852 (12.2)	322 (12.9)	593 (12.6)	358 (16.0)	272 (13.2)	33 (11.4)	3256 (12.5)
	IT벤더	8 (.1)	9 (.1)	1 (.0)	10 (.2)	2 (.1)	1 (.0)	0 (0.0)	31 (0.1)
	합계	7329	6969	2492	4701	2234	2053	290	26068

〈표 5-9〉 산업별 빅데이터 관련 분야 버즈 현황

		N(%)											
분야	속성	의료 헬스	패션 의류	정부공 공기관	게임	영화	통신	제조	보험 금융	선거 정치	자동차 교통	IT 벤더	합계
업무	고객관리 마케팅	172 (24.4)	149 (40.3)	781 (25.8)	251 (37.7)	280 (36.8)	306 (31.1)	241 (31.8)	464 (36.1)	273 (35.9)	357 (28.5)	2 (25.0)	3276 (31.0)
	생산관리	116 (16.4)	91 (24.6)	609 (20.1)	140 (21.0)	152 (20.0)	203 (20.7)	188 (24.8)	211 (16.4)	144 (18.9)	326 (26.0)	4 (50.0)	2184 (20.6)
	인사재무	85 (12.0)	21 (5.7)	270 (8.9)	53 (8.0)	59 (7.8)	66 (6.7)	33 (4.3)	82 (6.4)	83 (10.9)	80 (6.4)	1 (12.5)	833 (7.9)
	연구개발	65 (9.2)	5 (1.4)	289 (9.5)	36 (5.4)	37 (4.9)	70 (7.1)	23 (3.0)	45 (3.5)	15 (2.0)	90 (7.2)	0 (0.0)	675 (6.4)
	의사결정	74 (10.5)	26 (7.0)	310 (10.2)	49 (7.4)	84 (11.0)	108 (11.0)	93 (12.3)	128 (10.0)	68 (8.9)	117 (9.3)	1 (12.5)	1058 (10.0)
	컨설팅	79 (11.2)	39 (10.5)	400 (13.2)	65 (9.8)	55 (7.2)	143 (14.5)	124 (16.3)	193 (15.0)	57 (7.5)	106 (8.5)	0 (0.0)	1261 (11.9)
	통계	115 (16.3)	39 (10.5)	372 (12.3)	72 (10.8)	94 (12.4)	87 (8.9)	57 (7.5)	161 (12.5)	120 (15.8)	177 (14.1)	0 (0.0)	1294 (12.2)
	합계	706	370	3031	666	761	983	759	1284	760	1253	8	10581
기술	분석기술	276 (26.3)	90 (33.2)	1086 (26.3)	349 (28.7)	302 (30.9)	570 (30.7)	246 (25.0)	424 (26.1)	274 (34.5)	483 (29.9)	6 (20.0)	4106 (28.2)
	처리기술	255 (24.3)	101 (37.3)	1042 (25.3)	298 (24.5)	297 (30.4)	448 (24.1)	282 (28.6)	418 (25.8)	240 (30.2)	435 (26.9)	7 (23.3)	3823 (26.3)
	저장기술	433 (41.2)	75 (27.7)	1729 (41.9)	470 (38.6)	326 (33.4)	696 (37.5)	372 (37.8)	626 (38.6)	261 (32.9)	588 (36.4)	12 (40.0)	5588 (38.4)
	수집기술	7 (.7)	1 (.4)	12 (.3)	1 (.1)	4 (.4)	13 (.7)	3 (.3)	13 (.8)	2 (.3)	9 (.6)	0 (0.0)	65 (.4)

분야	속성	의료 헬스	패션 의류	정부공 공기관	게임	영화	통신	제조	보험 금융	선거 정치	자동차 교통	IT 벤더	합계
	활용기술	36 (3.4)	3 (1.1)	188 (4.6)	64 (5.3)	32 (3.3)	97 (5.2)	72 (7.3)	115 (7.1)	12 (1.5)	66 (4.1)	4 (13.3)	689 (4.7)
	기타기술	43 (4.1)	1 (0.4)	68 (1.6)	35 (2.9)	15 (1.5)	33 (1.8)	10 (1.0)	26 (1.6)	5 (.6)	36 (2.2)	1 (3.3)	273 (1.9)
	합계	1050	271	4125	1217	976	1857	985	1622	794	1617	30	14544
기반	프라이 버시	210 (7.8)	44 (4.9)	961 (8.0)	209 (7.5)	183 (7.2)	352 (8.7)	171 (7.5)	325 (8.2)	148 (6.5)	298 (6.9)	8 (15.4)	2909 (7.7)
	분석 전문가	219 (8.2)	82 (9.1)	1011 (8.4)	193 (6.9)	208 (8.1)	269 (6.7)	169 (7.4)	340 (8.6)	238 (10.5)	336 (7.7)	1 (1.9)	3066 (8.1)
	정책전략	550 (20.5)	173 (19.2)	2906 (24.1)	529 (18.9)	485 (19.0)	776 (19.2)	461 (20.3)	831 (21.0)	597 (26.4)	894 (20.6)	12 (23.1)	8214 (21.7)
	기반구축	753 (28.0)	230 (25.5)	3349 (27.8)	794 (28.4)	735 (28.7)	1134 (28.1)	678 (29.9)	1119 (28.3)	616 (27.2)	1271 (29.3)	15 (28.8)	10694 (28.2)
	비용	204 (7.6)	78 (8.6)	635 (5.3)	158 (5.6)	195 (7.6)	266 (6.6)	184 (8.1)	259 (6.6)	143 (6.3)	303 (7.0)	1 (1.9)	2426 (6.4)
	서비스	596 (22.2)	213 (23.6)	2436 (20.2)	574 (20.5)	500 (19.5)	867 (21.5)	431 (19.0)	826 (20.9)	363 (16.0)	956 (22.0)	14 (26.9)	7776 (20.5)
	콘텐츠	93 (3.5)	61 (6.8)	513 (4.3)	270 (9.6)	202 (7.9)	275 (6.8)	74 (3.3)	162 (4.1)	130 (5.7)	196 (4.5)	0 (0.0)	1976 (5.2)
	품질	60 (2.2)	22 (2.4)	238 (2.0)	72 (2.6)	51 (2.0)	102 (2.5)	101 (4.5)	92 (2.3)	29 (1.3)	91 (2.1)	1 (1.9)	859 (2.3)
	합계	2685	903	12049	2799	2559	4041	2269	3954	2264	4345	52	37920

2) 보건의료 및 공공 분야 빅데이터 버즈 현황

〈표 5-10〉과 같이 빅데이터와 관련하여 정부공공 분야가 84.6%, 보건의료 분야는 15.4% 언급된 것으로 나타났다. 빅데이터 기술 수요자의 감성은 52.3%, 공급자의 감성은 47.7%로 나타났다. 기반관련 버즈는 기반구축, 정책전략, 서비스 등의 순으로 나타났으며, 업무는 고객관리마케팅, 생산관리, 컨설팅, 통계 등의 으로 나타났다. 기술관련 버즈는 저장기술, 분석기술, 처리기술 등의 순으로 나타났으며, 국가는 한국, 글러벌, 미국, 유럽, 일본 등의 순으로 언급된 것으로 나타났다. 수요공급에 관한 영향요인으로는 성장, 관심, 효율, 혁신, 우려 등의 순으로 나타났다.

〈표 5-10〉 보건의료 및 공공 분야 빅데이터 버즈 현황

구분	항목	N(%)	구분	항목	N(%)
수용태도	수요자	1,543(52.3)	기술	분석기술	1,236(26.3)
	공급자	1,409(47.7)		처리기술	1,178(25.0)
	계	2,952		저장기술	1,974(42.0)
분야	보건의료	1,092(15.4)		수집기술	19(0.4)
	공공	5,989(84.6)		활용기술	205(4.4)
	계	7,081		기타기술	93(2.0)
기반	프라이버시	1,052(7.8)	업무	계	4,705
	분석전문가	1,102(8.2)		고객관리마케팅	880(26.0)
	정책전략	3,198(23.8)		생산관리	665(19.6)
	기반구축	3,766(28.1)		인사재무	318(9.4)
	비용	736(5.5)		연구개발	308(9.1)
	서비스	2,759(20.6)		의사결정	352(10.4)
	콘텐츠	549(4.1)		컨설팅	435(12.8)
	품질	261(1.9)		통계	429(12.7)
	계	13,423		계	3,387
수요공급 영향요인	성장요인	3,021(23.0)	국가	한국	2,627(29.7)
	관심요인	3,554(27.1)		글로벌	2,278(25.8)
	효율요인	2,261(17.2)		유럽	876(9.9)
	혁신요인	1,833(14.0)		미국	1,548(17.5)
	경제적요인	531(4.0)		일본	729(8.2)
	우려요인	886(6.8)		중국	679(7.7)
	신속요인	717(5.5)		싱가포르	106(1.2)
	신뢰요인	318(2.4)		계	8,843
	계	13,121			

〈표 5-11〉과 같이 보건의료와 공공 분야 빅데이터 기술은 저장기술, 분석기술, 처리기술, 활용기술 등의 순으로 나타났다. 보건의료 분야의 빅데이터 분석기술은 고객관리, 생산관리, 통계, 인사관리, 컨설팅 등의 순으로 활용되고 있으며, 공공 분야의 빅데이터 분석기술은 고객관리, 생산관리, 컨설팅, 의사결정 등에 활용되고 있는 것으로 나타났다. 보건의료 분야의 빅데이터 처리기술은 고객관리, 통계, 의사결정, 생산관리 등

의 순으로 활용되고 있으며, 공공 분야의 빅데이터 처리기술은 고객관리, 생산관리, 컨설팅, 통계 등에 활용되고 있는 것으로 나타났다. 보건의료 분야의 빅데이터 저장기술은 인사관리, 고객관리, 생산관리, 통계 등의 순으로 활용되고 있으며, 공공 분야의 빅데이터 저장기술은 고객관리, 생산관리, 컨설팅, 통계 등에 활용되고 있는 것으로 나타났다. 보건의료 분야의 빅데이터 수집기술은 통계, 고객관리(생산관리) 등의 순으로 활용되고 있으며, 공공 분야의 빅데이터 수집기술은 생산관리, 의사결정에 활용되고 있는 것으로 나타났다. 보건의료 분야의 빅데이터 활용기술은 고객관리, 의사결정, 통계, 컨설팅(생산관리) 등의 순으로 활용되고 있으며, 공공 분야의 빅데이터 활용기술은 고객관리, 컨설팅, 생산관리, 의사결정, 통계 등에 활용되고 있는 것으로 나타났다.

〈표 5-11〉 업무별 빅데이터 기술 활용 현황

N(%)

분야	기술	고객관리	생산관리	인사관리	연구개발	의사결정	컨설팅	통계	합계
보건 의료	분석기술	62 (19.7)	60 (19.1)	51 (16.2)	17 (5.4)	35 (11.1)	37 (11.8)	52 (16.6)	314 (33.2)
	처리기술	52 (23.2)	36 (16.1)	22 (9.8)	16 (7.1)	38 (17.0)	20 (8.9)	40 (17.9)	224 (23.7)
	저장기술	66 (19.8)	52 (15.6)	56 (16.8)	38 (11.4)	38 (11.4)	37 (11.1)	46 (13.8)	333 (35.2)
	수집기술	2 (25.0)	2 (25.0)	-	-	-	-	3 (37.5)	8 (.8)
	활용기술	15 (31.9)	5 (10.6)	3 (6.4)	2 (4.3)	9 (19.1)	5 (10.6)	8 (17.0)	47 (5.0)
	기타기술	4 (20.0)	5 (25.0)	3 (15.0)	-	3 (15.0)	1 (5.0)	4 (20.0)	20 (2.1)
	합계	201	160	135	73	124	100	153	946
정부 공공 기관	분석기술	290 (28.8)	212 (21.1)	84 (8.3)	77 (7.7)	113 (11.2)	121 (12.0)	109 (10.8)	1006 (29.6)
	처리기술	240 (28.1)	178 (20.8)	55 (6.4)	61 (7.1)	102 (11.9)	110 (12.9)	108 (12.6)	854 (25.1)

분야	기술	고객관리	생산관리	인사관리	연구개발	의사결정	컨설팅	통계	합계
	저장기술	332 (26.1)	227 (17.8)	97 (7.6)	135 (10.6)	126 (9.9)	216 (17.0)	141 (11.1)	1274 (37.4)
	수집기술	-	3 (75.0)	-	-	1 (25.0)	-	-	4 (.1)
	활용기술	53 (24.4)	35 (16.1)	14 (6.5)	11 (5.1)	34 (15.7)	40 (18.4)	30 (13.8)	217 (6.4)
	기타기술	13 (27.7)	11 (23.4)	3 (6.4)	4 (8.5)	7 (14.9)	4 (8.5)	5 (10.6)	47 (1.4)
	합계	928	666	253	288	383	491	393	3402

〈표 5-12〉와 같이 보건의료 분야 빅데이터 기반은 기반구축, 서비스, 정책전략, 비용 등의 순으로 나타났으며, 공공 분야는 기반구축, 정책전략, 서비스, 분석전문가 등의 순으로 나타났다. 보건의료 분야의 빅데이터 기반구축은 고객관리, 생산관리, 통계 등의 순으로 나타났으며, 공공 분야의 빅데이터 기반구축은 고객관리, 생산관리, 컨설팅 등의 순으로 나타났다. 보건의료 분야의 빅데이터 분석전문가는 고객관리, 인사관리(통계) 등의 순으로 활용되고 있으며, 공공 분야의 빅데이터 분석전문가는 고객관리, 생산관리, 통계 등에 활용되고 있는 것으로 나타났다. 보건의료 분야의 빅데이터 서비스는 고객관리, 통계, 생산관리 등의 순으로 활용되고 있으며, 공공 분야의 빅데이터 서비스는 고객관리, 생산관리, 컨설팅 등에 활용되고 있는 것으로 나타났다. 보건의료 분야의 빅데이터 프라이버시는 생산관리, 고객관리, 통계 등의 순으로 활용되고 있으며, 공공 분야의 빅데이터 프라이버시는 고객관리, 통계 등에 활용되고 있는 것으로 나타났다.

〈표 5-12〉 업무별 빅데이터 기반 활용 현황

N(%)									
분야	기반	고객관리	생산관리	인사관리	연구개발	의사결정	컨설팅	통계	합계
보건 의료	프라이버시	43 (21.9)	48 (24.5)	14 (7.1)	12 (6.1)	18 (9.2)	22 (11.2)	39 (19.9)	196 (8.0)
	분석전문가	52 (23.0)	27 (11.9)	36 (15.9)	16 (7.1)	29 (12.8)	30 (13.3)	36 (15.9)	226 (9.3)
	정책전략	107 (22.0)	84 (17.2)	63 (12.9)	49 (10.1)	55 (11.3)	51 (10.5)	78 (16.0)	487 (20.0)
	기반구축	146 (24.5)	103 (17.3)	75 (12.6)	54 (9.0)	66 (11.1)	66 (11.1)	87 (14.6)	597 (24.5)
	비용	58 (24.5)	45 (19.0)	34 (14.3)	15 (6.3)	33 (13.9)	21 (8.9)	31 (13.1)	237 (9.7)
	서비스	118 (23.6)	83 (16.6)	62 (12.4)	31 (6.2)	54 (10.8)	67 (13.4)	84 (16.8)	499 (20.5)
	콘텐츠	28 (24.8)	22 (19.5)	11 (9.7)	11 (9.7)	15 (13.3)	14 (12.4)	12 (10.6)	113 (4.6)
	품질	13 (16.0)	17 (21.0)	10 (12.3)	14 (17.3)	8 (9.9)	10 (12.3)	9 (11.1)	81 (3.3)
	계	565	429	305	202	278	281	376	2436
공공	프라이버시	202 (28.9)	153 (21.9)	56 (8.0)	38 (5.4)	65 (9.3)	92 (13.2)	93 (13.3)	699 (7.5)
	분석전문가	252 (26.1)	174 (18.0)	93 (9.6)	83 (8.6)	112 (11.6)	123 (12.7)	128 (13.3)	965 (10.4)
	정책전략	535 (26.0)	426 (20.7)	150 (7.3)	215 (10.4)	231 (11.2)	255 (12.4)	248 (12.0)	2060 (22.1)
	기반구축	588 (24.7)	498 (20.9)	174 (7.3)	230 (9.7)	267 (11.2)	337 (14.2)	285 (12.0)	2379 (25.6)
	비용	170 (27.1)	139 (22.2)	64 (10.2)	47 (7.5)	75 (12.0)	69 (11.0)	63 (10.0)	627 (6.7)
	서비스	484 (26.2)	364 (19.7)	135 (7.3)	161 (8.7)	186 (10.1)	297 (16.1)	217 (11.8)	1844 (19.8)
	콘텐츠	125 (27.4)	93 (20.4)	42 (9.2)	46 (10.1)	39 (8.5)	57 (12.5)	55 (12.0)	457 (4.9)
	품질	59 (21.6)	58 (21.2)	27 (9.9)	32 (11.7)	31 (11.4)	39 (14.3)	27 (9.9)	273 (2.9)
	계	2415	1905	741	852	1006	1269	1116	9304

나. 빅데이터 수요공급에 미치는 영향 요인

1) 전체 분야 빅데이터 수요공급 영향요인

성장과 혁신 요인은 수요자 보다 공급자에 정적(+)인 영향을 미치는 것으로 나타나, 성장과 혁신을 위해 공급자가 빅데이터를 수용할 확률이 높은 것으로 나타났다. 관심, 경제적, 신속 요인은 수요자 보다 공급자에 부정적(-)인 영향을 미치는 것으로 나타나, 관심, 경제적, 신속을 위해 수요자가 빅데이터를 수용할 확률이 높은 것으로 나타났다(표 5-13 참조).

〈표 5-13〉 빅데이터 수요자와 공급자에 미치는 영향요인*

변수	b ^a	S.E. ^b	OR(95%CI) ^c	P
성장요인	.248	.045	1.281(1.174~1.399)	.000
관심요인	-.126	.050	.882(.800~.973)	.012
효율요인	-.069	.045	.933(.854~1.020)	.128
혁신요인	.272	.047	1.313(1.197~1.439)	.000
경제적요인	-.172	.078	.842(.723~.980)	.027
우려요인	-.058	.072	.944(.820~1.086)	.418
신속요인	-.296	.060	.744(.661~.836)	.000
신뢰요인	.006	.088	1.006(.846~1.195)	.950
상수항	-.120	.049	.887	.014

주: * 기본범주: 수요자, ^a Standardized coefficients, ^b Standard error,
^c Adjusted odds ratio(95% Confidence interval)

2) 보건의료 및 공공 분야 빅데이터 수요공급에 미치는 영향 요인

보건의료분야는 혁신요인과 우려요인에 정적(+) 영향을 미치는 것으로 나타나, 혁신과 우려로 인해 공급자가 빅데이터를 수용할 확률이 높은 것으로 나타났다. 반면, 성장요인은 부적(-) 영향을 미치는 것으로 나타나, 성장으로 인해 공급자보다 수요자가 빅데이터를 수용할 확률이 높은 것으로 나타났다. 공공분야는 혁신요인과 성장요인에 정적(+) 영향을 미치는 것으로 나타나, 혁신과 성장으로 인해 수요자 보다 공급자가 빅데이터를 수용할 확률이 높은 것으로 나타났다. 반면, 신속요인은 부적(-) 영향을 미치는 것으로 나타나, 신속요인으로 인해 공급자보다 수요자가 빅데이터를 수용할 확률이 높은 것으로 나타났다(표 5-14 참조).

〈표 5-14〉 보건의료 및 공공 분야 빅데이터 수요자와 공급자에 미치는 영향요인*

변수	보건의료 분야				공공 분야			
	b ^a	S.E. ^b	OR ^c	P	b ^a	S.E. ^b	OR ^c	P
성장요인	-.394	.182**	.674	.030	.342	.084***	1.408	.000
관심요인	.192	.218	1.211	.381	-.024	.091	.976	.789
효율요인	-.273	.175	.761	.119	-.079	.081	.924	.327
혁신요인	.470	.178***	1.600	.008	.410	.084***	1.507	.000
경제적요인	-.031	.258	.969	.904	-.105	.126	.900	.405
우려요인	.456	.261*	1.577	.081	-.098	.103	.907	.341
신속요인	.152	.220	1.164	.490	-.420	.110***	.657	.000
신뢰요인	-.225	.284	.799	.429	.139	.148	1.149	.349
상수항	-.177	.222	.838	.424	-.328	.093*	.720	.000

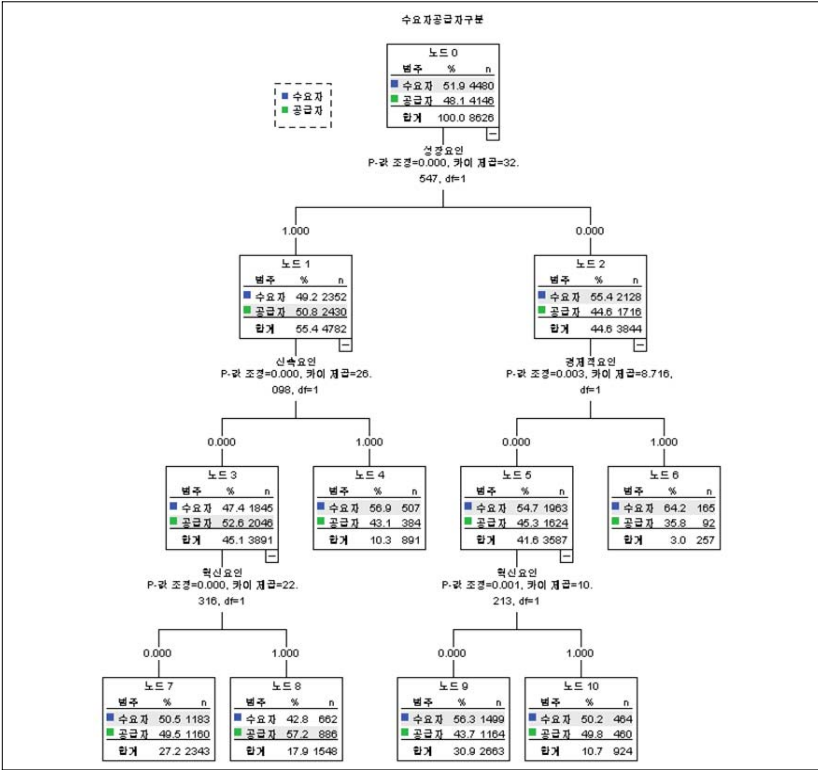
주: * 기본범주: 수요자, ^a Standardized coefficients, ^b Standard error, ^c Adjusted odds ratio
 *** p< .01, ** p< .05, * p< .1

다. 빅데이터 수요공급 예측모형

1) 전체 분야 빅데이터 수요공급 예측모형

본 연구에서는 감성요인, 기반요인, 업무요인이 국내의 빅데이터 수요 공급 예측모형에 미치는 영향에 대해 살펴보았다. 감성요인이 빅데이터 수요공급 예측모형에 미치는 영향은 [그림 5-2]과 같다. 나무구조의 최상위에 있는 네모는 루트노드로서, 예측변수(독립변수)가 투입되지 않은 종속변수(수요자, 공급자)의 빈도를 나타낸다. 루트노드에서 빅데이터 수요는 51.9%(4,480건), 공급은 48.1%(4,146건)으로 나타났다. 루트노드의 하단의 가장 상위에 위치하는 요인이 빅데이터 수요공급 예측에 가장 영향력이 높은(관련성이 깊은)요인으로 ‘성장요인’의 영향력이 가장 큰 것으로 나타났다. ‘성장요인’이 높을 경우 빅데이터 수요가 이전의 51.9%에서 49.2%로 감소한 반면, 빅데이터 공급은 이전의 48.1%에서 50.8%로 증가하였다. ‘성장요인’이 많고 ‘신속요인’이 높은 경우 빅데이터 수요가 이전의 49.2%에서 56.9%로 증가한 반면, 빅데이터 공급은 이전의 50.8%에서 43.1%로 감소하였다. <표 5-15>의 감성요인의 수요공급에 대한 이익도표와 같이 빅데이터 수요에 가장 영향을 높은 경우는 ‘성장요인’이 적고 ‘경제적요인’이 많은 조합으로 나타났다. 즉, 6번 노드의 지수(index)가 123.6%로 뿌리마디와 비교했을 때 6번 노드의 조건을 가진 집단이 빅데이터 수요자일 확률이 약 1.24배로 나타났다. 빅데이터 공급에 가장 영향을 높은 경우는 ‘성장요인’이 높고 ‘신속요인’이 낮고, ‘혁신요인’이 높은 조합으로 나타났다. 즉, 8번 노드의 지수(index)가 119.1%로 뿌리마디와 비교했을 때 8번 노드의 조건을 가진 집단이 빅데이터 공급자일 확률이 약 1.19배로 나타났다.

[그림 5-2] 빅데이터 수요공급 예측모형(감성요인)



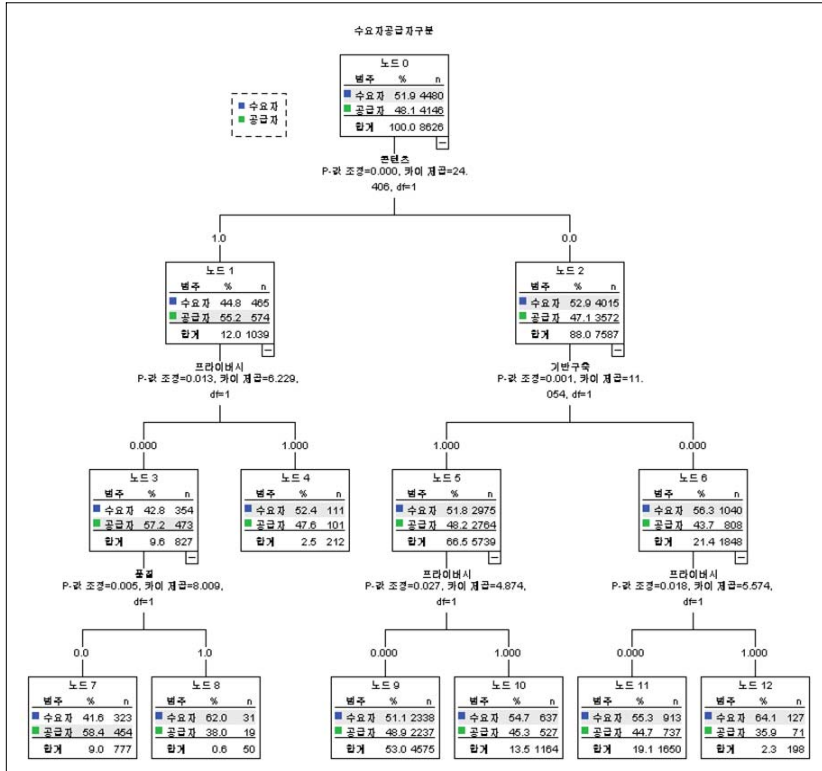
<표 5-15> 빅데이터 수요공급 예측모형에 대한 이익도표(감성요인)

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
수요자	6	257	3.0	3.7	123.6	165	3.7	64.2	123.6
	4	891	10.3	11.3	109.6	672	15.0	58.5	112.7
	9	2663	30.9	33.5	108.4	2171	48.5	57.0	109.7
	7	2343	27.2	26.4	97.2	3354	74.9	54.5	104.9
	10	924	10.7	10.4	96.7	3818	85.2	53.9	103.9
	8	1548	17.9	14.8	82.3	4480	100.0	51.9	100.0

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
공급자	8	1548	17.9	21.4	119.1	886	21.4	57.2	119.1
	10	924	10.7	11.1	103.6	1346	32.5	54.4	113.3
	7	2343	27.2	28.0	103.0	2506	60.4	52.0	108.3
	9	2663	30.9	28.1	90.9	3670	88.5	49.1	102.1
	4	891	10.3	9.3	89.7	4054	97.8	48.4	100.8
	6	257	3.0	2.2	74.5	4146	100.0	48.1	100.0

기반요인이 빅데이터 수요공급 예측모형에 미치는 영향은 [그림 5-3]과 같다. 기반요인에서 빅데이터 수요공급 예측에 ‘콘텐츠요인’의 영향력이 가장 큰 것으로 나타났다. ‘콘텐츠요인’이 높을 경우 빅데이터 수요가 이전의 51.9%에서 44.8%로 감소한 반면, 빅데이터 공급은 이전의 48.1%에서 55.2%로 증가하였다. ‘콘텐츠요인’이 높고 ‘프라이버시요인’이 높은 경우 빅데이터 수요가 이전의 44.8%에서 52.4%로 증가한 반면, 빅데이터 공급은 이전의 55.2%에서 47.6% 감소하였다. <표 5-16>의 기반요인의 수요공급에 대한 이익도표와 같이 빅데이터 수요에 가장 영향을 높은 경우는 ‘콘텐츠요인’이 낮고 ‘기반구축요인’이 낮으며 ‘프라이버시요인’이 높은 조합으로 나타났다. 즉, 12번 노드의 지수가 123.5%로 뿌리마디와 비교했을 때 12번 노드의 조건을 가진 집단이 빅데이터 수요자일 확률이 약 1.24배로 나타났다. 빅데이터 공급에 가장 영향을 높은 경우는 ‘콘텐츠요인’이 높고 ‘프라이버시요인’이 낮고, ‘품질요인’이 낮은 조합으로 나타났다. 즉, 7번 노드의 지수가 121.6%로 뿌리마디와 비교했을 때 7번 노드의 조건을 가진 집단이 빅데이터 공급자일 확률이 약 1.22배로 나타났다.

[그림 5-3] 빅데이터 수요공급 예측모형(기반요인)



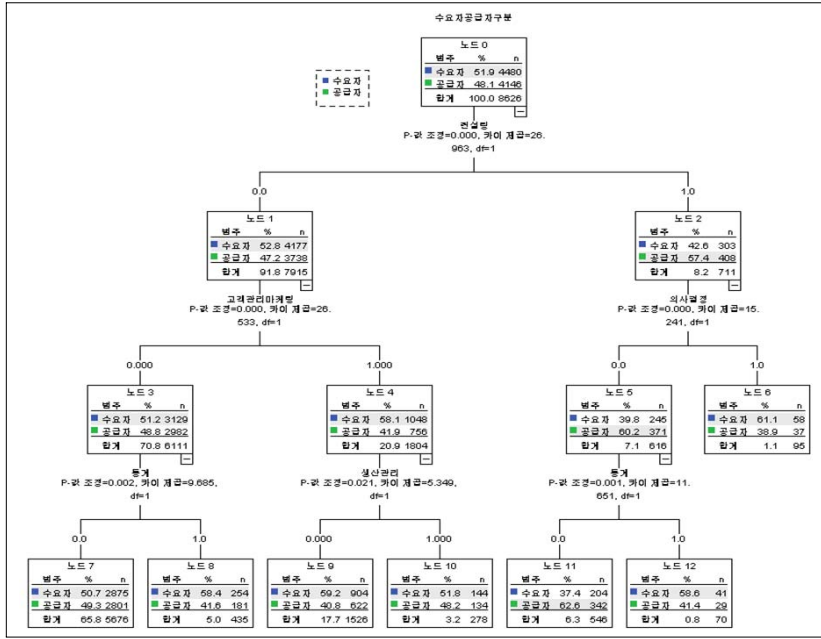
<표 5-16> 빅데이터 수요공급 예측모형에 대한 이익도표(기반요인)

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
수요자	12	198	2.3	2.8	123.5	127	2.8	64.1	123.5
	8	50	.6	.7	119.4	158	3.5	63.7	122.7
	11	1650	19.1	20.4	106.5	1071	23.9	56.4	108.6
	10	1164	13.5	14.2	105.4	1708	38.1	55.8	107.4
	4	212	2.5	2.5	100.8	1819	40.6	55.6	107.0
	9	4575	53.0	52.2	98.4	4157	92.8	53.0	102.0
	7	777	9.0	7.2	80.0	4480	100.0	51.9	100.0

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
공급자	7	777	9.0	11.0	121.6	454	11.0	58.4	121.6
	9	4575	53.0	54.0	101.7	2691	64.9	50.3	104.6
	4	212	2.5	2.4	99.1	2792	67.3	50.2	104.4
	10	1164	13.5	12.7	94.2	3319	80.1	49.3	102.6
	11	1650	19.1	17.8	92.9	4056	97.8	48.4	100.7
	8	50	.6	.5	79.1	4075	98.3	48.4	100.6
	12	198	2.3	1.7	74.6	4146	100.0	48.1	100.0

업무요인이 빅데이터 수요공급 예측모형에 미치는 영향은 [그림 5-4]와 같다. 업무요인에서 빅데이터 수요공급 예측에는 ‘컨설팅요인’의 영향력이 가장 큰 것으로 나타났다. ‘컨설팅요인’이 높을 경우 빅데이터 수요가 이전의 51.9%에서 42.6%로 감소한 반면, 빅데이터 공급은 이전의 48.1%에서 57.4%로 증가하였다. ‘컨설팅요인’이 높고 ‘의사결정요인’이 높은 경우 빅데이터 수요가 이전의 42.6%에서 61.1%로 증가한 반면, 빅데이터 공급은 이전의 57.4%에서 38.9% 감소하였다. <표 5-17>의 업무요인의 수요공급에 대한 이익도표와 같이 빅데이터 수요에 가장 영향을 높은 경우는 ‘컨설팅요인’이 높고 ‘의사결정요인’이 높은 조합으로 나타났다. 즉, 6번 노드의 지수가 117.6%로 뿌리마디와 비교했을 때 6번 노드의 조건을 가진 집단이 빅데이터 수요자일 확률이 약 1.18배로 나타났다. 빅데이터 공급에 가장 영향을 높은 경우는 ‘컨설팅요인’이 높고 ‘의사결정요인’이 낮고, ‘통계요인’이 낮은 조합으로 나타났다. 즉, 11번 노드의 지수가 130.3%로 뿌리마디와 비교했을 때 11번 노드의 조건을 가진 집단이 빅데이터 공급자일 확률이 약 1.30배로 나타났다.

[그림 5-4] 빅데이터 수요공급 예측모형(업무요인)



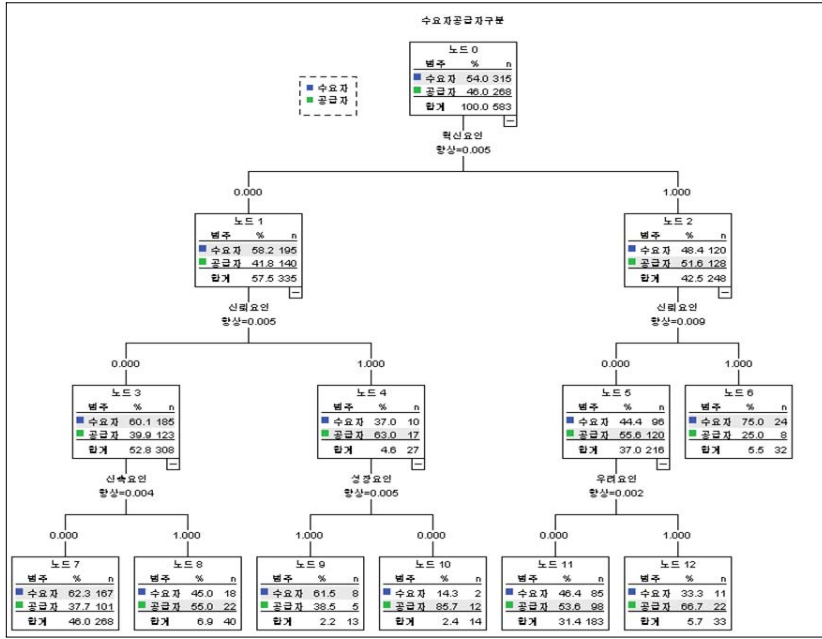
<표 5-17> 빅데이터 수요공급 예측모형에 대한 이익도표(업무요인)

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
수요자	6	95	1.1	1.3	117.6	58	1.3	61.1	117.6
	9	1526	17.7	20.2	114.1	962	21.5	59.3	114.3
	12	70	.8	.9	112.8	1003	22.4	59.3	114.2
	8	435	5.0	5.7	112.4	1257	28.1	59.1	113.8
	10	278	3.2	3.2	99.7	1401	31.3	58.3	112.2
	7	5676	65.8	64.2	97.5	4276	95.4	52.9	101.9
	11	546	6.3	4.6	71.9	4480	100.0	51.9	100.0
공급자	11	546	6.3	8.2	130.3	342	8.2	62.6	130.3
	7	5676	65.8	67.6	102.7	3143	75.8	50.5	105.1
	10	278	3.2	3.2	100.3	3277	79.0	50.4	104.9
	8	435	5.0	4.4	86.6	3458	83.4	49.9	103.7
	12	70	.8	.7	86.2	3487	84.1	49.8	103.6
	9	1526	17.7	15.0	84.8	4109	99.1	48.2	100.2
	6	95	1.1	.9	81.0	4146	100.0	48.1	100.0

2) 보건의료 및 공공 분야 빅데이터 수요공급 예측모형

본 연구에서는 감성요인과 기반요인이 보건의료 및 공공 분야의 빅데이터 수요공급 예측모형에 미치는 영향에 대해 살펴보았다. 보건의료분야의 감성요인이 빅데이터 수요공급 예측모형에 미치는 영향은 [그림 5-5]와 같다. 나무구조의 최상위에 있는 네모는 루트노드로서, 예측변수(독립변수)가 투입되지 않은 종속변수(수요자, 공급자)의 빈도를 나타낸다. 루트노드에서 빅데이터 수요는 54.0%(315건), 공급은 46.0%(268건)으로 나타났다. 루트노드의 하단의 가장 상위에 위치하는 요인이 빅데이터 수요공급 예측에 가장 영향력이 높은(관련성이 깊은)요인으로 ‘혁신요인’의 영향력이 가장 큰 것으로 나타났다. ‘혁신요인’이 높을 경우 빅데이터 수요가 이전의 54.0%에서 48.4%로 감소한 반면, 빅데이터 공급은 이전의 46.0%에서 51.6%로 증가하였다. ‘혁신요인’이 많고 ‘신뢰요인’이 높은 경우 빅데이터 수요가 이전의 48.4%에서 75.0%로 크게 증가한 반면, 빅데이터 공급은 이전의 51.6%에서 25.0%로 크게 감소하였다. <표 5-18>의 감성요인의 수요공급에 대한 이익도표와 같이 빅데이터 수요에 가장 영향을 높은 경우는 ‘혁신요인’이 많고 ‘신뢰요인’이 많은 조합으로 나타났다. 즉, 6번 노드의 지수(index)가 138.8%로 뿌리마디와 비교했을 때 6번 노드의 조건을 가진 집단이 빅데이터 수요자일 확률이 약 1.39배로 나타났다. 빅데이터 공급에 가장 영향을 높은 경우는 ‘혁신요인’이 적고 ‘신뢰요인’이 많고, ‘성장요인’이 적은 조합으로 나타났다. 즉, 10번 노드의 지수(index)가 186.5%로 뿌리마디와 비교했을 때 10번 노드의 조건을 가진 집단이 빅데이터 공급자일 확률이 약 1.87배로 나타났다.

[그림 5-5] 보건의료분야 빅데이터 수요공급 예측모형(감성요인)



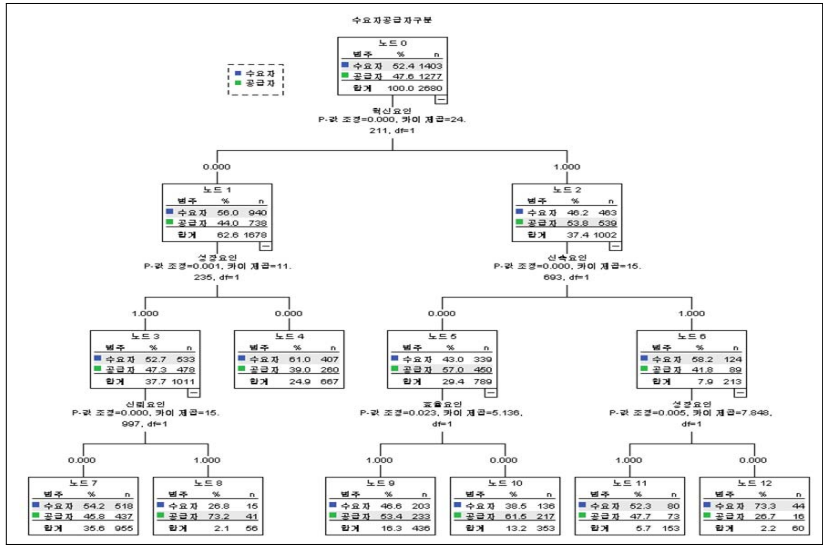
<표 5-18> 보건의료분야 빅데이터 수요공급 예측모형에 대한 이익도표(감성요인)

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
수요자	6	32	5.5	7.6	138.8	32	5.5	7.6	138.8
	7	268	46.0	53.0	115.3	300	51.5	60.6	117.8
	9	13	2.2	2.5	113.9	313	53.7	63.2	117.7
	11	183	31.4	27.0	86.0	496	85.1	90.2	106.0
	8	40	6.9	5.7	83.3	536	91.9	95.9	104.3
	12	33	5.7	3.5	61.7	569	97.6	99.4	101.8
	10	14	2.4	.6	26.4	583	100.0	100.0	100.0
공급자	10	14	2.4	4.5	186.5	14	2.4	4.5	186.5
	12	33	5.7	8.2	145.0	47	8.1	12.7	157.4
	8	40	6.9	8.2	119.6	87	14.9	20.9	140.0
	11	183	31.4	36.6	116.5	270	46.3	57.5	124.1
	9	13	2.2	1.9	83.7	283	48.5	59.3	122.2
	7	268	46.0	37.7	82.0	551	94.5	97.0	102.6
	6	32	5.5	3.0	54.4	583	100.0	100.0	100.0

공공분야의 감성요인이 빅데이터 수요공급 예측모형에 미치는 영향은 [그림 5-6]과 같다. 빅데이터 수요공급 예측에 ‘혁신요인’의 영향력이 가장 큰 것으로 나타났다. ‘혁신요인’이 높을 경우 빅데이터 수요가 이전의 52.4%에서 46.2%로 감소한 반면, 빅데이터 공급은 이전의 47.6%에서 53.8%로 증가하였다. ‘혁신요인’이 많고 ‘신속요인’이 많은 경우 빅데이터 수요가 이전의 46.2%에서 59.2%로 증가한 반면, 빅데이터 공급은 이전의 53.8%에서 41.8%로 감소하였다. <표 5-19>의 감성요인의 수요공급에 대한 이익도표와 같이 빅데이터 수요에 가장 영향을 높은 경우는 ‘혁신요인’이 많고 ‘신속요인’이 많고 ‘성장요인’이 적은 조합으로 나타났다. 즉, 12번 노드의 지수(index)가 140.1%로 뿌리마디와 비교했을 때 12번 노드의 조건을 가진 집단이 빅데이터 수요자일 확률이 약 1.40배로 나타났다. 빅데이터 공급에 가장 영향을 높은 경우는 ‘혁신요인’이 적고 ‘성장요인’이 많고, ‘실패요인’이 적은 조합으로 나타났다. 즉, 8번 노드의 지수(index)가 153.7%로 뿌리마디와 비교했을 때 8번 노드의 조건을 가진 집단이 빅데이터 공급자일 확률이 약 1.54배로 나타났다. 보건의료분야의 기반요인이 빅데이터 수요공급 예측모형에 미치는 영향은 [그림 5-7]과 같다. 빅데이터 수요공급 예측에 ‘품질요인’의 영향력이 가장 큰 것으로 나타났다. ‘품질요인’이 많을 경우 빅데이터 수요가 이전의 54.0%에서 40.8%로 감소한 반면, 빅데이터 공급은 이전의 46.0%에서 59.2%로 증가하였다. ‘품질요인’이 많고 ‘비용요인’이 높은 경우 빅데이터 수요가 이전의 40.8%에서 20.0%로 크게 감소한 반면, 빅데이터 공급은 이전의 59.2%에서 80.0%로 크게 증가하였다. <표 5-20>의 기반요인의 수요공급에 대한 이익도표와 같이 빅데이터 수요에 가장 영향을 높은 경우는 ‘품질요인’이 많고 ‘비용요인’이 낮고 ‘프라이버시요인’이 높은 조합으로 나타났다. 즉, 12번 노드의 지수(index)가 156.6%로 뿌리마디와 비교했을

때 12번 노드의 조건을 가진 집단이 빅데이터 수요자일 확률이 약 1.57배로 나타났다. 빅데이터 공급에 가장 영향을 높은 경우는 ‘품질요인’이 적고 ‘비용요인’이 많고, ‘정책전략요인’이 적은 조합으로 나타났다. 즉, 13번 노드의 지수(index)가 217.5%로 뿌리마디와 비교했을 때 13번 노드의 조건을 가진 집단이 빅데이터 공급자일 확률이 약 2.18배로 나타났다.

[그림 5-6] 공공분야 빅데이터 수요공급 예측모형(감성요인)



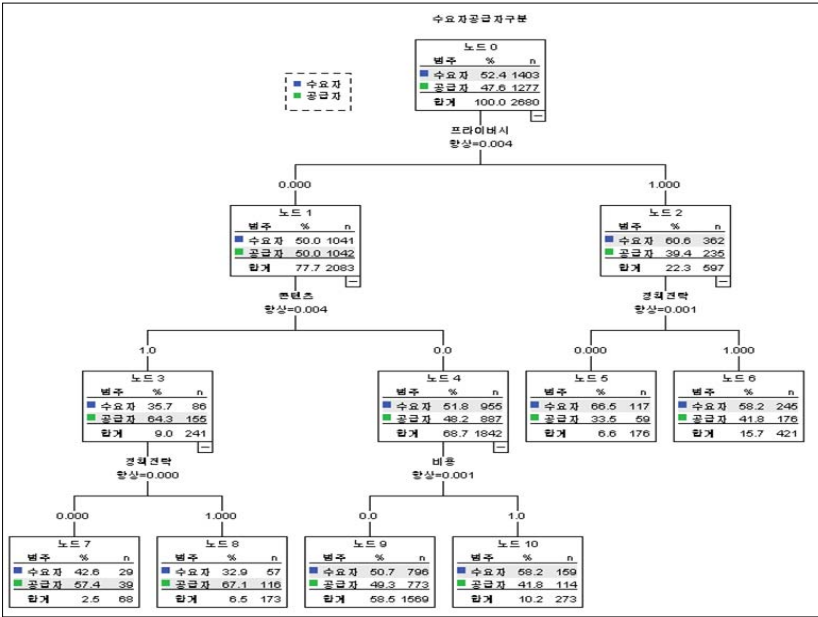
<표 5-19> 공공분야 빅데이터 수요공급 예측모형에 대한 이익도표(감성요인)

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
수요자	12	60	2.2	3.1	140.1	60	2.2	3.1	140.1
	4	667	24.9	29.0	116.6	727	27.1	32.1	118.5
	7	955	35.6	36.9	103.6	1682	62.8	69.1	110.0
	11	153	5.7	5.7	99.9	1835	68.5	74.8	109.2
	9	436	16.3	14.5	88.9	2271	84.7	89.2	105.3
	10	353	13.2	9.7	73.6	2624	97.9	98.9	101.0
	8	56	2.1	1.1	51.2	2680	100.0	100.0	100.0

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
	14	14	2.4	1.3	52.9	577	99.0	100.0	101.0
	13	6	1.0	0.0	0.0	583	100.0	100.0	100.0
공급자	13	6	1.0	2.2	217.5	6	1.0	2.2	217.5
	14	14	2.4	3.7	155.4	20	3.4	6.0	174.0
	11	16	2.7	4.1	149.6	36	6.2	10.1	163.2
	9	16	2.7	3.0	108.8	52	8.9	13.1	146.4
	7	331	56.8	60.4	106.5	383	65.7	73.5	111.9
	8	80	13.7	11.6	84.3	463	79.4	85.1	107.1
	10	107	18.4	14.2	77.3	570	97.8	99.3	101.5
	12	13	2.2	.7	33.5	583	100.0	100.0	100.0

공공분야의 기반요인이 빅데이터 수요공급 예측모형에 미치는 영향은 [그림 5-8]과 같다. 빅데이터 수요공급 예측에 ‘프라이버시요인’의 영향력이 가장 큰 것으로 나타났다. ‘프라이버시요인’이 높을 경우 빅데이터 수요가 이전의 52.4%에서 60.6%으로 증가한 반면, 빅데이터 공급은 이전의 47.6%에서 39.4%로 감소하였다. ‘프라이버시요인’이 많고 ‘정책전략요인’이 많은 경우 빅데이터 수요가 이전의 60.62%에서 58.2%로 감소한 반면, 빅데이터 공급은 이전의 39.4%에서 41.8%로 증가하였다. <표 5-21>의 기반요인의 수요공급에 대한 이익도표와 같이 빅데이터 수요에 가장 영향을 높은 경우는 ‘프라이버시요인’이 많고 ‘정책전략요인’이 적은 조합으로 나타났다. 즉, 5번 노드의 지수(index)가 127.0%로 뿌리마디와 비교했을 때 5번 노드의 조건을 가진 집단이 빅데이터 수요자일 확률이 약 1.27배로 나타났다. 빅데이터 공급에 가장 영향을 높은 경우는 ‘프라이버시요인’이 적고 ‘콘텐츠요인’이 많고, ‘정책전략요인’이 많은 조합으로 나타났다. 즉, 8번 노드의 지수(index)가 140.7%로 뿌리마디와 비교했을 때 8번 노드의 조건을 가진 집단이 빅데이터 공급자일 확률이 약 1.41배로 나타났다.

[그림 5-8] 공공분야 빅데이터 수요공급 예측모형(기반요인)



[표 5-21] 공공분야 빅데이터 수요공급 예측모형에 대한 이익도표(기반요인)

구분	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
수요자	5	176	6.6	8.3	127.0	176	6.6	8.3	127.0
	10	273	10.2	11.3	111.3	449	16.8	19.7	117.4
	6	421	15.7	17.5	111.2	870	32.5	37.1	114.4
	9	1569	58.5	56.7	96.9	2439	91.0	93.9	103.1
	7	68	2.5	2.1	81.5	2507	93.5	95.9	102.6
	8	173	6.5	4.1	62.9	2680	100.0	100.0	100.0
공급자	8	173	6.5	9.1	140.7	173	6.5	9.1	140.7
	7	68	2.5	3.1	120.4	241	9.0	12.1	135.0
	9	1569	58.5	60.5	103.4	1810	67.5	72.7	107.6
	6	421	15.7	13.8	87.7	2231	83.2	86.5	103.9
	10	273	10.2	8.9	87.6	2504	93.4	95.4	102.1
	5	176	6.6	4.6	70.4	2680	100.0	100.0	100.0

제3절 보건복지 정책 수요예측¹⁴⁾

1. 서론

새정부 출범이후 2013년부터 건강보험 보장성 강화를 위한 국민적 요구가 증대하고 고령화로 진료비 지출증가 등이 예상(신현웅, 2013)됨에 따라 진료비 지출을 억제하고, 건강보험의 안정적 운영을 위한 정책개발이 요구되고 있다. 특히, 인구고령화가 사회적 문제로 대두됨에 따라 전체 인구의 고령화와 함께 장애인구 역시 고령화되는 현상을 보임에 따라 노인과 장애인을 위한 정책 수요는 증가할 것으로 보인다(김성희, 2013). 그리고 고령화로 인한 노인인구의 양적 증가와 특성의 다양화로 인한 보건복지 서비스에 대한 관심의 증가와 함께 향후 노인의 보건복지서비스에 대한 욕구도 폭발적으로 늘어날 것으로 예측하고 있다(이윤경 외, 2013). 그동안 정부의 보육정책은 저소득층 위주의 경제적 부담완화에 초점을 맞추어 왔으나 새정부 출범과 함께 전 계층 무상보육과 양육수당 지급정책이 도입됨에 따라 대상자들의 정책 체감도는 상당히 높아질 것으로 예견되고 있다(김은정, 2013). 또한 2014년 7월부터 상대적으로 형편이 어려운 노인에게 차등 지급하는 기초연금에 관한 정책이 결정되어 많은 국민이 공적연금의 혜택을 받을 수 있게 되었다. 그러나 영리병원을 비롯한 의료민영화 정책과 원격의료 정책은 아직 정부와 의료계의 합의가 이루어지지 못하고 있다. 이와 같이 대상자별, 분야별로 다양한 보건복지 정책이 요구됨에 따라 정부는 보건복지 정책 수요를 사전에 예측하여 보건복지 욕구를 충족시키기 위한 노력이 필요하다.

14) 본 절은 본 과제의 일환으로 수행되었으며 보건복지포럼(2014년 7월) '소셜 빅데이터를 활용한 보건복지정책 동향 분석'에 게재된 내용임을 밝힌다.

한편, 스마트기기의 급속한 보급과 소셜미디어의 확산으로 데이터량이 기하급수적으로 증가하고 데이터의 생산, 유통, 소비 체계에 큰 변화를 주면서 데이터가 경제적 자산이 될 수 있는 빅데이터¹⁵⁾ 시대를 맞이하게 되었다¹⁶⁾. 세계 각국의 정부와 기업들은 빅데이터가 향후 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하고 있으며, McKinsey, The Economist, Gartner 등은 빅데이터를 활용한 시장변동 예측과 신사업 발굴 등 경제적 가치창출 사례 및 효과를 제시하고 있다¹⁷⁾. 특히, 많은 국가와 기업에서는 SNS를 통하여 생산되는 소셜 빅데이터의 활용과 분석을 통하여 새로운 경제적 효과와 일자리 창출은 물론 사회적 문제의 해결을 위하여 적극적으로 노력하고 있다. 한국은 최근 정부 3.0¹⁸⁾과 창조경제의 추진과 실현을 위하여 모든 분야에 빅데이터의 효율적 활용을 적극적으로 모색하고 있다. 이와 같이 많은 국가에서 빅데이터가 공공과 민간에 미치는 파급효과를 전망함에 따라 빅데이터의 활용은 정부의 정책을 효율적으로 추진하기 위한 새로운 동력이 될 것으로 보여진다. 국민이 요구하는 보건복지 정책 수요를 분석하기 위해서는 다양한 산업의 종사자나 일반인을 대상으로 설문조사를 실시해야 한다. 그러나 기존에 실시하던 횡단적 조사나 종단적 조사 등을 대상으로 한 연구는 정해진 변인들에 대한 개인과 집단의 관계를 보는 데에는 유용하나, 사이버 상에서 언급된 개인별 담론(buzz)에서 논의된 관련 변인의 상호간의 연관관계를 밝히고 원인을 파악하는 데는 한계가 있다. 따라서 본 연구는

15) 빅데이터란 기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다(위키백과, 2014. 5. 24).

16) 송태민(2012). 보건복지 빅데이터 효율적 활용방안. 보건복지포럼, 통권 제193호.

17) 송태민(2013). 우리나라 보건복지 빅데이터 동향 및 활용방안. 과학기술정책, 통권 제192호.

18) 정부 3.0이란 공공정보를 적극 개방·공유하고, 부처 간 칸막이를 없애 소통·협력함으로써 국민 맞춤형 서비스를 제공하고, 일자리 창출과 창조 경제를 지원하는 새로운 정부 운영 패러다임을 의미함

2014년 1월과 2월 우리나라 온라인 뉴스사이트, 블로그, 카페, SNS, 게시판 등에서 수집한 소셜 빅데이터를 바탕으로 보건복지 수요를 예측하고자 한다.

2. 연구방법

가. 연구대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅 데이터를 대상으로 하였다. 본 분석에서는 116개의 온라인 뉴스사이트, 4개의 블로그(네이버, 네이버, 다음, 티스토리), 2개의 카페(네이버, 다음), 2개의 SNS(트위터, 미투데이), 4개의 게시판(네이버지식인, 네이버지식, 네이버톡, 네이버판) 등 총 128개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹문서(버즈)를 소셜 빅데이터로 정의하였다. 보건복지 관련 토픽의 수집은 2014. 1. 1~2014. 2. 28까지 해당 채널에서 요일별, 주말, 휴일을 고려하지 않고 매 시간단위로 수집하였으며, 수집된 총 111,596건(1월: 57,830건, 2월: 53,766건)¹⁹⁾의 텍스트(Text) 문서를 본 연구의 분석에 포함시켰다. 본 연구를 위한 소셜 빅데이터의 수집²⁰⁾은 크롤러(Crawler)를 사용하였고, 토픽의 분류는 주제분석(text mining) 기법을 사용하였다. 보건복지 토픽은 모든 관련 문서를 수집하기 위해 '보건', '복지', 그리고 '보건복지'를 사용하였다.

19) 본 연구에서 수집된 문서는 SNS 77.7%(트위터: 75.9%(84,721건), 미투데이: 1.8%(2,010건), 온라인 뉴스사이트 7.4%(8,345건), 카페 6.8%(7,508건), 블로그 5.9%(6,704건), 게시판 2.2%(2,317건) 순으로 나타났다.

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

나. 연구도구

보건복지와 관련하여 수집된 버즈는 주제분석(text mining)과 요인분석(factor analysis)의 과정을 거쳐 다음과 같이 정형화 데이터로 코드화하여 사용하였다.

1) 보건복지 관련 수요

본 연구의 종속변수인 보건복지 수요(찬성, 반대)의 정의는 주제분석과 요인분석의 과정을 거쳐 ‘운영, 지원, 계획, 예정, 강화, 실시, 확대, 진행, 이용, 사용, 도입, 추진, 참여’은 찬성의 감정으로, ‘문제, 지적, 반대, 거짓말, 논란, 비판, 걱정, 억울, 외면’은 반대의 감정으로 정의하였다.

2) 보건복지 관련 정책

보건복지 관련 정책의 정의는 주제분석과 요인분석의 과정을 거쳐 ‘연금(기초연금, 국민연금), 기초생활(최저생계비, 기초생활보장, 기초생활수급권), 양육수당(양육수당, 보조금, 맞춤형), 무상정책(무상보육, 무상의료, 무상급식), 의료민영화(포괄수가제, 의료상업화, 민영화, 의료민영화), 건강보험(국민건강보험, 민간보험, 건강보험), 원격의료(법인약국, 의료영리화, 원격의료, 원격의료), 전세대책(전세대책, 서민정책), 반값등록금(반값등록금, 4대강), 퇴직연금(희망누리, 퇴직연금, 산재보험), 증세(증세, 세법개정안, 부자증세, 부자감세), 과다복지(방만경영, 과다복지, 복리후생비, 복지카드), 행복온도, 육아휴직(출산휴가, 보건휴가, 육아휴직), 중독법(중독법, 게임중독법)’의 15개 정책으로 해당 정책이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

3) 보건복지 관련 공공기관

보건복지 관련 공공기관의 정의는 주제분석과 요인분석의 과정을 거쳐 ‘국회(보건복지위원회, 국회), 보건복지부(보건복지부, 질병관리본부), 청와대, 고용여성부처(고용노동부, 여성가족부), 공기업(공기업, 공공기관, 근로복지공단, 근로복지넷), 교육농림부처(교육부, 농림축산식품부), 지자체(보건소, 주민센터)’의 7개 기관으로 정의하였다. 정의된 모든 공공기관은 해당 공공기관이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

4) 보건복지 관련 민간기관

보건복지 관련 민간기관의 정의는 주제분석과 요인분석의 과정을 거쳐 ‘요양병원(인증요양병원, 요양병원, 의료기관), 관련협회(약사회, 의사협회, 보건의료노조, 보건의료단체, 대한의사협회), 기업(대기업, 중소기업, 회사, 기업, 삼성, 복지시설), 대학(대학, 학교), 시민단체(시민단체, 병원), 사회복지시설(장애인복지시설, 아동복지시설, 사회복지시설, 장애인복지단체)’의 6개 기관으로 정의하였다. 정의된 모든 민간기관은 해당 민간기관이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

5) 보건복지 관련 대상

보건복지 관련 대상의 정의는 주제분석과 요인분석의 과정을 거쳐 ‘가족(모녀, 부자, 가족, 할머니, 자식들, 아이), 청소년(아동, 청소년), 환자, 의사, 저소득층(저소득층, 취약계층, 기초생활수급자, 소외계층, 빈곤층),

장애인(장애인, 노약자), 여성, 노동자(노동자, 노동자들, 근로자), 서민(서민, 서민들), 중산층, 노인(노인들, 노인, 어르신들, 어르신), 피해자, 공무원(공무원, 군인), 비정규직, 학생(대학생, 학생, 청년), 노숙자(노숙자, 노숙자들), 외국인'의 17개 대상으로 정의하였다. 정의된 모든 대상은 해당 대상이 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

6) 보건복지 관련 분야

보건복지 관련 분야의 정의는 주제분석의 과정을 거쳐 '주거(주거복지, 주거안정), 교육, 사회복지, 보건의료, 경제, 문화, 환경, 통일, 가정, 노동, 보육, 범죄, 안보, 다문화'의 14개로 정의하였다. 정의된 모든 분야는 해당 분야가 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

7) 보건복지 관련 주요이슈

보건복지 관련 주요이슈의 정의는 주제분석과 요인분석의 과정을 거쳐 '중증질환, 의료비, 담배, 건강, 일자리(일자리, 취업), 의료수가, 자살, 등록금(등록금, 학자금), 세금, 행복지수(행복지수, 국민행복지수), 개인 정보, 부동산, 정규직, 위안부, 결혼, 출산율, 양극화, 성폭행, 반려동물'의 19개 이슈로 정의하였다. 정의된 모든 이슈는 해당 대상이 있는 경우는 '1', 없는 경우는 '0'으로 코드화 하였다.

다. 분석 방법

본 연구에서는 한국의 보건복지 정책 수요를 설명하는 가장 효율적인 예측모형을 구축하기 위해 특별한 통계적 가정이 필요하지 않은 데이터

마이닝의 의사결정나무 분석방법을 사용하였다. 데이터마이닝의 의사결정나무 분석은 방대한 자료 속에서 종속변인을 가장 잘 설명하는 예측모형을 자동적으로 산출해 줌으로써 각기 다른 속성은 가진 보건복지 정책스요에 대한 요인을 쉽게 파악할 수 있다. 본 연구의 의사결정나무 형성을 위한 분석 알고리즘은 CHAID(Chi-squared Automatic Interaction Detection)²¹⁾를 사용하였다. CHAID는 이산형인 종속변수의 분리기준으로 카이제곱(χ^2 -검정)을 사용하며, 모든 가능한 조합을 탐색하여 최적분리를 찾는다. 정지규칙(stopping rule)으로 관찰치가 충분하여 상위노드(부모마디)의 최소케이스 수는 100으로 하위노드(자식마디)의 최소 케이스 수는 50으로 설정하였고, 나무깊이는 3수준으로 정하였다. 본 연구의 기술분석, 다중응답분석, 로지스틱 회귀분석, 의사결정나무분석은 SPSS 22.0을 사용하였다.

3. 연구 결과

가. 보건복지 관련 분야 버즈 현황

보건복지 관련 수요는 찬성의 감정을 가진 버즈가 62.6%로 나타났다. 보건복지 관련 주요정책으로는 의료민영화의 버즈가 36.9%로 가장 높게 나타났으며, 과다복지(8.4%), 증세(8.1%), 양육수당(7.2%), 원격의료(6.9%), 연금(6.2%), 건강보험(5.8%)의 순으로 나타났다. 보건복지 관련 주요이슈로는 일자리의 버즈가 20.8%로 가장 높게 나타났으며, 건강(16.4%), 세금(15.7%), 결혼(6.4%), 자살(5.9%), 의료비(5.8%), 출산율

21) Kass, G., "An exploratory technique for investigating large quantities of categorical data", Applied Statistics, Vol. 292, 1980, pp. 119-127.

(4.6%) 등의 순으로 나타났다. 보건복지 관련 분야의 버즈로는 교육(19.2%), 사회복지(18.1%), 보건의료(15.7%), 경제(13.8%), 문화(7.0%), 환경(6.2%), 가정(4.9%) 등의 순으로 나타났다. 보건복지 관련 대상의 버즈로는 노인(15.5%), 가족(15.2%), 공무원(9.4%), 저소득층(8.9%), 장애인(7.7%), 노동자(6.5%), 학생(5.6%) 등의 순으로 나타났다. 보건복지 관련 공공기관의 버즈로는 청와대(34.4%), 보건복지부(32.1%), 공기업(17.5%), 국회(7.0%), 지자체(6.0%) 등의 순으로 나타났다. 보건복지 관련 민간기관의 버즈로는 기업(32.8%), 대학(16.6%), 사회복지시설(16.5%), 관련협회(15.6%) 등의 순으로 나타났다.

〈표 5-22〉 보건복지 관련 버즈 현황

구분	항목	N(%)	구분	항목	N(%)	구분	항목	N(%)
이슈	중증질환	266(1.9)	정책	연금	1,419(6.2)	수요	반대	9,912(37.4)
	의료비	823(5.8)		기초생활	693(3.0)		찬성	16,594(62.6)
	담배	283(2.0)		양육수당	1,644(7.2)		계	26,506
	건강	2,345(16.4)		무상정책	968(4.2)	민간 기관	요양병원	902(5.8)
	일자리	2,970(20.8)		의료민영화	8,456(36.9)		관련협회	2,430(15.6)
	의료수가	154(1.1)		건강보험	1,322(5.8)		기업	5,104(32.8)
	자살	844(5.9)		원격의료	1,573(6.9)		대학	2,592(16.6)
	등록금	502(3.5)		전세대책	117(5)		시민단체	1,982(12.7)
	세금	2,245(15.7)		반값등록금	705(3.1)		사회복지시설	2,563(16.5)
	행복지수	499(3.5)		퇴직연금	855(3.7)		계	15,573
	개인정보	307(2.1)		증세	1,856(8.1)	대상	가족	4,300(15.2)
	부동산	356(2.5)		과다복지	1,923(8.4)		청소년	1,516(5.4)
	정규직	262(1.8)		행복온도	733(3.2)		환자	1,310(4.6)
	위안부	239(1.7)		육아휴직	366(1.6)		의사	1,063(3.8)
	결혼	909(6.4)		중독법	267(1.2)		저소득층	2,511(8.9)
	출산율	662(4.6)		계	22,897		장애인	2,166(7.7)

구분	항목	N(%)	구분	항목	N(%)	구분	항목	N(%)
공공기관	양극화	205(1.4)	분야	주거	429(2.0)		여성	1,514(5.4)
	성폭행	240(1.7)		교육	4,019(19.2)		노동자	1,838(6.5)
	반려동물	182(1.3)		사회복지	3,790(18.1)		서민	1,473(5.2)
	계	14,293		보건의료	3,299(15.7)		중산층	278(1.0)
	국회	2,183(7.0)		경제	2,884(13.8)		노인	4,400(15.5)
	보건복지부	10,041(32.1)		문화	1,458(7.0)		피해자	250(.9)
	청와대	10,778(34.4)		환경	1,297(6.2)		공무원	2,652(9.4)
	고용여성부	488(1.6)		통일	870(4.1)		비정규직	417(1.5)
	공기업	5,469(17.5)		가정	1,019(4.9)		학생	1,591(5.6)
	교육농림부	468(1.5)		노동	802(3.8)		외국인	501(1.8)
	지자체	1,890(6.0)		보육	427(2.0)		노숙자요인	516(1.8)
	계	31,317		범죄	241(1.1)		계	28,296
				안보	233(1.1)			
				다문화	197(.9)			
				계	20,965			

〈표 5-23〉과 같이 청와대와 관련한 정책 버즈는 의료민영화, 증세, 연금, 양육수당, 과다복지 등의 순으로 나타났다. 국회와 관련한 정책 버즈는 의료민영화, 연금, 원격의료, 중독법, 양육수당, 건강보험 등의 순으로 나타났다. 보건복지부와 관련한 정책 버즈는 의료민영화, 건강보험, 행복 온도, 연금, 원격의료, 양육수당 등의 순으로 나타났다. 청와대와 관련한 주요이슈 버즈는 세금, 일자리, 의료비, 건강, 개인정보, 등록금 등의 순으로 나타났다. 국회와 관련한 주요이슈 버즈는 일자리, 건강, 의료비, 중증질환, 세금, 담배 등의 순으로 나타났다. 보건복지부와 관련한 주요이슈 버즈는 건강, 일자리, 중증질환, 의료비, 세금, 담배 등의 순으로 나타났다.

〈표 5-23〉 공공기관별 보건복지 정책 및 이슈 버즈 현황

N(%)

분야	속성	국회	보건 복지부	청와대	고용 여성부처	공기업	교육 농림부처	지자체	합계
정책	연금	245(16.4)	615(9.9)	479(9.5)	25(11.7)	89(3.6)	14(18.9)	57(10.4)	1524
	기초생활	61(4.1)	115(1.9)	57(1.1)	17(7.9)	45(1.8)	10(13.5)	75(13.7)	380
	양육수당	122(8.2)	268(4.3)	362(7.2)	88(41.1)	214(8.6)	17(23.0)	190(34.8)	1261
	무상정책	59(4.0)	48(8)	79(1.6)	4(1.9)	92(3.7)	9(12.2)	6(1.1)	297
	의료민영화	358(24.0)	2763(44.6)	2484(49.4)	6(2.8)	495(20.0)	10(13.5)	58(10.6)	6174
	건강보험	112(7.5)	893(14.4)	127(2.5)	31(14.5)	74(3.0)	7(9.5)	76(13.9)	1320
	원격의료	218(14.6)	430(6.9)	369(7.3)	6(2.8)	19(8)	4(5.4)	46(8.4)	1092
	전세대책	2(1)	3(0)	4(1)	1(5)	2(1)	-	-	12
	반값등록금	29(1.9)	145(2.3)	65(1.3)	2(9)	32(1.3)	2(2.7)	5(9)	280
	퇴직연금	3(2)	26(4)	9(2)	13(6.1)	708(28.6)	-	4(7)	763
	증세	92(6.2)	29(5)	624(12.4)	2(9)	184(7.4)	1(1.4)	2(4)	934
	과다복지	13(9)	21(3)	310(6.2)	2(9)	484(19.5)	-	20(3.7)	850
	행복온도	-	732(11.8)	-	-	-	-	-	732
	육아휴직	13(9)	21(3)	54(1.1)	15(7.0)	38(1.5)	-	7(1.3)	148
	중독법	166(11.1)	82(1.3)	2(0)	2(9)	-	-	-	252
	계	1493	6191	5025	214	2476	74	546	16019
이슈	중증질환	63(10.0)	151(9.4)	72(3.5)	22(7.3)	50(5.6)	-	51(8.0)	409
	의료비	76(12.0)	142(8.8)	326(16.1)	12(4.0)	131(14.7)	2(1.8)	83(13.0)	772
	담배	43(6.8)	96(6.0)	25(1.2)	1(3)	27(3.0)	-	30(4.7)	222
	건강	103(16.3)	376(23.4)	121(6.0)	54(18.0)	83(9.3)	29(26.4)	236(36.9)	1002
	일자리	123(19.4)	339(21.1)	423(20.8)	154(51.3)	127(14.2)	37(33.6)	117(18.3)	1320
	의료수가	16(2.5)	34(2.1)	16(8)	-	3(3)	2(1.8)	1(2)	72
	자살	14(2.2)	32(2.0)	70(3.4)	5(1.7)	7(8)	2(1.8)	9(1.4)	139
	등록금	17(2.7)	28(1.7)	76(3.7)	4(1.3)	154(17.3)	14(12.7)	14(2.2)	307
	세금	52(8.2)	126(7.9)	548(27.0)	7(2.3)	192(21.5)	5(4.5)	16(2.5)	946
	행복지수	1(2)	2(1)	4(2)	-	1(1)	-	-	8
	개인정보	31(4.9)	57(3.6)	99(4.9)	6(2.0)	27(3.0)	4(3.6)	20(3.1)	244
	부동산	14(2.2)	28(1.7)	41(2.0)	4(1.3)	17(1.9)	1(9)	7(1.1)	112
	정규직	15(2.4)	18(1.1)	33(1.6)	4(1.3)	21(2.4)	2(1.8)	7(1.1)	100
	위안부	8(1.3)	20(1.2)	9(4)	5(1.7)	4(4)	2(1.8)	8(1.3)	56
	결혼	24(3.8)	76(4.7)	70(3.4)	13(4.3)	27(3.0)	4(3.6)	18(2.8)	232
	출산율	5(8)	21(1.3)	16(8)	3(1.0)	3(3)	2(1.8)	10(1.6)	60
	양극화	17(2.7)	16(1.0)	61(3.0)	2(7)	12(1.3)	3(2.7)	-	111
	성폭행	11(1.7)	43(2.7)	20(1.0)	3(1.0)	5(6)	1(9)	10(1.6)	93
	반려동물	-	-	-	1(3%)	1(1)	-	2(3)	4
	계	633	1605	2030	300	892	110	639	6209

〈표 5-24〉과 같이 기업과 관련한 정책 버즈는 양육수당, 의료민영화, 연금, 건강보험, 무상정책, 원격의료 등의 순으로 나타났다. 시민단체와 관련한 정책 버즈는 의료민영화, 원격의료, 건강보험, 연금, 양육수당, 기초생활 등의 순으로 나타났다. 기업과 관련한 주요이슈 버즈는 일자리, 건강, 세금, 결혼, 의료비, 등록금 등의 순으로 나타났다. 시민단체와 관련한 주요이슈 버즈는 건강, 일자리, 의료비, 의료 수가, 중증질환, 등록금 등의 순으로 나타났다.

〈표 5-24〉 민간기관별 보건복지 정책 및 이슈 버즈 현황

N(%)

분야	속성	요양병원	관련협회	기업	대학	시민단체	사회복지 시설	합계
정책	연금	57(10.3)	18(1.1)	136(11.9)	81(15.7)	97(8.3)	20(9.8)	409
	기초생활	44(7.9)	-	51(4.5)	47(9.1)	44(3.8)	19(9.3)	205
	양육수당	87(15.7)	14(9)	213(18.6)	103(20.0)	90(7.7)	118(57.6)	625
	무상정책	8(1.4)	10(6)	90(7.9)	39(7.6)	31(2.7)	13(6.3)	191
	의료민영화	76(13.7)	1071(65.0)	183(16.0)	29(5.6)	325(27.9)	8(3.9)	1692
	건강보험	131(23.6)	93(5.6)	102(8.9)	76(14.8)	199(17.1)	13(6.3)	614
	원격의료	111(20.0)	436(26.5)	83(7.2)	11(2.1)	284(24.4)	-	925
	전세대책	-	-	4(3)	1(2)	3(3)	-	8
	반값등록금	15(2.7)	1(1)	38(3.3)	31(6.0)	17(1.5)	-	102
	퇴직연금	9(1.6)	-	50(4.4)	14(2.7)	15(1.3)	7(3.4)	95
	중세	2(4)	4(2)	70(6.1)	41(8.0)	22(1.9)	7(3.4)	146
	과다복지	8(1.4)	-	78(6.8)	21(4.1)	13(1.1)	-	120
	육아휴직	6(1.1)	-	46(4.0)	21(4.1)	11(9)	-	84
	중독법	-	-	2(2)	-	12(1.0)	-	14
계		554	1647	1146	515	1163	205	5230
이슈	중증질환	63(10.8)	17(6.4)	36(2.1)	43(3.6)	53(4.6)	8(2.9)	220
	의료비	104(17.9)	64(24.2)	111(6.6)	71(6.0)	205(17.8)	18(6.6)	573
	담배	27(4.6)	11(4.2)	38(2.3)	42(3.5)	27(2.3)	2(7)	147
	건강	158(27.1)	60(22.7)	265(15.8)	198(16.7)	267(23.2)	62(22.8)	1010
	일자리	101(17.4)	35(13.3)	417(24.9)	376(31.7)	261(22.7)	80(29.4)	1270
	의료수가	19(3.3)	28(10.6)	15(9)	3(3)	82(7.1)	00.0%	147
	자살	4(7)	6(2.3)	18(1.1)	26(2.2)	25(2.2)	9(3.3)	88

분야	속성	요양병원	관련협회	기업	대학	시민단체	사회복지 시설	합계
	등록금	16(2.7)	-	90(5.4)	167(14.1)	44(3.8)	6(2.2)	323
	세금	12(2.1)	10(3.8)	233(13.9)	61(5.1)	43(3.7)	18(6.6)	377
	행복지수	-	-	5(3)	4(3)	2(2)	1(4)	12
	개인정보	24(4.1)	7(2.7)	26(1.6)	17(1.4)	25(2.2)	16(5.9)	115
	부동산	1(2)	-	61(3.6)	29(2.4)	13(1.1)	5(1.8)	109
	정규직	10(1.7)	2(8)	53(3.2)	24(2.0)	20(1.7)	2(7)	111
	위안부	5(9)	6(2.3)	32(1.9)	23(1.9)	7(6)	2(7)	75
	결혼	13(2.2)	5(1.9)	162(9.7)	59(5.0)	33(2.9)	8(2.9)	280
	출산율	9(1.5)	-	21(1.3)	8(7)	13(1.1)	2(7)	53
	양극화	14(2.4)	9(3.4)	36(2.1)	21(1.8)	19(1.6)	3(1.1)	102
	성폭행	2(3)	4(1.5)	51(3.0)	13(1.1)	11(1.0)	29(10.7)	110
	반려동물	-	-	5(3)	2(2)	2(2)	1(4)	10
	계	582	264	1675	1187	1152	272	5132

나. 보건복지 정책 수요에 미치는 영향 요인

보건복지 정책에서 전세대책, 양육수당, 기초생활, 건강보험, 퇴직연금, 중독법, 원격의료 요인은 정적(+)인 영향을 미치는 것으로 나타나, 동 요인들에 대한 찬성의 확률이 높은 것으로 나타났다. 과다복지, 증세, 반값등록금, 무상정책, 의료민영화, 연금 요인은 부적(-)인 영향을 미치는 것으로 나타나, 동 요인들에 대한 반대의 확률이 높은 것으로 나타났다 (표 5-25 참조).

〈표 5-25〉 보건복지 정책 수요에 미치는 영향요인*

변수	b ^a	S.E. ^b	OR(95%CI) ^c	P
연금	-.831	.097	.436(.36-.58)	.000
기초생활	2.640	.315	14.017(7.55-26.01)	.000
양육수당	2.751	.170	15.657(11.22-21.85)	.000
무상정책	-.966	.123	.380(.30-.48)	.000
의료민영화	-.967	.039	.380(.35-.41)	.000

변수	b ^a	S.E. ^b	OR(95%CI) ^c	P
건강보험	2.258	.145	9.565(7.19-12.72)	.000
원격의료	.276	.083	1.318(1.12-1.55)	.001
전세대책	3.409	.482	30.240(11.76-77.79)	.000
반값등록금	-1.997	.168	.136(.10-.19)	.000
퇴직연금	2.242	.470	9.416(3.75-23.65)	.000
증세	-1.609	.094	.200(.17-.24)	.000
과다복지	-2.523	.084	.080(.07-.09)	.000
행복온도	.667	1.118	1.949(.22-17.44)	.551
육아휴직	.324	.280	1.382(.80-2.39)	.248
중독법	.803	.254	2.233(1.36-3.67)	.002
상수	.719	.016	2.05	0.000

주: *기준범주: 반대, ^a Standardized coefficients, ^b Standard error,

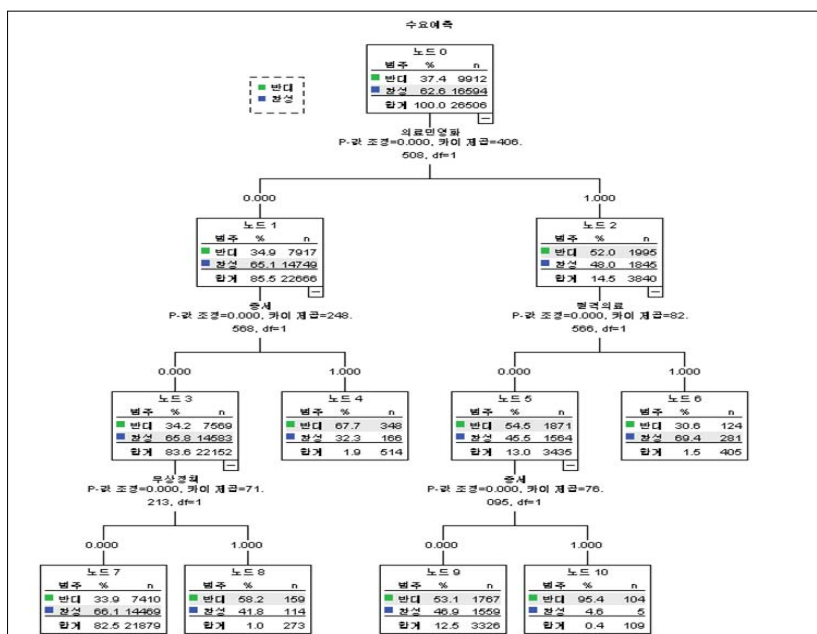
^c Adjusted odds ratio(95% Confidence interval)

다. 보건복지 정책 수요 예측모형

본 연구에서는 주요정책에 대한 수요를 예측하기 위하여 연금, 무상정책, 의료민영화, 원격의료, 증세 요인에 대한 데이터마이닝 분석을 실시하였다. 보건복지 정책 수요 예측모형에 미치는 영향은 [그림 5-9]와 같다. 나무구조의 최상위에 있는 네모는 루트노드로서, 예측변수(독립변수)가 투입되지 않은 종속변수(찬성, 반대)의 빈도를 나타낸다. 루트노드에서 보건복지 전체 정책의 찬성은 62.6%(16,594건), 반대는 37.4%(9,912건)으로 나타났다. 루트노드의 하단의 가장 상위에 위치하는 요인이 보건복지 정책 수요 예측에 가장 영향력이 높은(관련성이 깊은)요인으로 ‘의료민영화요인’의 영향력이 가장 큰 것으로 나타났다. ‘의료민영화요인’이 높을 경우 보건복지 정책 찬성이 이전의 62.6%에서 48.0%로 감소한 반면, 보건복지 정책의 반대는 이전의 37.4%에서 52.0%로 증가하였다. ‘의료민영화요인’이 높고 ‘원격의료요인’이 높은

경우 보건복지 정책의 찬성이 이전의 48.0%에서 69.4%로 증가한 반면, 보건복지 정책의 반대는 이전의 52.0%에서 30.6%로 감소하였다. <표 5-26>의 보건복지 정책 수요 예측에 대한 이익도표와 같이 보건복지 정책의 찬성에 가장 영향력이 높은 경우는 ‘의료민영화요인’이 높고 ‘원격의료요인’이 높은 조합으로 나타났다. 즉, 6번 노드의 지수(index)가 110.8%로 뿌리마디와 비교했을 때 6번 노드의 조건을 가진 집단이 보건복지 정책을 찬성할 확률이 약 1.11배로 나타났다. 보건복지 정책의 반대에 가장 영향력이 높은 경우는 ‘의료민영화요인’이 높고 ‘원격의료요인’이 낮고, ‘증세요인’이 높은 조합으로 나타났다. 즉, 10번 노드의 지수가 255.1%로 뿌리마디와 비교했을 때 10번 노드의 조건을 가진 집단이 보건복지 정책을 반대할 확률이 약 2.56배로 나타났다.

[그림 5-9] 보건복지 정책 수요 예측모형



〈표 5-26〉 보건복지정책 수요 예측모형에 대한 이익도표

구분	노드	이익지수				누적지수			
		노드(n)	노드(%)	이익(%)	지수(%)	노드(n)	노드(%)	이익(%)	지수(%)
찬성	6	405	1.5	1.7	110.8	405	1.5	1.7	110.8
	7	21879	82.5	87.2	105.6	22284	84.1	88.9	105.7
	9	3326	12.5	9.4	74.9	25610	96.6	98.3	101.7
	8	273	1.0	.7	66.7	25883	97.6	99.0	101.4
	4	514	1.9	1.0	51.6	26397	99.6	100.0	100.4
	10	109	.4	.0	7.3	26506	100.0	100.0	100.0
반대	10	109	.4	1.0	255.1	109	.4	1.0	255.1
	4	514	1.9	3.5	181.1	623	2.4	4.6	194.0
	8	273	1.0	1.6	155.7	896	3.4	6.2	182.4
	9	3326	12.5	17.8	142.1	4222	15.9	24.0	150.6
	7	21879	82.5	74.8	90.6	26101	98.5	98.7	100.3
	6	405	1.5	1.3	81.9	26506	100.0	100.0	100.0



제6장

소셜 빅데이터를 활용한 사회위험요인 예측

제1절 비만 빅데이터 수집을 위한 관련 주제 분류

제2절 소셜 빅데이터를 이용한 비만 담론 분석

제3절 소셜 빅데이터를 이용한 청소년 자살위험 예측요인

6

소셜 빅데이터를 활용한 << 사회위험요인 예측

SNS는 청소년들의 일상생활 속에서 갖는 우울한 감정이나 스트레스, 고민을 들을 수 있고 행태를 이해할 수 있는 장소로 SNS상에서 나타나는 자살에 대한 감정표현이나 심리적 위기 행태들을 분석하게 되면 위험징후와 유의미한 패턴을 감지하여 자살을 예방하는데 긍정적 효과가 발휘될 수 있다. 우리나라는 급격한 사회·경제적 변화 속에 자살률이 2004년부터 OECD(Organization for Economic Cooperation and Development) 회원국 중 최고의 수준이며, 특히 청소년계층의 자살 문제가 사회적 이슈로 대두되면서 정부차원의 적극적인 대책이 시급한 실정이다. 또한, 사이버따돌림에 노출된 청소년들이 자살을 선택하거나 폭력의 가해자가 됨에 따라 심각한 사회문제로 떠오르고 있다. 이와 같은 사회위험에 대한 원인과 관련요인을 규명하기 위하여 기존에 실시하던 횡단적 조사나 종단적 조사 등을 대상으로 한 연구는 정해진 변인들에 대한 개인과 집단의 관계를 보는데는 유용하나 사이버 상에서 언급된 개인별 버즈(buzz)가 사회적 현상들과 어떻게 얼마나 연관되어 있는지를 밝히는 데는 한계가 있다.

이러한 점에서 소셜 빅데이터를 활용한 다변량 분석은 새로운 상관관계나 패턴 등을 발견함으로써 사회위험 요인과 같이 인간행동의 복잡하고 역동적인 현상에서 발생하는 다양한 원인들의 상호작용 관계를 효과적으로 분석할 수 있다. 따라서 본 연구는 우리나라 온라인 뉴스사이트, 블로그, 카페, 게시판 등에서 수집된 소셜 빅데이터를 바탕으로 본 연구에서는 비만과 청소년 자살을 사회위험요인 사례로 제시하였다.

제1절 비만 빅데이터 수집을 위한 관련 주제 분류²²⁾

1. 연구의 필요성

최근 스마트폰과 모바일 인터넷 서비스의 활성화로 인해, 사용자의 위치정보와 온라인 및 모바일 사용기록, SNS 상의 사용자의 일상생활과 의견에 대한 저장으로 정보량이 폭증하게 되었다(김지현 외, 2014; 송태민, 2012). 또한 정보 통신기술의 발전으로 많은 데이터를 빠르게 저장 및 처리하는 것이 가능해지면서 빅데이터(BigData)에 대한 관심이 증가하고 있다(김정태 외, 2013). 빅데이터는 일반적인 데이터베이스나 소프트웨어로는 관리하기 어려운 정도의 큰 규모를 가진 대용량 데이터를 의미하며, 최근에는 대용량 자료의 수집, 저장, 분석, 체계화를 위한 도구, 플랫폼, 분석기법 등을 포괄하는 용어로 변화하고 있으며, 대용량 자료로부터 가치 있는 정보를 찾아내고 생성된 지식에 근거하여 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술을 의미한다(강만모 외, 2012). 즉, 빅데이터는 데이터 자체에 대한 제공과 분석 서비스를 제공하는 두 가지 특징을 가지고 있다.

사람들은 소셜 네트워킹 서비스(SNS)를 통해 자신의 경험이나 신념, 선호하는 것들에 대해 자유롭게 이야기하고, 다른 이들과 이야기를 공유 및 전파하여 사회 전반에 큰 영향력을 행사하고 있다(노진석, 2012). 소셜 분석의 대표적인 기술 중의 하나는 텍스트마이닝이다. 텍스트마이닝은 말 그대로 텍스트에서 의미를 찾아내는 것을 말하며, 컴퓨터로 문자가

22) 본 절은 본 과제의 일환으로 비만 빅데이터 수집을 위해 서울대학교 간호대학 박현애교수 연구팀과 공동으로 수행되었으며, 'Ae Ran Kim, Tae Min Song, Hyeoun-Ae Park(2014). Development of an obesity ontology for collection and analysis of big data related to obesity. TBC 2014' 에 게재하였음을 밝힌다.

지닌 의미와 구조를 분석하는 자연어 처리(Natural Language Processing) 기술을 이용해 다양한 SNS 플랫폼에서 자발적으로 생성되고 있는 수십억 건의 메시지를 분석하여 사회 현상을 찾아내는 것을 가능하게 하였다(노진석, 2012).

2011년 현재 우리나라의 19세 이상 성인비만율(BMI 25 이상)은 전체 인구의 31.9% 수준으로 남자는 35.2%, 여자는 28.6%를 차지한다(질병관리본부, 2011). 이는 1998년 26.0%에서 2005년 31.3%로 5.3% 증가한 이후에 지속적으로 높은 수준을 유지하고 있는 추세이다(질병관리본부, 2011). 비만으로 인한 건강보험 진료비 지출 규모는 2007년 1조 8,971억원에서 2011년 2조 6,919억원으로 41.9% 증가하였고, 이는 건강보험 전체 진료비의 5.8%에 해당하는 규모이다(건강보험공단, 2012). 비만과 관련한 질병의 진료비 지출 비중은 고혈압(36.2%), 당뇨(20.1%), 뇌졸중(12.0%), 허혈성 심장질환(9.2%), 골관절염(7.9%) 등의 순으로 높게 나타났다(건강보험공단, 2012). 이와 같이 비만 문제의 급속한 증가로 인한 사회 경제적 손실과 부담감은 매우 크다고 볼 수 있다. 비만을 예방하기 위한 건강한 생활 습관의 실천을 위해서는 비만의 폐해 및 실태에 대한 일반 국민의 인식 제고와 자기관리 능력의 향상이 필요하다. 비만과 관련된 요인은 복합적인 것으로 비만과 관련된 개별적인 변인과 함께 사회환경적 변인과의 관계 분석이 필요하다. 비만에 대한 관심 증가와 함께 대중들은 인터넷과 소셜 미디어를 이용하여 비만에 대한 정보를 얻고 공유하며 의견을 교환한다. 이에, 소셜 미디어 상에서 생성되고 있는 비만 관리 주제와 관련된 빅데이터를 수집하고 분석한다면 비만과 관련된 대중의 인식과 대처 방법, 사회 현상을 보다 실제적으로 파악하여 그 의미를 도출해 낼 수 있을 것이다.

소셜 미디어에서 표현되는 언어들은 주로 사람들의 일상 대화에서 쓰

이는 구어체 문장으로 이루어진 비정형 데이터이기 때문에(노진석, 2012), 이를 보다 효과적으로 수집 및 분석하기 위한 분석틀이 필요하다. 분석틀의 내용으로는 관련 주제가 어떠한 개념 영역들로 구성되어 있는지와 각 개념 간 관계에 대한 정의가 필요하기 때문에, 이를 반영한 온톨로지(ontology)가 개발될 필요가 있다. 온톨로지(ontology)는 관심 주제의 공유된 개념(shared concepts)을 형식화하고(formalizing) 표현하기 위한(representing), 컴퓨터가 해석 가능한 지식 모델(computer-interpretable knowledge model)이다(Kim et al., 2013). 온톨로지를 통하여, 단어(word)를 사용해 어떤 영역(domains)을 개념화하고 명시적으로 표현하며, 개념의 종류와 그들의 연관 관계(relationship)를 명백하게 기술할 수 있다(이경일, 2006). 또한 수집되는 빅데이터 자료는 비정형적으로 다양하게 표현됨으로써 온톨로지를 구성하는 개념을 설명하는 용어와 그 유의어를 정의하여 기술하여 용어체계를 마련하는 것이 필요하다. 이에 본 연구에서는 비만 주제에 대해 수집된 빅데이터 자료를 식별하고 활용하기 위한 분석틀로서, 비만 관리 관련 주제를 분류하고, 비만 관리 온톨로지와 용어체계를 개발하고자 한다.

2. 연구목표

비만 빅데이터의 수집과 분석을 위한 분석틀을 개발한다.

가. 비만 빅데이터 수집을 위한 분류틀을 개발한다.

나. 비만 관리 주제를 구성하는 내용 영역과 영역간 관계를 설명하는 비만 관리 온톨로지를 개발한다.

다. 비만 관리 온톨로지를 구성하는 영역의 개념을 표현하는 비만 관리 용어체계를 개발한다.

3. 연구방법

비만 빅데이터 수집과 분석을 위한 분석틀의 개발 과정은 구체적으로 다음과 같다.

가. 공신력 있는 기관에서 발표한 임상실무지침을 포함한 관련 문헌 조사를 통해 비만 주제를 설명하는 개념 내용 및 분류 기준을 파악하였다.

나. 선행 문헌 고찰 결과, 본 연구에서는 NICE의 지침(NICE, 2006)과 ‘비만치료지침 2010 권고안’(대한내분비학회, 2010)에서 다루고 있는 비만 관리에 대한 내용을 근거로 하여 비만 관리 주제의 대분류, 중분류와 소분류 영역을 도출하였다.

다. 비만 관리에 대한 영역간 관계를 기술한 비만 관리 온톨로지를 개발하였다.

라. 관련 문헌과 인터넷 키워드 검색 등을 통해, 비만 주제의 대분류-중분류-소분류 영역별 지식 진술문에서 개념 및 용어를 추출하여 비만 관리 용어체계를 개발하였다.

4. 연구결과

가. 비만 관련 임상실무지침

비만을 관리하기 위한 임상실무지침으로는 NHLBI(1998), NICE(2006), AMA(2003)에서 개발한 지침들이 널리 알려져 있다(전은주, 2013). NICE(NICE, 2006)는 성인과 아동에서의 비만과 과체중의 예방, 진단, 사정과 관리에 대한 지침을 제시하였다. NHLBI(NHLBI, 1998)는 알고리즘 유형의 임상실무지침으로 조건에 따른 의사 결정을 보

여주는 도식 형태의 지침으로 되어 있고(전은주, 2013), AMA(2003)에서 개발한 임상실무지침은 의료인이 성인 비만 환자에게 의료 서비스를 제공하는 데에 사용 가능한 입문서이다(전은주, 2013).

비만 관리 임상실무지침들에서는 비만의 정의와 분류 기준, 위험요인, 치료와 예방 등의 내용을 공통적으로 다루고 있었다. NICE(NICE, 2006)는 아동과 성인 대상의 비만에 대한 임상 관리 지침(clinical care pathway)과 공공 보건 건강 지도(public health map)를 각각 제시하고 있으며, 비만 예방부터 치료에 이르기까지의 관리에 대하여 이전 선행연구들의 근거와 함께 상세히 설명하고 있다. 대한내분비학회와 대한비만학회에서는 한국인 대상의 ‘비만치료지침 2010 권고안’을 제시하였다(대한내분비학회, 2010). 본 연구에서는 NICE의 지침(NICE, 2006)과 ‘비만치료지침 2010 권고안’(대한내분비학회, 2010)에 근거하여 비만 빅데이터 수집을 위한 분석틀로서, 비만 관리 주제에 대한 분류틀(대분류, 소분류, 중분류), 비만관리 온톨로지와 용어체계를 개발하였다.

나. 비만 관리 주제 분류틀

비만 주제와 관련된 임상실무지침 및 선행 문헌 자료를 조사한 결과, 비만 관리를 설명하는 영역의 대분류에는 비만, 인구학적 특성, 위험요인, 증상 및 징후, 합병증, 진단, 치료와 예방이 포함되었다. 각 대분류 영역에 대한 중분류와 소분류의 설명과 근거는 다음과 같다.

1) 비만(정의 및 분류)

비만이란 체지방의 과잉 축적으로 정의되며, 우리나라 성인의 경우에

보통 체질량지수 (BMI) 25 kg/m^2 이상인 경우를 말한다(대한비만학회, 2010). 성인의 경우에, BMI $25.0 \sim 29.9(\text{kg/m}^2)$ 는 과체중(overweight), $30.0 \sim 34.9$ 는 1단계 비만(class I), $35.0 \sim 39.9$ 는 2단계 비만(class II), 그리고 40 이상은 3단계 비만 또는 초고도 비만(extreme obesity, class III)로 분류된다(NHLBI, 1998; NICE, 2006; 최중명, 2011). 수정된 WHO의 정의에 의하면, BMI $35 \sim 40 \text{ kg/m}^2$ 의 경우 severe obesity로 정의하였다(Wikipedia, 2014). BMI 35 이상이면서 비만과 관련된 건강 문제를 동반하거나, BMI $40 \sim 44.9$ 의 경우에는 morbid obesity, BMI 45~50 이상의 경우에는 super obesity(초고도 비만)로 분류된다(Wikipedia, 2014). 한국인의 비만 기준은 BMI 25 kg/m^2 이상으로 한다(대한내분비학회, 2010). 소아 청소년의 경우에는, 2세 이상의 모든 소아 청소년에서 2007년 소아청소년 표준 성장도표를 기준으로 연령별, 성별 BMI가 95 백분위수 이상 혹은 BMI 25 이상을 비만으로 진단하고 85-94 백분위수는 과체중(비만위험군)으로 판정한다. 또는 비만도 20% 이상을 비만으로 진단하며, 이 중에서 20-29%는 경도 비만, 30-49%는 중등도 비만, 50% 이상을 고도 비만으로 분류한다(대한내분비학회, 2010).

비만은 원인에 따라 원발성 비만과 이차성 비만으로 분류된다(대한내분비학회, 2010), 원발성 비만은 단순성 비만이라 불리기도 하며, 개체가 섭취한 에너지의 총량과 소비한 에너지의 균형이 무너진 상태에서 체지방이 증가하여 발생하는 것으로 설명된다(대한비만학회, 2010). 이는 식습관, 생활습관, 연령, 유전적 요인 등의 다양한 위험 요인이 복합적으로 관여하는 경우가 많다(대한비만학회, 2010). 2차성 비만은 내분비 질환이나 유전, 약제 등이 원인이 된다(대한비만학회, 2010).

부위에 따라 전신 비만과 부분 비만으로 분류될 수 있다. 부분 비만의 예

로써, 복부 비만은 복부에 지방세포가 축적되는 비만(사과형 비만, android)으로 남성에게서 많고, 둔부, 허벅지 등의 하체에 지방세포가 축적되는 하체 비만(배형 비만, gynoid)은 여성에게서 흔히 나타난다(민경훈·이태용, 2011).

〈표 6-1〉 비만 영역 분류틀

대분류	중분류	소분류
비만	BMI에 따른 분류	과체중, 위험체중, 1단계 비만, 2단계 비만, 3단계 비만, severe obesity, morbid obesity, super obesity
	부위에 따른 분류	전신 비만, 부분 비만(상체 비만, 하체 비만, 복부 비만, 내장 비만)
	원인에 따른 분류	원발성 비만, 이차성 비만

2) 인구학적 특성

비만의 대상자는 인구학적 특성을 가지며, 이는 성별, 연령별, 세팅별 영역으로 구성된 중분류로 분류하고 있다. 성별에 따른 소분류는 남성에서의 비만과 여성에서의 비만이 포함되고 있다.

질병관리본부(질병관리본부, 2011)에서 발표한 ‘국민건강통계: 국민건강영양조사’에 의하면 19세 이상 성인에서 2011년도 우리나라 만 19세 이상 국민의 비만 유병률(BMI 25 이상)은 전체 31.9%, 남자 35.2%, 여자 28.6%로 남자가 여자보다 6.6%p 높았으나, 60대 이후에는 여자가 남자보다 약 10%p 높았다. 허리둘레 기준 비만 유병률은 전체 26.3%, 남자 27.6%, 여자 25.0%이었고, 남자, 여자 모두 60대에서 가장 높았다(질병관리본부, 2011). BMI와 허리둘레 기준 모두에서, 40대 이전에는 남자가 여자보다 높고, 50대 이후에는

여자가 남자보다 높은 양상을 나타내었다(질병관리본부, 2011). 2011년도 기준 소아청소년(만2-18세)의 비만 유병률은 전체 9.7%, 남자 11.0%, 여자 8.3%이었다(질병관리본부, 2011).

연령에 따라 태아, 소아, 청소년, 학생, 성인, 임신 전후, 폐경의 시기로 분류하고 있다. NICE (NICE, 2006)의 지침에서는 생애주기에 따라서 체중 증가가 최대화되는 시기가 있고, 이러한 시기가 과잉 체중 증가와 연결될 수 있다고 하였다. 태아 시기의 영양은 성장, 체형, 비만, 에너지 조절에 영구적인 영향을 미칠 수 있다(NICE, 2006). 소아 청소년기 비만은 성인 비만으로 이행하기 쉬우며 소아청소년 및 성인기의 만성질환을 유발할 가능성이 크다(대한내분비학회, 2010). 청소년기에는 자율성의 증가와 함께 불규칙한 식사, 식사 습관 변화와 신체활동 감소 등으로 인한 체중증가로 비만이 발생할 수 있다(NICE, 2006). 임신 기간 동안의 과잉 체중 증가는 출산 후의 체중 증가와도 연결된다(NICE, 2006). 폐경기에는 호르몬 변화로 인해 특히 복부 부위를 포함한 체중 증가의 가능성이 높아진다(NICE, 2006). 세팅별 분류기준을 살펴보면, NICE(NICE, 2006)의 지침에 따라 임상(clinical), 공공 보건(public health)과 취약계층에 대한 대상으로 분류하였다. 임상 세팅에는 일차 진료(primary care), 특정 질병을 동반한 대상자에 대한 이차 진료(secondary care), 그리고 비만에 대한 전문가의 진료(specialist obesity services)가 있으며, 공공 보건에는 일반적인 대중, 학령 전기 아동, 학생, 직장, 지역사회가 해당된다(NICE, 2006). 취약계층에 해당하는 건강 불평등(inequalities)과 관련하여서는 사회 상태, 소수 민족(ethnicity), 그리고 장애(disabilities)가 있거나 심한 정신 건강 문제, 보호 시설의 어린이 등을 언급하였다(NICE, 2006).

〈표 6-2〉 인구학적 특성 영역 분류들

대분류	중분류	소분류
대상자	성별	남성, 여성
	연령별	태아, 소아(학령 전기, 학령기 등), 청소년, 학생(여학생, 남학생, 초등학생, 중학생, 여중생, 고등학생, 여고생, 대학생, 여대생 등), 성인(청년, 장년, 노년), 임신, 산후, 폐경
	세팅별	임상, 공공 보건(일반 대중, 보육시설, 학교, 지역사회, 직장), 취약계층(사회적 수준, 교육 수준, 인종, 불구, 고아 등)

3) 위험요인

원발성 비만은 전체의 90% 이상으로(대한비만학회, 2010), 섭취 에너지가 소비 에너지를 초과하여 체지방이 증가된 상태를 의미하며, 식습관, 생활 습관, 연령, 인종, 유전적 요인 등의 다양한 위험요인이 복합적으로 관여한다(대한내분비학회, 2010; 대한비만학회, 2010). 일부 비만 환자에서 유전 및 선천성 장애, 약물, 신경 및 내분비계 질환, 정신 질환 등과 관련되어 이차성 비만이 발견된다(대한비만학회, 2010). 이러한 경우에는 정확한 원인 감별을 통한 효과적인 체중 감량이 필요하다(대한비만학회, 2010). 대한비만학회(대한비만학회, 2010)에서 다음과 같이 제시한 원발성 비만과 이차성 비만의 원인에 따라 위험요인 분류에 대한 중분류와 소분류를 작성하였다. NICE가 clinical care pathway에서 제시한 사정 평가 항목(NICE, 2006)에서는 다음의 위험요인들을 사정해야 한다고 제시하고 있으며, 이는 대한비만학회(대한비만학회, 2010)에서 제시한 원발성 비만과 이차성 비만의 위험요인과 유사하다. ∴ 과체중 또는 비만의 증상과 기저 원인을 파악한다, 변화하고자 하는 의지 또는 동기화 정도를 파악한다, 위험 요소와 동반 질환, 정신적 스트레스를 파악한다, 과체중 또는 비만 그리고 동반 질환의 가족력을 파악한다, 식이와 신체적 활동과 같은 생활

습관을 사정한다, 환경적·사회적·가족적 요소를 파악한다, 정신심리적 문제 또는 의학적 문제와 투약 상태를 사정한다, 아동의 경우에는 성장 상태를 사정한다.

〈표 6-3〉 위험요인 영역 분류들

대분류	중분류	소분류
위험 요인	식이 및 식사 습관	고지방식이(지방, 과자, 삼겹살구이, 튀김 등), 고열량식이, 고탄수화물식이, 가공 식품(패스트푸드, 인스턴트, 통조림, 과자, 식품첨가물, 탄산음료 등), 섭식 장애(식사 횟수 감소, 빠른 식사 시간, 과식, 다이어트, 야식증후군 등)
	신체적 활동	활동 부족(TV 시청, 컴퓨터게임, 대중교통 발달, 놀이공간 협소 등)
	기타 생활습관	수면 부족, 금연, 흡연, 음주 등
	인종	인종
	사회 경제적 요인	낮은 사회 경제적 지위 등
	유전적 요인	가족력, 비만 유발 유전자(ob, db, POMC, MC4R 등), 비만 관련 선천성 장애(프라더-윌리 증후군, 로렌스-문-비들 증후군, 알스트롬 증후군, 코헨 증후군, 카펜터 증후군 등)
	약물	항정신병약물(티오리다진, 클로자핀, 올란자핀, 퀴에타핀, 리스페리돈 등), 알파-2 길항제(이미프라민, 미트라자핀 등), 선택적 세로토닌 재흡수 억제제(패록세틴), 항전간제(발프로에이트, 카바마제핀, 가바펜틴 등), 당뇨병 치료제(인슐린, 설폰요소제, 글리나이드제제, 타아졸리디네돈 등), 세로토닌 길항제(피조티펜 등), 항히스타민제(사이프로헵타딘 등), 베타 차단제(프로프라놀롤 등), 알파 차단제(테라조신 등), 스테로이드 제제 (경구 피임제, 당질 코르티코이드 등)
	신경 및 내분비계 질환	시상하부성 비만(외상, 종양, 감염성 질환, 수술, 뇌압 상승 등), 쿠싱증후군, 갑상선 기능 저하증, 인슐린종, 다낭난소증후군, 성인 성장호르몬 결핍증
	심리적 요인	스트레스, 강박감, 불안, 외로움, 사랑 결핍, 관심 결핍
	정신질환	과식장애, 계절성 정서 장애

4) 증상 및 징후

비만은 무증상으로 나타나기도 하지만, 비만으로 인해 신체적 그리고 정신적으로 다양한 자각 증상과 징후를 경험할 수 있다. 1개 병원 건강검진센터에서 검진을 받고 비만으로 판정 받은 대상자의 일반적 특성과 주요 증상들을 조사한 이전의 연구(탁기천·류규수, 2001)에 의하면, 대상자가 호소한 주요 신체적 증상으로는 트림, 소화불량, 목의 통증, 호흡곤란, 갈증, 소변량 증가, 잦은 소변, 현기증, 요통, 두통, 불규칙한 월경, 발진, 시력장애, 청력감소 등이 있었다. 외모를 강조하는 사회 분위기 속에서, 특히 여성들은 체중 증가를 부끄러워하며 불안 및 우울, 사회적 고립 등의 정신적 증상을 나타내며, 체중 증가를 막기 위한 강박적인 시도를 하기도 한다(강원섭·김종우, 2010). 이에 비만의 증상과 징후에 대한 분류를 신체적 측면과 정신적 측면으로 구분하고 이에 해당되는 증상 및 징후들을 소분류로 분류하였다.

〈표 6-4〉 증상 및 징후 영역 분류틀

대분류	중분류	소분류
증상 및 징후	신체적 측면	무증상, 트림, 소화불량, 목의 통증, 호흡곤란, 갈증, 소변량 증가, 잦은 소변, 현기증, 요통, 두통, 불규칙한 월경, 발진, 시력장애, 청력감소 등
	정신적 측면	스트레스, 자살충동, 고민, 불안, 공포, 불면, 대인기피, 우울, 신경성 식욕부진, 신경성 폭식증, 자존감 저하, 대인관계 위축, 차별감 등

5) 합병증

비만 환자는 비만하지 않은 건강인에 비하여 고혈압, 심혈 관계 질환, 뇌졸중, 고지혈증, 당뇨병, 담낭질환, 신장질환, 간 기능 부전, 근골격계

질환, 관절염, 수면장애, 악성신생물 등의 각종 질병에 중복 이환될 위험성이 높다(최중명·김춘배, 2011). 여러 문헌들에서 비만과 관련하여 나타날 수 있는 연관 질병과 건강문제들에 대해 기술하고 있다(NICE, 2006; 대한내분비학회, 2010; 대한비만학회, 2010; 최중명·김춘배, 2011). 본 연구에서는 비만의 합병증에 해당하는 질환들을 내분비계, 소화기계, 심혈관계, 근골격계, 호흡기계, 생식/호르몬계, 소화기계, 암, 기타 질환으로 분류하여 중분류와 소분류를 작성하였다.

〈표 6-5〉 합병증 영역 분류들

대분류	중분류	소분류
합병증	내분비계 질환	당뇨병, 대사증후군, 인슐린 저항성 등
	심혈관계 질환	관상동맥질환, 부정맥, 울혈성 심부전, 비후성 심근증, 돌연사, 고혈압, 고지혈증 등
	근골격계 질환	골관절염, 요통, 통풍 등
	호흡기계 질환	수면 무호흡증, 호흡 저하, 숨가쁨, 코골이 등
	생식/호르몬계 질환	성기능 장애, 호르몬 장애, 불임, 다낭성 난소 증후군 등
	소화기계 질환	담낭 질환, 담석증, 지방간 등
	암	유방암, 대장암, 담낭암, 췌장암, 신장암, 방광암, 자궁경부암, 전립선암, 자궁내막암 등
	기타	마취 관련 위험 증가, 정신 장애, 모성 비만과 관련된 태아 결함, 뇌졸중, 고요산혈증, 요요현상 등

6) 진단

비만 진단을 위해 성인에서 가장 흔히 사용되는 방법은 체질량지수(BMI)와 허리 둘레를 측정하는 것이다(NICE, 2006). 체질량지수는 체중을 키의 제곱으로 나누어 계산된다. 허리 둘레의 측정은 허리와 엉덩이의 비율을 계산함으로써 개인의 전반적인 체형을 판단할 수 있다(NICE, 2006). 허리둘레로 본 복부 비만의 기준은 남자에서는 90 cm 이상, 여자에서는 85 cm 이상을 권고하며(대한내분비학회, 2010), 허리 둘레에 따라 동반질환의 위험도에는 차이가 있다(최중명·김춘배, 2011).

실제 비만은 체지방의 과다상태로 체지방량을 기준으로 판단해야 하는데, 최근에는 생체전기저항분석(bioelectrical impedance analysis, BIA) 방식의 체성분 분석기를 이용하여 비교적 정확한 체지방률 및 체지방률, 근육량 등을 측정하는 방법도 있다(민경훈·이태용, 2011). 하지만 이러한 직접적인 계측 방법은 시간, 비용, 고가의 장비 등이 요구되는 복잡한 과정이다(NICE, 2006; 민경훈·이태용, 2011). 실제로 임상에서는 한 가지의 방법보다는 신체계측법과 기기를 이용하는 방법을 병용하여 사용한다(대한비만학회, 2010). 본 연구에서는 비만의 진단방법에 대하여, 신체계측 방법과 기기를 이용한 평가로 분류하여, 각 중분류에 해당하는 소분류를 추출하고 설명하였다.

〈표 6-6〉 진단 영역 분류틀

대분류	중분류	소분류
진단	신체계측	BMI, 허리둘레, 엉덩이둘레, 허리-엉덩이 둘레 비율, 신장, 체중, 케틀레지수 체지방량, 피부주름두께, 복부내장지방량, 체형 등
	기기 이용 평가	체성분분석기, 이중에너지 X-선 흡수법, 임피던스, 브로카 변법, 복부 CT, 복부 MRI 등

7) 치료

NICE의 지침에 의하면, 비만의 치료를 위한 과체중과 비만 관리 전략으로 생활 양식에 대한 중재, 행동 중재, 신체적 활동, 식이에 대한 조언, 약물 중재, 수술적 치료를 제시하고 있다(Kim et al., 2013; NICE, 2006). NICE의 지침(NICE, 2006)에 제시된 임상 실무 경로(clinical care pathways)에 의하면, 아동의 경우에는 신체적 활동의 증가, 식이 행위의 향상, 건강한 식이를 격려하기 위한 행동 치료를 사용한 복합적인 중재를 제공할 것을 제시하고 있다. 성인의 경우에는 BMI, 허리 둘레, 동반 질환 여부와 잠재적인 건강 이득에 따른 비만의 위험 수준에 근거하여 관리 강도가 결정된다고 하였다(NICE, 2006).

NICE의 지침(NICE, 2006)에 제시된 식생활 지침으로는 건강한 식생활 조언(healthy eating advice), 1일 600kcal 미만의 저칼로리 다이어트 [600kcal/day deficit or low-fat diet, very-low-calorie diet(VLCD)], 단백질 절약형 변형단식[protein-sparing modified fast(PSMF)], 저탄수화물식(low-carbohydrate), 고도의 불포화지방 식이(high-monounsaturated-fat-diet), 염분 제한(salt restriction)이 있다(NICE, 2006). 한국인의 비만관리지침(대한내분비학회, 2010)의 내용을 살펴보면, 효과적인 체중감량을 위해 열량과 지방, 당질의 섭취를 제한하고, 열량 제한에 따른 체단백 손실과 단백질 부족을 예방하기 위해 단백질 섭취를 권장하며, 비타민과 무기질 섭취가 부족하지 않도록 보충할 것을 권장하고 있다. 음주 역시 제한해야 하며, 개별화된 영양교육을 지속적이고 체계적으로 시행해야 한다고 제시하고 있다(대한내분비학회, 2010). 신체적 활동은 단독으로 진행하거나 식사 치료와 병행하였을 때에 체중감소에 기여한

다(NHLBI, 1998; 대한내분비학회, 2010). 비만인은 운동 전에 운동 능력의 파악과 운동 실시 여부를 결정하기 위해 운동 전 건강검사를 권장하며, 운동처방은 운동형태, 강도, 시간, 빈도, 기간 등으로 적절하게 구성되어야 한다(대한내분비학회, 2010). 운동유형에는 유산소운동, 근력운동 및 유연성 운동 등의 복합적인 방법들이 권장된다(대한내분비학회, 2010).

평생 지속되어야 하는 체중조절을 위해서는 식습관을 비롯한 생활습관의 개선이 필요하며, 비만 치료에서 가장 효과적인 방법은 행동치료를 포함한 식사 조절과 운동치료의 병합치료이다(대한내분비학회, 2010). 비만치료에 흔히 사용되는 행동치료 기법에는 자극 조절기법, 식사행동조절, 보상제도(양성 강화물의 사용), 자기관찰, 영양교육, 신체활동조절, 대체행동기법, 인지적 재구성기법 및 사회적 지지의 이용 등이 있다(대한내분비학회, 2010; 대한비만학회, 2010; 최중명, 2011). 대한비만학회 Fact Sheet(대한비만학회, 2010)에는 약물 요법은 비만의 기본적인 치료 방법인 식사, 운동 및 행동수정 요법에 대한 보조적인 치료법이 제시되어 있다(대한내분비학회, 2010). 비만치료제에는 장기 사용이 가능한 시부트라민(sibutramine)과 올리스타트(orlistat), 그리고 단기 사용하는 펜터민(phentermine) 등이 있다(대한내분비학회, 2010; 최중명·김춘배, 2011). 시부트라민은 중추신경계에 작용하는 식욕억제제이다(대한비만학회, 2010). 일반 대중들 사이에서 비만 또는 과체중을 조절하고자 한약이나 건강보조식품 등이 사용되고 있다. 고도비만은 체형의 변화 뿐 아니라 동반 합병증으로 사망률을 증가시키고 삶의 질을 파괴시켜 심각한 사회문제의 하나로 볼 수 있다(대한비만학회, 2010). 고도비만의 비약물적 치료 성적은 양호하지 못하고, 다시 체중증가를 경험하거나 비수술

적 치료에 대한 부작용으로 새로운 질병을 경험하기도 한다(대한비만학회, 2010). 이러한 고도비만에 대해 수술적 치료를 시행한다(대한비만학회, 2010). 복강경 수술이 고도비만수술에 도입되면서 비만 대사수술의 새로운 장을 열고 있다(대한비만학회, 2010). 본 연구에서는 비만의 치료에 대한 중분류를 식사치료, 운동치료, 행동치료, 약물치료(제품), 수술(시술)로 분류하였고 해당되는 소분류 영역을 추출하였다.

〈표 6-7〉 치료 영역 분류틀

대분류	중분류	소분류
치료	식사	식사요법(단일식품 다이어트, 칼로리 제한 다이어트, 저탄수화물 다이어트 등), 식이요법 교육, 제한식품(열량, 지방, 당질, 음주 등), 허용식품(단백질, 비타민, 무기질 등), 저칼로리 식사(채식, 잡곡 등), 저지방 식사(기름기 없는 고기, 저지방식, 저지방 우유 등), 고단백 식사(단백질 파우더, 계란 등), 섬유질(야채, 과일, 샐러드, 잡곡류 등), 비타민(야채, 해조류 등), 무기질(칼슘, 철, 인, 요오드, 나트륨 등)
	운동 (신체적 활동)	저항성 운동(바벨, 운동밴드, 짐볼 등), 유산소 운동(자전거타기, 에어로빅 댄스, 조깅 등), 복합성 운동, 유연성 운동(스트레칭, 요가 등), 운동처방, 신체활동 안내
	행동치료	자극조절기법, 식사행동조절, 보상, 자기관찰(일기, 수첩), 영양교육, 신체활동조절, 문제 해결, 대체행동 기법(산보, 영화, 전화 등), 인지적 재구조화, 사회적 지지 프로그램(동호회, 자조모임 등), 기타 행동치료(건강한 생활방식 유지, 자가 감시, 스트레스 관리, 생활습관 교정 등)
	약물치료/제품	비만치료제(시부트라민, 오리스타트, 펜터민 등), 한약, 건강기능식품 등
	수술/ 시술	수술(복강경 수술, 조절형 위밴드 설치술, 루와이 우회술 등), 시술(지방흡입술, 주사 등)

8) 예방

비만의 예방을 위한 방법으로는 ‘치료’ 분류에서 기술하였던 식사, 운동, 행동치료와 함께 체중유지(에너지 균형)와 인식고취를 들 수 있다(NICE, 2006). 비만은 완치되는 것이 아니라 평생 조절해야 하는 병이므로 체중 감량에 초점을 맞추기보다는 식사를 비롯한 한 생활습관을 교정하여 오랫동안 감량된 체중을 유지하고, 비만으로 인한 대사 질환을 예방하는 것이 진정한 체중감량의 목표임을 인식시키는 것이 중요하다(대한비만학회, 2010; 최중명·김춘배, 2011). 체중 유지 전략에 대한 근거를 비만치료지침 2010 권고안(대한내분비학회, 2010)에서 제시하고 있는데, 그 내용은 다음과 같다. : 실현 가능한 체중 감량 목표를 정한다, 오랫동안 지속적으로 체중을 감량한다, 적절한 신체활동을 유지한다, 행동치료(식사조절, 스트레스 관리 등)를 병행한다, 체중 감량 이후에도 지속적으로 치료한다. 이러한 내용에 근거하여, 체중 유지(에너지 균형)에 대한 소분류 영역을 추출하였다. NICE(NICE, 2006)의 지침에서는 비만 예방의 인식 고취를 위한 방법으로, 이전의 연구들에서 교육 중재 또는 multi-component 중재 형식의 캠페인을 제공하였던 근거들을 기술하였다. 비만과 관련된 정보나 지식이 여러 주체에서 제공되고 있어 일관성을 가지지 못하고 양산되고 있으며, 신뢰성 있는 정보를 얻는 것이 쉽지 않다(김혜련 외, 2009). 본 연구에서는 비만에 대한 효과적이고 대대적인 캠페인, 교육 홍보자료와 비만예방 매뉴얼을 개발하여 보급하고 비만 예방에 대한 인식을 강조할 필요성이 있다(김혜련, 2009). 이에 비만의 예방에 대하여, 체중유지와 인식 고취에 대한 항목을 추가하고 치료에서 설명한 식사, 운동, 행동 치료와 함께 중분류로 분류하였고,

해당 소분류 영역을 추출하였다.

〈표 6-8〉 예방 영역 분류틀

대분류	중분류	소분류
예방	식사	식사요법(단일식품 다이어트, 칼로리 제한 다이어트, 저탄수화물 다이어트 등), 식이요법 교육, 제한식품(열량, 지방, 당질, 음주 등), 허용식품(단백질, 비타민, 무기질 등), 저칼로리 식사(채식, 잡곡 등), 저지방 식사(기름기 없는 고기, 저지방식, 저지방 우유 등), 고단백 식사(단백질 파우더, 계란찜, 삶은 계란 등), 섬유질(야채, 과일, 샐러드, 잡곡류 등), 비타민(야채, 해조류 등), 무기질(칼슘, 철, 인, 요오드, 나트륨 등)
	운동 (신체적 활동)	저항성 운동(바벨, 운동밴드, 짐볼 등), 유산소 운동(자전거타기, 에어로빅 댄스, 조깅 등), 복합성 운동, 유연성 운동(스트레칭, 요가 등), 운동처방, 신체활동 안내
	행동치료	자극조절기법, 식사행동조절, 보상, 자기관찰(일기, 수첩), 영양교육, 신체활동조절, 문제 해결, 대체행동 기법(산보, 영화, 전화 등), 인지적 재구조화, 사회적 지지 프로그램(동호회, 자조모임 등), 기타 행동치료(동기 부여, 건강한 생활방식 유지, 자가 감시, 스트레스 관리 등)
	체중유지	체중 감량 목표, 신체활동 유지, 행동치료 유지, 식사 조절, 스트레스 관리, 정기적인 체중 측정, 정기적 추적관리
	인식 고취	캠페인(식사, 운동, 행동치료, 생활습관 교정)

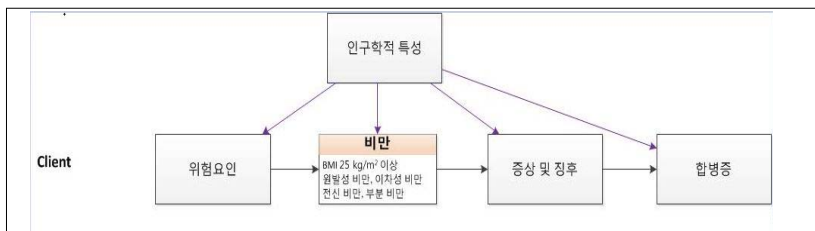
5. 비만 관리 온톨로지

본 연구에서 개발된 비만 관리 온톨로지는 [그림 6-1]과 같다. 비만 관리 온톨로지는 비만(정의 및 분류), 대상자의 인구학적 특성, 위험요인, 증상 및 징후, 합병증, 예방, 진단과 치료의 영역으로 구성되었다. 이는 대상자(client)가 보유한 비만 관련 특성을 설명하는

부분과 건강관리 제공자(healthcare provider)에 의해 제공될 수 있는 비만 관리 서비스를 설명하는 부분으로 나누어 설명할 수 있다.

대상자 측면(client)에서 비만은 어떤 인구학적 특성을 가진 대상자가 식이 및 식사습관, 신체적 활동, 유전적 요인 등을 포함하는 비만 관련 위험요인이 존재할 때에 비만으로 이행되어, 신체적 또는 정신적 증상을 발현시키며, 비만 정도가 심각해지는 경우에는 증상 및 징후를 넘어 비만으로 인한 여러 합병증을 나타나게 할 수 있다. 즉 비만과 증상 및 징후, 합병증의 발생 과정에는 시간적 경과에 따른 변화가 있다고 볼 수 있다. 비만의 위험요인, 증상 및 징후, 합병증은 대상자가 보유한 특성이며, 이는 대상자의 인구학적 특성에도 영향을 받는다. 앞서 대상자의 인구학적 특성 영역 분류틀의 설명에서 제시하였던 것처럼 대상자의 생애주기에 따라서 체중 증가가 최대화되는 시기가 있고 (NICE, 2006) 이러한 특성에 따라 비만 발생 가능성의 정도가 달라지기 때문이다(그림 6-1 참조).

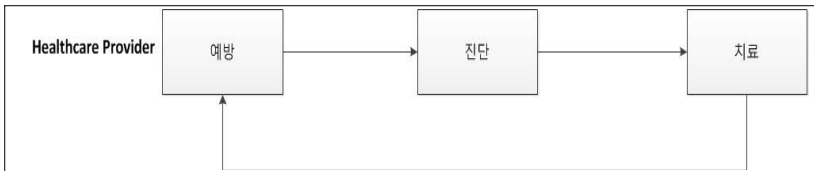
[그림 6-1] 비만 온톨로지 인구학적 특성



건강관리 제공자(health care provider)가 비만 관리 서비스를 제공하는 측면에서 비만 관리 개념을 살펴보았을 때에는 비만의 예방과 진단, 치료 개념을 생각할 수 있다. 비만의 예방은 대상자가 보유하고 있는 비만의 위험요인을 예방하는 것이며, 비만이 진단된 이후에

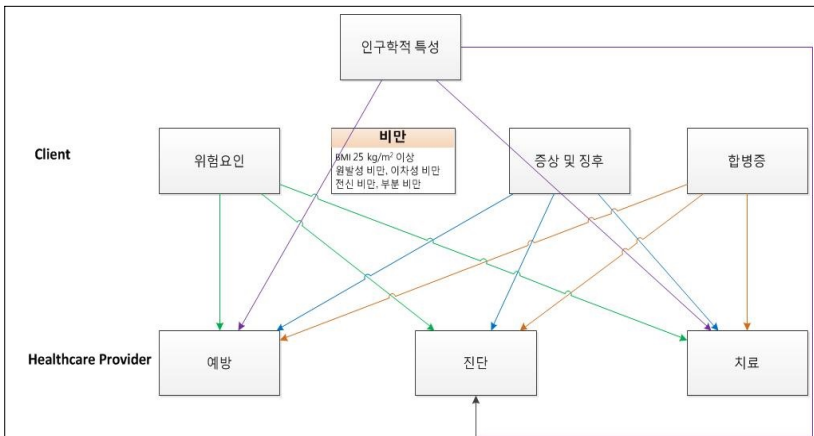
는 치료의 과정을 거치기 때문에 비만 관리 서비스는 단계적 흐름(process)을 가진다고 볼 수 있다. 치료가 끝난 이후에는 다시 예방 단계로 갈 수 있다(그림 6-2 참조).

[그림 6-2] 비만 관리 서비스 단계적 흐름



비만의 진단, 예방과 치료 방법은 대상자가 보유한 인구학적 특성, 위험요인, 증상 및 징후, 합병증 유무의 영향을 받는다(그림 6-3 참조).

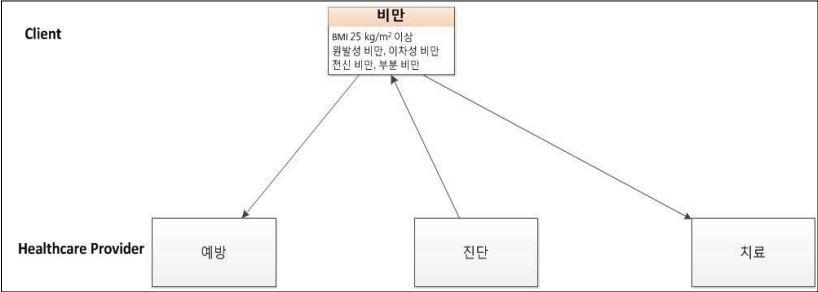
[그림 6-3] 비만 온톨로지 비만의 진단, 예방과 치료 프로세스



대상자의 인구학적 특성과 비만 관련 특성(위험요인, 증상 및 징후, 합병증 유무)의 영향을 받아 비만의 정도와 비만 분류 기준이 진단된 이후

에는, 대상자의 상태에 따라 적합한 치료 또는 예방 방법이 선택될 수 있다(그림 6-4 참조).

[그림 6-4] 비만 대상자의 상태에 따른 치료 또는 예방 진단

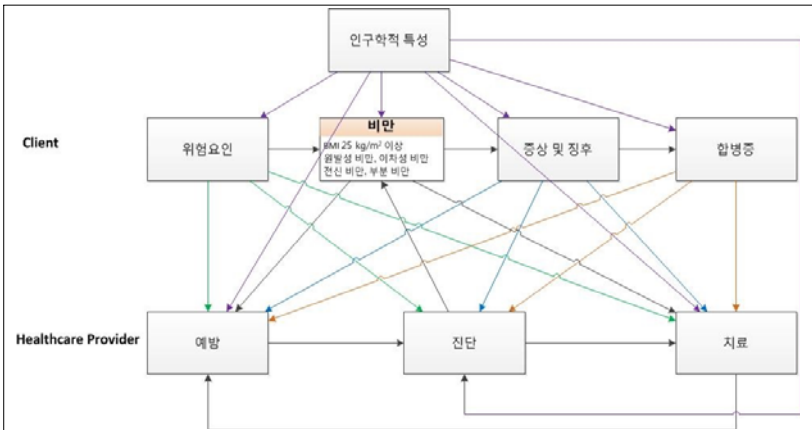


<표 6-9> 비만 온톨로지 분류에 따른 영역수준

대분류	중분류	소분류 1	소분류 2	동의어 유의어	영역 수준
위험 요인				위험인자, 위험요소	대분류(위험요인)
	식이 및 식사 습관			식사, 식이, 식이요법, 식사습관	대분류(위험요인)>중분류(식이 및 식사습관)
		고지방식이		고지방식, 지방 과다섭취	대분류(위험요인)>중분류(식이 및 식사습관)>소분류1(고지방식이)
			지방	기름기, 기름부위, 유지류	대분류(위험요인)>중분류(식이 및 식사습관)>소분류1(고지방식이)>소 분류2(지방)
			과자	과자류, 비스킷, 스낵	대분류(위험요인)>중분류(식이 및 식사습관)>소분류1(고지방식이)>소 분류2(과자)
			삼겹살구이	돼지고기 구이, 삼겹살	대분류(위험요인)>중분류(식이 및 식사습관)>소분류1(고지방식이)>소 분류2(삼겹살구이)
			튀김	튀김요리, 튀김류	대분류(위험요인)>중분류(식이 및 식사습관)>소분류1(고지방식이)>소 분류2(튀김)
			...		이 외 분류를 포함할 수 있음.

지금까지 설명한 개념 간 관계를 전체적으로 표시하여 개발된 비만 관리 온톨로지는 [그림 6-5]과 같다.

[그림 6-5] Ontology for obesity management



6. 비만 관리 용어체계

비만 관리 주제를 설명하는 분류틀에 해당하는 용어에 대하여, 대분류-중분류-소분류의 각 영역 수준별로 용어를 추출하여 영역 수준별로 제시하였다. 각 용어별로 인터넷 검색과 선행 문헌 검색 등의 방법을 이용하여 동의어와 유사어의 예시를 제시하였고, 그 결과의 예시는 <표 6-9>와 같다. 전체 영역의 용어체계는 부록 3에 제시하였다²³⁾.

23) 부록 3참조

7. 결론

본 연구에서는 소셜 미디어 상의 비만 주제와 관련된 메시지를 수집하여 분석하기 위한 분석틀을 마련하고자 하였다. 그 결과로 비만 관리 주제를 설명하는 대분류-중분류-소분류의 영역 분류틀과 각 영역간의 관계를 설명하는 비만관리 온톨로지를 개발하였다. 이와 함께 각 영역을 설명하는 선행 문헌과 인터넷 검색 진술문으로부터 용어 자료를 수집하여 각 용어에 대한 동의어와 유사어와 용어의 영역 수준을 설명하는 용어체계를 개발하였다. 본 연구에서 개발된 분석틀에서 비만 관리를 설명하는 영역은 비만(정의 및 분류), 인구학적 특성, 위험요인, 증상 및 징후, 합병증, 예방, 진단과 치료로 구성되었다. 이는 대상자 측면의 영역과 건강관리 제공자 측면에서의 비만 관리 서비스 차원으로 나누어 설명할 수 있었다. 비만 현상은 시간적인 경과에 따라 위험요인으로부터 합병증으로 진행되었고, 비만 관리는 예방에서 진단 치료로, 그리고 다시 예방으로 순환되는 과정적(process) 특성이 있었다. 본 연구에서 개발된 개념 및 용어 수준에 따른 분석틀은 비만의 대상자, 위험요인, 증상, 합병증, 진단 및 치료와 예방 개념을 설명하기 위하여, 소셜 미디어상의 빅데이터를 보다 체계적으로 수집하고 분석하는 데에 활용될 수 있을 것이다.

제2절 소셜 빅데이터를 이용한 비만 담론 분석

2011년 현재 우리나라 19세 이상 성인 비만율(BMI 25이상)은 남자는 35.1%, 여자는 27.1%로 전체 인구의 31.4% 수준으로 2005년 이후 급속한 증가세를 보이고 있으며, 건강생활실천을 위해서는 비만의 폐해 및

실태에 대한 일반 국민의 인식 제고 및 자가관리 능력의 필요성이 제기되고 있다. 본 연구에서는 소셜 빅 데이터 분석에 기반을 두어 ‘비만 언급에 대한 패턴과 원인’을 찾고 대응체계 수립을 위한 시사점을 도출코자 한다.

1. 조사 설계

본 연구는 2011년 1월 1일부터 2013년 12월 31일까지 최근 3년 동안 블로그, 카페, 게시판 및 주요 커뮤니티, SNS, 인터넷 뉴스 및 미디어 사이트 채널 등에서 ‘비만’, ‘다이어트’라는 키워드를 수집하였다. 비만과 관련된 세부 키워드의 수집은 앞 절에서 개발한 비만 온톨로지를 사용하였다.

〈표 6-10〉 소셜 빅데이터 활용한 비만 담론 분석 조사설계

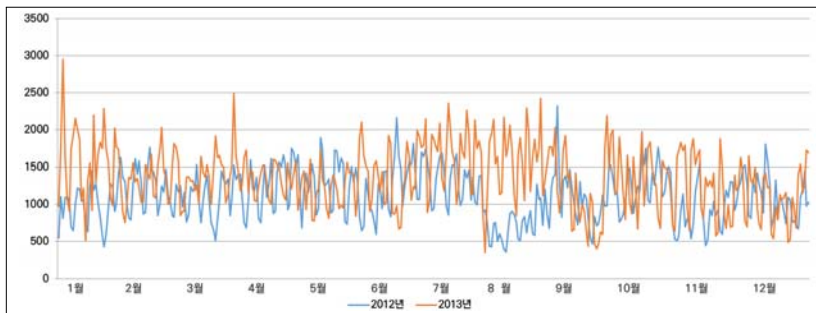
분석기간	2011.01.01.~2013.12.31.(3년)
조사방법	SKT Smart Insight 플랫폼을 이용한 Social Media Research (Online Buzz 분석)
분석 대상 키워드	비만, 다이어트
분석 대상	총 1,207,531건(2011년: 300,867건, 2012년: 411,011건, 2013년: 495,653건)
수집 채널	1. 뉴스/미디어: 온라인 상에 게재되는 217개 뉴스 사이트 2. 블로그: 네이버, 다음, 네이버, 티스토리 3. 카페: 네이버카페, 다음 카페 4. SNS: 트위터, 미투데이 5. 게시판: 네이버 지식인, 네이버 지식, 네이버 톡, 네이버 판, 다음 지식인, 디시인사이드_갤러리, 일간베스트, 다음 미즈백_생활에피소드, 뷰티톡, 뷰티블, SLR 클럽

2. 조사 분석 결과

가. 버즈 현황 분석

[그림 6-6]과 같이 비만 및 다이어트에 관한 Buzz는 특정 시즌이 아닌 연중 나타나고 있으며, 특히 여름 보다는 겨울에서부터 봄 시즌에 비만과 다이어트에 관한 언급이 증가하는 추세로 나타났다.

[그림 6-6] 연도별 버즈 현황: 2012-2013



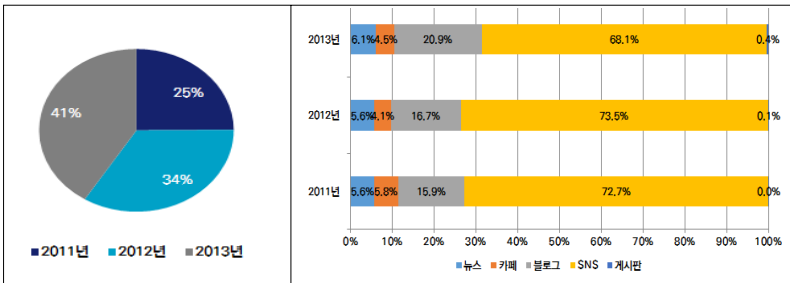
주: 트위터 문서의 수집은 2011년 9월부터 시작되어 2011년 Buzz Trend에 Bias를 줄 수 있어 전체 Buzz 분석에서는 2011년 데이터는 제외

나. 버즈 점유율 분석

비만 및 다이어트에 관한 관심이 증가하면서 매년 온라인 상에서 언급량은 증가하고 있는 것으로 분석되었다. 2011년부터 2013년 까지 3개년 모두 SNS에서의 언급이 2012년 73.5%, 2013년 72.7%, 2011년 68.1%순으로 높았으며, 그 다음으로 블로그에서 많이 언급된 것으로 나타났다. SNS의 경우 비만 및 다이어트에 대한 걱정, 감정 등을 짧게 남기는 반면 블로그, 카페 등의 채널에서는 다이어트 방법, 후

기 등을 작성하고 공유하는 형태의 버즈가 많은 것으로 분석되었다 (그림 6-7 참조).

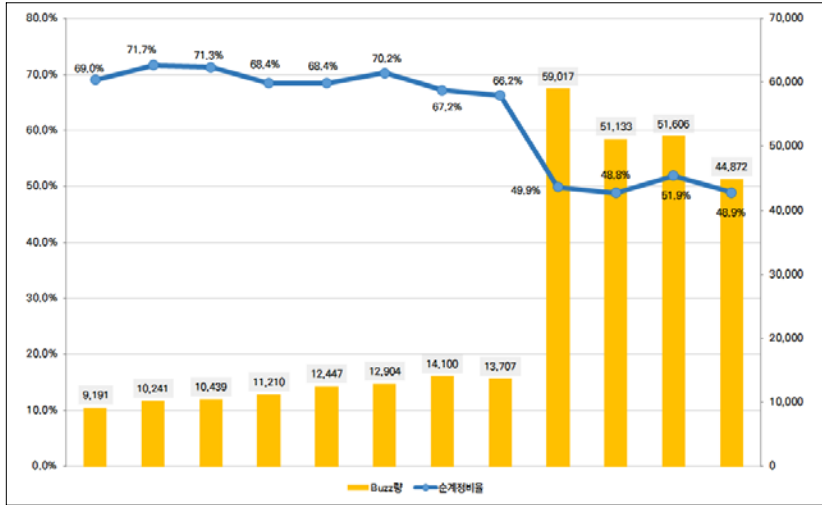
[그림 6-7] 연도별 버즈 점유율 현황: 2012-2013



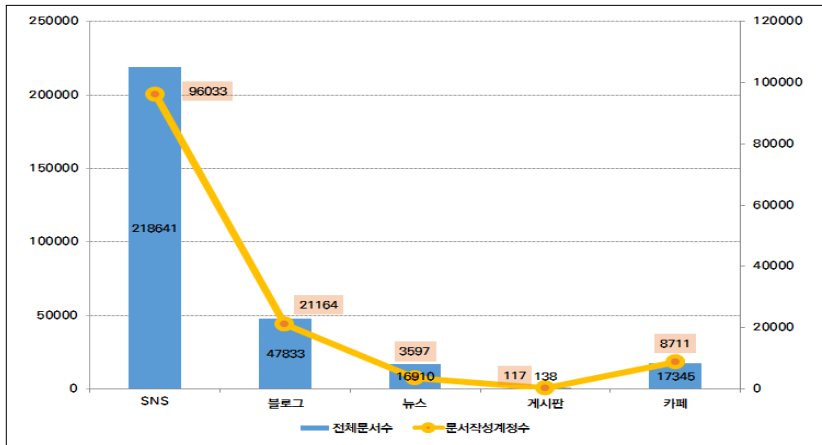
다. 계정활동 분석

2011년 버즈량과 순계정비율은 [그림 6-8]과 같이 2011년 계정 활동 분석 결과 트위터를 제외한 블로그, 게시판, 카페의 버즈량이 많았던 1월~8월의 순계정 비율은 약 70% 정도였으나, 9월 이후 트위터 문서가 다수 발생하며 작성 계정의 비율은 감소한 것으로 분석되었다. 전체문서에 따른 작성계정수는 [그림 6-9]와 같이 게시판과 카페의 경우 발생 버즈 대비 언급자의 수가 많았으나 트위터 작성자의 경우 언급자가 전체 문서 대비 44%로 특정 계정만이 비만과 다이어트에 대한 언급을 하고 있는 것으로 나타났다.

[그림 6-8] 2011년 버즈량과 순계정 비율



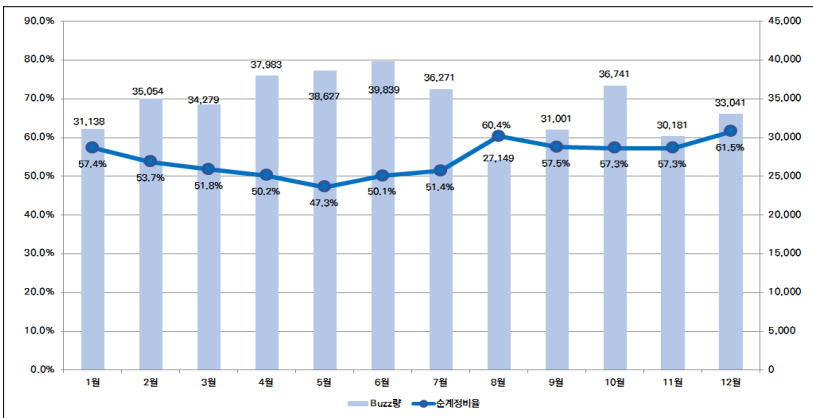
[그림 6-9] 2011년 전체 문서수와 문서작성계정 수



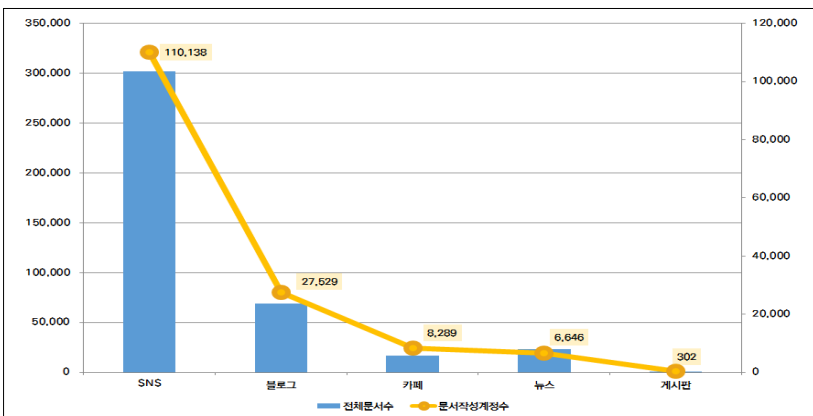
2012년 버즈량과 순계정비율은 [그림 6-10]과 같이 2012년에는 월 평균 35,000건의 버즈, 작성 계정수는 월 평균 15,000건이 발생

하였다. 특히 SNS의 채널의 경우 순계정 비율이 30% 미만으로 매우 낮아져 특정 계정의 반복 언급이 있는 것으로 분석되었다(그림 6-11 참조).

[그림 6-10] 2012년 버즈량과 순계정 비율

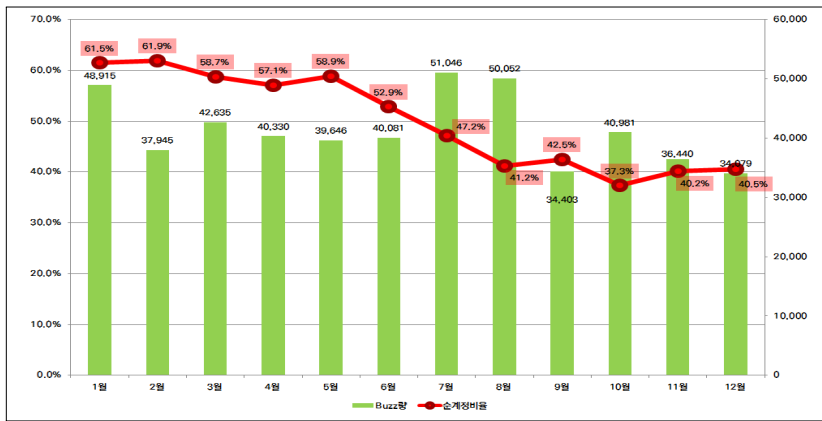


[그림 6-11] 2012년 전체 문서수와 문서작성계정 수

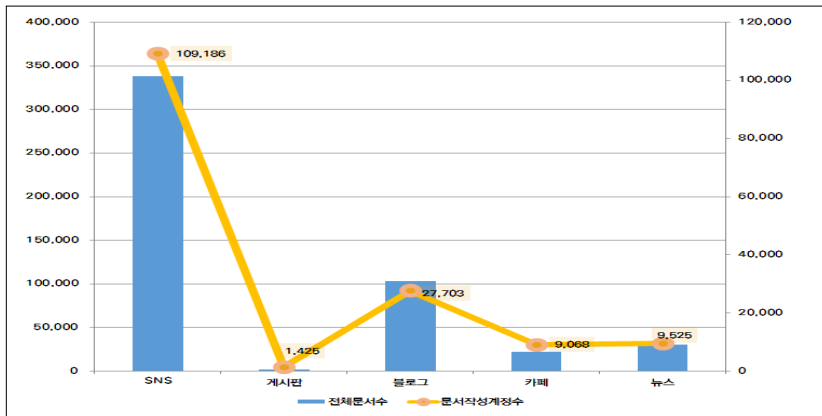


2013년 버즈량과 순계정 비율은 [그림 6-12]와 같이 비만관련 버즈량이 증가함에도 불구하고 순계정 비율과 작성 계정 수가 감소한 것으로 분석되었다. 버즈량과 순계정 분석 결과 비만에 대한 다수의 관심과 문서의 증가 사이의 인과관계는 크지 않는 것으로 나타났다(그림 6-13 참조).

[그림 6-12] 2013년 버즈량과 순계정 비율



[그림 6-13] 2013년 전체 문서수와 문서작성계정 수



라. 트위터 유형 분석

연도별 트위터 유형 분석 결과 2011년에는 리트윗(RT, 전파형)이나 정보링크 보다는 비만에 대한 직접 멘션을 남기는 독백형 및 대화형이 30%이상의 점유율을 차지하고 있으며, 2012년은 직접 멘션을 남기는 독백형, 대화형이 약 50%를 차지했으며, 콘텐츠 링크가 멘션 내에 포함되어 있는 정보링크형도 25% 차지하였다. 2013년 트위터의 언급 형태는 전파형(RT)을 중심으로 다이어트 약품, 운동 업체 등의 이벤트 버즈가 많은 부분을 차지한 것으로 분석되었다.

[그림 6-14] 연도별 트위터 유형 분석



마. 키워드 분석

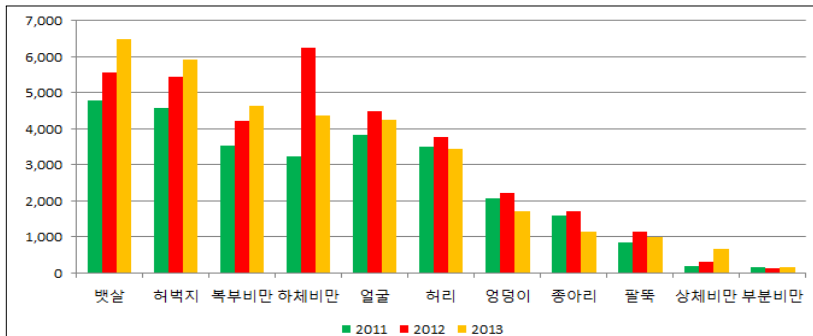
1) 연도별 비만 부위별 언급 키워드 분석

비만 언급자들의 부위별 언급 키워드 분석을 한 결과 <표 6-11>과 같이 뱃살과 허벅지는 언급이 상위에 있는 것으로 나타났으며, 2011년 하체비만 언급이 3,236회로 전체 언급의 6위였던 반면 2012년 6,239회(1위) 언급, 2013년에는 4,369회 언급(4위)으로 상위 키워드로 분석되었다.

〈표 6-11〉 연도별 비만 부위별 언급 키워드 현황

순위	2011년		2012년		2013년	
	키워드	빈도	키워드	빈도	키워드	빈도
1	뱃살	4,795	하체비만	6,239	뱃살	6,482
2	허벅지	4,572	뱃살	5,553	허벅지	5,909
3	얼굴	3,823	허벅지	5,450	복부비만	4,637
4	복부비만	3,528	얼굴	4,500	하체비만	4,369
5	허리	3,493	복부비만	4,229	얼굴	4,254
6	하체비만	3,236	허리	3,765	허리	3,437
7	엉덩이	2,059	엉덩이	2,213	엉덩이	1,703
8	종아리	1,599	종아리	1,699	종아리	1,139
9	팔뚝	832	팔뚝	1,134	팔뚝	1,001
10	상체비만	189	상체비만	316	상체비만	656
11	부분비만	170	부분비만	126	부분비만	167

[그림 6-15] 연도별 비만 부위별 언급 키워드 트렌드



2) 연도별 비만 관련 약물치료 언급 키워드 분석

비만과 관련하여 치료 관련 언급된 키워드를 분석한 결과는 〈표 6-12〉와 같다. 가장 많이 언급된 건강식품으로는 ‘허벌라이프’, ‘쉐이크’ 등의 대체 식품 종류로 나타났으며, 한약, 식욕억제제, 비만치료제 등 식욕을 억제하는 약물도 다수 언급이 되었다. 뿐만 아니라 다이어트 시 건강 유지를 위한 약물, 기능식품에 대한 언급도 나타났다.

〈표 6-12〉 연도별 비만 치료 방법(약물치료) 언급 키워드 현황

순위	2011년		2012년		2013년	
	키워드	빈도	키워드	빈도	키워드	빈도
1	허벌라이프	3,305	쉐이크	2,463	쉐이크	3,237
2	쉐이크	2,107	허벌라이프	2,316	한약	2,586
3	한약	784	알로에겔	1,200	허벌라이프	1,544
4	알로에겔	532	건강기능식품	662	알로에겔	571
5	식욕억제제	143	한약	639	건강기능식품	174
6	비만치료제	92	비만치료제	309	식욕억제제	168
7	건강기능식품	86	시부트라민	80	비만치료제	146
8	시부트라민	38	식욕억제제	70	시부트라민	80
9			펜터민	6		
10			orlistat	6		
11			phentermine	4		

주: 의약품의 국문명과 영문명의 빈도는 합하여 표기함.

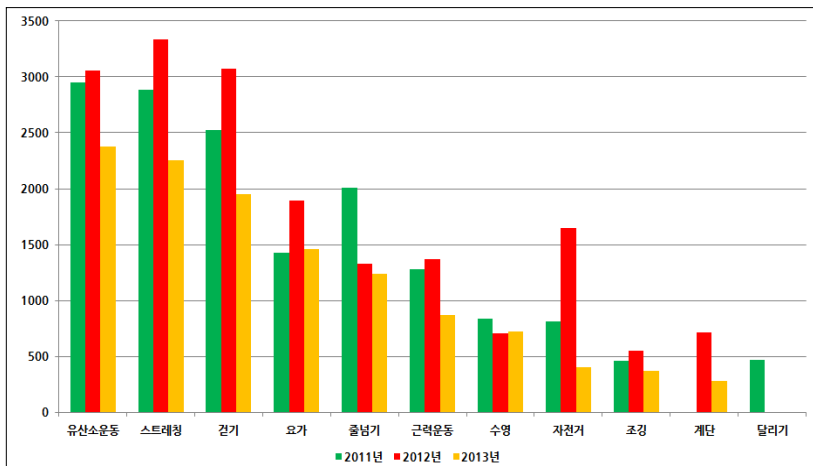
3) 연도별 비만 관련 운동치료 언급 키워드 분석

비만과 관련하여 비만 치료를 위한 운동치료 키워드 분석 결과는 〈표 6-13〉과 같다. 비만 치료를 위한 운동은 유산소 운동, 스트레칭 언급 빈도가 가장 높았으며, 운동 종목 역시 걷기, 줄넘기, 계단, 수영, 조깅 등 유산소 운동이 상위에 언급 되었으며, 실내에서 혹은 운동기구 없이 간편하게 할 수 있는 운동이 비만과 함께 언급 된 것으로 분석되었다(그림 6-16 참조).

〈표 6-13〉 연도별 비만 치료 방법(운동치료) 언급 키워드 현황

순위	2011년		2012년		2013년	
	키워드	빈도	키워드	빈도	키워드	빈도
1	유산소운동	2,948	스트레칭	3,333	유산소운동	2,380
2	스트레칭	2,881	걷기	3,068	스트레칭	2,255
3	걷기	2,523	유산소운동	3,059	걷기	1,953
4	줄넘기	2,010	요가	1,895	요가	1,459
5	요가	1,425	자전거	1,646	줄넘기	1,242
6	근력운동	1,282	근력운동	1,367	근력운동	868
7	수영	836	줄넘기	1,330	수영	727
8	자전거	813	계단	714	자전거	404
9	달리기	469	수영	707	조깅	375
10	조깅	461	조깅	556	계단	284

〔그림 6-16〕 연도별 비만관련 운동치료 언급 키워드 트렌드



4) 연도별 비만 관련 식사치료 언급 키워드 분석

〈표 6-14〉과 같이 비만과 관련하여 식이 요법과 관련된 언급 키워드는 열량, 저칼로리 등 음식의 칼로리 관련 문서가 다수 언급 되었다. 지난 3

년동안 언급된 상위 키워드로는 ‘지방’이 상위 키워드로 가장 높은 빈도를 보였다. 다음으로는 원푸드 다이어트, 덴마크 다이어트, 황제 다이어트 등과 각 다이어트에 필요한 음식이 자주 언급 된 것으로 나타났다.

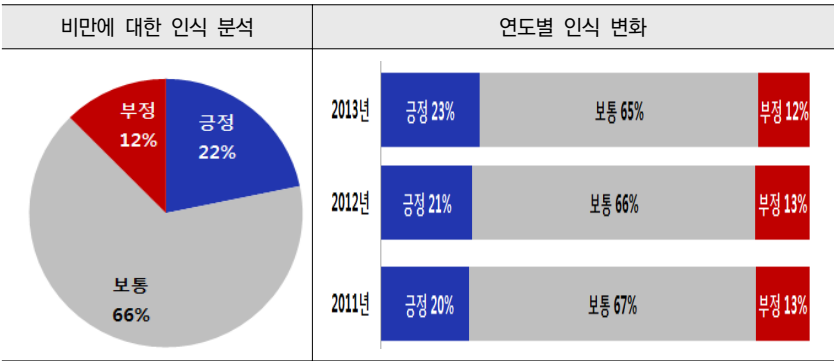
〈표 6-14〉 연도별 비만 치료 방법(식사치료) 언급 키워드 현황

순위	2011년		2012년		2013년	
	키워드	빈도	키워드	빈도	키워드	빈도
1	지방	10,969	지방	13,066	지방	12,786
2	가지	7,005	가지	7,119	가지	7,346
3	단백질	6,801	단백질	7,081	단백질	6,687
4	과일	5,015	과일	5,691	식이요법	5,137
5	식이요법	4,553	식이요법	4,649	비타민	4,577
6	야채	3,479	비타민	4,162	과일	4,513
7	샐러드	3,253	샐러드	3,736	열량	3,813
8	우유	3,139	채소	3,398	채소	2,975
9	채소	3,105	열량	3,283	야채	2,431
10	비타민	3,068	야채	3,271	샐러드	2,426

바. 감성분석

감성분석은 전체 수집 문서 1,207,531건 중에서 비만에 대한 긍정, 보통, 부정적 인식이 명확한 표현어를 대상으로 분석하였다. 비만, 다이어트에 대한 감성 분석 결과는 [그림 6-17]과 같이 보통 비율이 66%로 가장 높았으며, 부정적인 인식(12%)보다는 긍정적인 인식(22%)의 비중이 10% 가량 더 높은 것으로 분석 되었다. 최근 3년간의 감성 분석을 비교한 결과 긍정적인 비중의 인식이 높아지는 추이를 보이고 있는 것으로 나타났다.

[그림 6-17] 비만에 대한 감성분석 결과(N=1,441,939)



비만과 관련하여 긍정적인 언급 키워드를 분석한 결과 비만과 관련하여 ‘예방하다’, ‘탈출하다’ 등의 비만 문제에 대하여 인식하고 해결하고자 하는 표현어가 도출 되었다. 부정적인 언급 키워드로는 ‘마른 비만’, ‘초 고도 비만’ 등의 유형 및 각종 질병을 일으키는 원인으로써 부정적인 언급이 많은 것으로 나타났다(표 6-15 참조, 그림 6-18 참조).

<표 6-15> 비만에 대한 긍·부정 감성분석 결과(키워드 분석)

순위	긍정	부정
1	비만 예방하다	마른 비만
2	하체비만 탈출하다	비만 원인되다
3	하체비만 다이어트하다	비만 심하다
4	복부비만 해결하다	비만 심각하다
5	하체비만 해결하다	비만 증가하다
6	비만 관리하다	복부비만 심하다
7	비만 탈출하다	비만 증가하다
8	복부비만 빼다	하체비만 심하다
9	비만 벗어나다	복부비만 심각하다
10	복부비만 탈출하다	비만 위험하다

[그림 6-18] 비만에 대한 감성분석 결과(SNS 언급)

긍정 언급 SNS 언급	너무 말랐을 때는 그 나름대로 힘들었는데 살이 찌고 나면 그때가 얼마나 좋았는지 새삼 느끼게 되었다. 이제 복부 비만만 해결하면 된다! 조금만 더 힘내자 ㅋ 우리모두 일찍자고 일찍 일어나서 비만으로 부터 탈출해보아요.^^ 우리가 살이 찌는 이유들...?! 불규칙한 아침식사가 삶을 찌운다?! 해독주스를 통해 비만에서 탈출해봅시다 :-)	비만 판정을 받고 열심히 노력중입니다. 식단을 바꿔도 보고 운동도 열심히 하고 있는데요. 이렇게 비만을 지속적으로 관리할 수 있는 방법으로는 어떤 좋은 방법이 있을까요 cafe.naver.com/hanamein 적절한 비만의 예방을 위해선 나이가 들수록 운동과 식사요법 등의 생활습관 교정에 더욱 노력을 기울일 필요가 있습니다. 여러분도 건강관리에 신경쓰실거죠? 할머니 소박한 밥상차림이, 비만과 당뇨예방한다 durl.me/58szqp
	이거 걱정 아니데? 마른비만..임니다만? 뱃살만 쫌나쥌 비만이 위험한 이유, 부모 모두 비만이면 자녀도 위험: 최근 포털사이트에는 비만이 위험한 이유가 주요 검색어에 올라왔다. 비만이 위험한 이유는 비만 자체가 각종 성인병의 원인이기 때문이다. 비만인 ... bit.ly/1eLgzqc 88%는 과식 62%는 패스트푸드, 60%는 설탕이 함유된 음료를 비만의 원인이 된다고 말했다. newspeppermint.com/2013/12/12/hea...	비만=게으르다// 의 등호가 성립하기 때문에 게으름에 대한 비난이 되기도 하더라. 근데 비만인 친구들이 막 게으르고 안씻고 그런 경우가 왕왕있기는 해. 특히 복부비만. 다른데는 말랐는데 배만 튀어나왔다는가... 비만은 마른비만이 더 심각한거 다 아시죠? 건강검진 결과가 안좋아서요 ㅠ 모든 병은 비만과 관련있잖아요.. 거기다 전 쫌나 쫌똥함.. 요새 초고도 비만인 듯

다이어트에 대한 긍정 및 부정적인 언급 키워드를 분석한 결과에서도 다이어트 목표 및 방법에 관하여 언급하며 다이어트 성공에 대한 포부와 다짐에 대한 글의 빈도가 높았던 반면, 부정적인 언급 키워드 분석 결과에서는 무리한 다이어트로 인한 부작용과 다이어트로 인한 고통 및 실패 경험에 대한 부정적 표현어가 빈도가 높은 것으로 나타났다(표 6-16 참조). 다이어트에 대한 긍정적인 멘트로는 ‘성공하다’, ‘건강하다’, ‘좋다’, ‘도움되다’ 등이 상위 키워드로 나타난 반면, 부정적인 멘트로는 ‘무리하다’, ‘실패하다’, ‘잘못되다’, ‘무너지다’ 등의 순으로 부정적인 멘트가 상

위 키워드로 나타났다(그림 6-19 참조).

〈표 6-17〉은 운동에 대한 긍정 및 부정 언급 키워드 분석 결과이다. 운동에 대한 긍정적인 언급 키워드로는 ‘운동 좋다’, ‘살빼는 운동’, ‘운동 꾸준히 하다’, ‘운동 효과적이다’, ‘운동 추천하다’, ‘만들어 주는 운동’, ‘운동 가능하다’, ‘운동 도움 되다’, ‘운동 권유하다’, ‘운동 잘 하다’ 등의 건강한 다이어트 및 비만 예방을 위한 다양한 운동 방법을 공유하는 언급이 대다수인 것으로 나타난 반면에 부정적인 언급 키워드 분석 결과에서는 ‘운동 안 하다’, ‘운동 못하다’, ‘운동 힘들다’, ‘운동 싫어하다’, ‘운동 부족하다’, ‘운동 시간 없다’, ‘운동 과하다’ 등의 운동의 중요성은 인식을 하고 있지만 본인의 의지 및 시간 부족에 대한 어려움을 호소하며 무리한 운동을 자제하는 부정적인 성향의 언급 빈도가 높은 것으로 분석되었다(그림 6-20 참조).

〈표 6-16〉 다이어트에 대한 긍·부정 감성분석 결과(키워드 분석)

순위	긍정	부정
1	다이어트 성공하다	다이어트 무리하다
2	다이어트 건강하다	다이어트 실패하다
3	다이어트 좋다	다이어트 잘못되다
4	다이어트 도움되다	다이어트 무너지다
5	다이어트 효과적이다	다이어트 하지 말다
6	다이어트 효과 있다	다이어트 안되다
7	다이어트 올바르다	다이어트 혹독하다
8	다이어트 효과 빠르다	다이어트 포기하다
9	다이어트 추천하다	다이어트 못하다
10	다이어트 쉽다	다이어트 좋지 않다

[그림 6-19] 다이어트에 대한 감성분석 결과(SNS 언급)

SNS 언급 다이어트	영마 올, 채윤아 여자는 건널 만큼만 먹는 거야~ 영마는 왜 그리 자꾸 드심. ♫ 나 내년엔 꼭 다이어트 성공할꺼야~!! 진짜 건강한 다이어트는 평생하는 것 같다. 올바른 식이습관을 만들어가면서 원래 좋아하던 음식도 자주 즐기는 것, 단기로 쪽쪽 빼는 다이어트밖에 못들어본 사람들은 내가 똥썰 먹으면서 다이어트 일기 쓰는 걸 이해 못한다. 바보들 규칙적인 식사가 다이어트에도 도움이 된다는 거, 알고 있어? 차차, 어서 먹으러가라고!	다이어트 성공해서 이뻐지길~ 건강검진 진행 중- 인생 최고의 똥썰이 시절이고나~(내년부터 다시 건강을 위해 다이어트!! 다이어트 성공해서 이뻐지길~
	기관지 내시경 횡드네요. 다이어트 너무 심하게 하지 마세요. 상관없어 귀여운실수니까/죄구/ 다이어트는 자욱의 시작이지 나는 그 자욱을 내 발로 들어가겠어 다이어트는 포기다 내가 무슨 다이어트를 하겠어 휴	다이어트는 무슨 다이어트입니까, 당신은 지금의 모습으로도 충분히 아름다운 걸요. 그보단 당신의 건강이 우선입니다. 괜찮으시다면, 함께 식사하시겠습니까? 영광입니다. 하아.. 나도 다이어트 해야하는데 망했어.... TT TT 설마 다들 저녁 안먹은건 아니지? 무리한 다이어트는 좋지 않다고!

〈표 6-17〉 운동에 대한 긍·부정 감성분석 결과(키워드 분석)

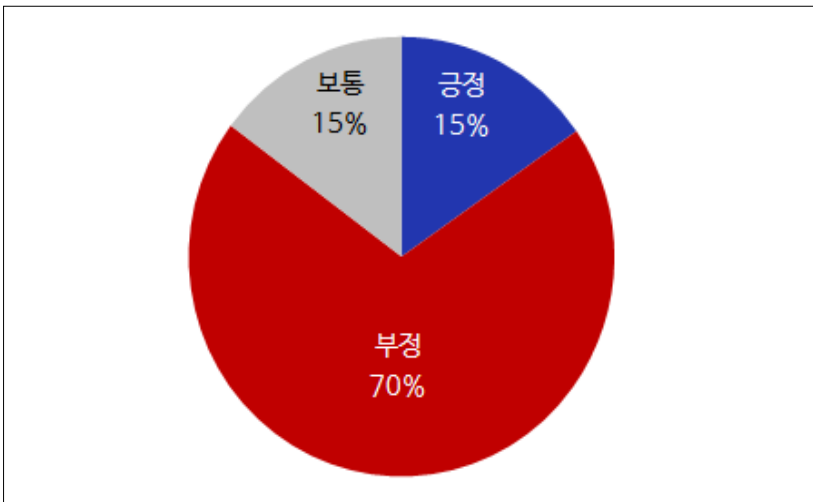
순위	긍정	부정
1	운동 좋다	운동 안 하다
2	살 빼는 운동	운동 못하다
3	운동 꾸준히 하다	운동 힘들다
4	운동 효과적이다	운동 싫어하다
5	운동 추천하다	운동 부족하다
6	만들어 주는 운동	운동 시간 없다
7	운동 가능 하다	운동 과하다
8	운동 도움 되다	운동 하지 못하다
9	운동 권유하다	운동 필요 없다
10	운동 잘 하다	운동 피하다

〔그림 6-20〕 운동에 대한 감성분석 결과(SNS 언급)

긍정 언급 SNS 언급	유산소 운동과 근육운동을 고루 배분해서 운동하는 것이 다이어트를 위한 가장 좋은 방법입니다.	조깅으로 비만예방의 효과를 보기 위해서는 30분 이상 실시해야 합니다. 운동 강도는 '걷기-빠르게 걷기-달리기'의 과정으로 주 3회 이상 실시하는 것이 좋습니다. 속도는 빠를수록 지방제거에 더욱 효과적입니다. ^^ 한국건강관리협회"
	@marrygch나 올 요가가 헬스만큼 다이어트되고 유연성이나 몸 체형변화에도 좋네요. 그리고 스쿼시가 하반신 라인 만드는데 좋고 운동도 많이되요.	다이어트 비법: 적게 먹고 운동 하는 것
	자전거타기는 비만한 사람들의 초기 운동으로서 가장 적당하며 더욱 건강한 생활을 하는데 도움이 됩니다.	비만의 사람이 날씬한 몸매를 만들려면 당연히 많은 운동이 필요하다. 특히, 워킹도 운동 중에서도 근력 운동이 가장 효과가 높다.
부정 언급 SNS 언급	운동은 하되 무리하지 마라. 피로가 쌓이면 오히려 다이어트에 복이 된다.	진짜 다이어트 해야하는데...ㅜ.ㅜ.... 식욕보다도 운동이 중요한데...왜 하기 싫니...
	무리한 운동 다이어트 실패 방법 과다운동증후군 (오버트레이닝)	오늘 부터 운동 다이어트 시작이다ㅇㅈㅇ!! 그냥 빨리걸기했을 뿐인데 힘들다아ㅈ
	나이 드니까 다이어트 운동도 힘들다. 와서 어깨나 좀 주물러라.	와~~ 힘들겠다~~~ 나는 요즘 운동할 시간도 없고 몸이 힘들어서 다이어트는 힘든데.. 화이팅!!

[그림 6-21]은 다이어트로 인한 정신적 증상에 대한 긍·부정 감성분석 결과이다. 전체 수집 문서 18,023건 중에 부정적인 언급에 대한 표현이 70%, 긍정이 15%, 보통이 15% 순으로 나타났다. 이는 스트레스와 관련된 부정적 표현어가 대다수를 차지하면서 부정적인 언급이 더 높은 것으로 판단된다. <표 6-18>은 정신적인 증상에 대한 긍·부정 언급 키워드 결과이다. 긍정적인 키워드로는 ‘스트레스 해소하다’, ‘스트레스 받지 않다’, ‘스트레스 해소 되다’, ‘스트레스 해소시키다’, ‘스트레스에 좋다’, ‘스트레스 효과적이다’ 등의 긍정적인 언급이 많았던 반면, 부정적인 언급으로는 ‘스트레스 받다’, ‘스트레스 과도하다’, ‘스트레스 받게되다’, ‘스트레스 쌓이다’, ‘먹는 스트레스’, ‘스트레스 심하다’ 등의 언급이 높았다. 이는 식욕을 통한 스트레스 해소가 비만의 원인이 되며, 다이어트 실패로 인하여 다시 스트레스가 축적이 된 것이라고 판단 할 수 있다.

[그림 6-21] 다이어트로 인한 정신적인 증상에 대한 감성분석 결과



〈표 6-18〉 다이어트로 인한 정신적인 증상에 대한 긍·부정 감성분석 결과(키워드 분석)

순위	긍정	부정
1	스트레스 해소 하다	스트레스 받다
2	스트레스 받지 않다	스트레스 인하다
3	스트레스 해소 되다	스트레스 과도하다
4	스트레스 해소시키다	스트레스 받게 되다
5	스트레스에 좋다	스트레스 쌓이다
6	스트레스 효과적이다	먹는 스트레스
7	스트레스 사라지다	스트레스 심하다
8	스트레스 완화하다	스트레스 주다
9	고민할 필요없다	스트레스 많다
10	스트레스 이기다	스트레스 되다

3. 결론

소셜 빅데이터를 이용한 비만, 다이어트 담론 분석 결과 ‘비만 및 다이어트’에 관한 버즈 발생은 특정 시즌이 아닌 1년 내내 나타나는 것으로 분석 되었으며, 3개년 동안 해마다 온라인 상의 언급량이 증가한 것으로 나타났다. 3개년 동안 SNS의 언급량이 가장 높았으며, 그 다음이 블로그에서 다수 언급된 것으로 나타났다. SNS 채널의 경우 비만 및 다이어트에 대한 걱정, 감정 등을 짧게 남기는 반면에 블로그, 카페 등의 게시판 형태의 채널에서는 다이어트 방법, 후기 등을 작성하고 공유하는 형태의 버즈가 많이 발생하였다. 트위터 채널에서도 3개년 간 순계정 비율이 약 40%로 특정 계정이 반복하여 비만 및 다이어트에 관한 언급을 하고 있는 것으로 분석 되었다. 트위터를 제외한 블로그, 게시판, 카페 채널은 순계정 비율은 높은 편이었으며 다수의 사람들이 활동 하는 것으로 나타났다. 각 연도별 트윗 유형으로는 2011년과 2012년은 직접 비만에 대한 멘션을 언급하는 독백형, 대화형이 상위를 차지하였던 반면에 2013년에는 트위터 언급 형태는 전파형(RT)를 중심으로 다이어트, 약품, 운동 업체 등의

이벤트 버즈가 상당 부분 차지 한 것으로 분석 되었다. 시간이 지난 수록 비만 치료에 대한 관련 업체의 홍보가 트위터 상에 증가 하는 것으로 나타났다. 키워드 및 감성 분석 결과에서도 비만 치료와 관련된 키워드는 허벌라이프, 셰이크 등의 대체식품 종류와 식욕 억제류 등의 약물이 상위 키워드로 언급 되었으며, 관련된 운동은 특별한 기구가 없어도 손쉽게 할 수 있는 조깅, 유산소 운동이 상위 키워드로 언급이 되었다. 다이어트 방법으로는 꾸준한 노력보다 쉽게 살을 뺄 수 있는 약물, 수술 치료 등에 관심을 보이는 경향이 있었으며, 최근 3년간 비만 및 다이어트에 관한 긍정적인 인식의 비중이 지속적으로 높아지는 것으로 나타났다. 비만을 예방하고 관리하기 위해서는 효과적인 운동, 식이요법을 공유하는 형태의 언급이 많았으며, 다이어트 성공을 다짐하는 멘션도 높았다. 스트레스 해소를 위한 음식 섭취가 비만으로 이어지고 다이어트 실패로 인한 스트레스를 받는 악순환에 관한 언급이 있었다. 이는 계속되는 다이어트 실패로 스트레스가 축적되어 부정적인 언급이 높아진 것이라 판단 할 수 있다.

제3절 소셜 빅데이터를 이용한 청소년 자살위험 예측요인²⁴⁾

1. 서론

우리나라는 자살에 의한 사망률(자살률)은 2012년 인구 10만명 당 29.1명으로 경제협력개발기구 회원국의 평균자살률 12.5명에 비해 회원국 중 가장 높게 나타나고 있으며, 2012년 자살로 인한 총 사망자

24) 본 절은 해외 학술지에 게재하기 위하여 '송태민, 송주영, 안지영, Hayman, 우종민'에 의해 작성된 논으로, 'Social Risk Factor Prediction Utilizing Social Big Data. Working Paper 2014-01, KIHASA'에 수록된 내용임을 밝힌다.

14,160명중 1,632명이 30세 미만으로 자살이 10대와 20대의 사망원인 1위를 차지하고 있다(통계청, 2013). 2012년 우리나라 청소년(13~24세)의 10명중 6명이 전반적인 생활과 학교생활로 인해 스트레스를 받고 있으며, 청소년의 11.2%가 지난 1년동안 한번이라도 자살하고 싶다는 생각을 해 본적이 있는 것으로 나타났다(통계청, 2013). 청소년 자살생각의 주된 이유로는 성적 및 진학문제(28.0%), 경제적어려움(20.5%), 외로움·고독(14.1%), 가정불화(13.6%), 직장문제(6.7%), 기타(17.1%) 순이며, 기타 이유로는 이성문제, 질환, 장애, 친구불화 등인 것으로 나타났다. 이와 같이 자살이 국내뿐만 아니라 많은 선진국들에서도 청소년 사망의 주요 원인으로 나타남에 따라, 전세계적으로 청소년 자살에 대한 다차원적인 예방과 개입의 필요성이 증가하고 있으며 청소년 자살예방에 대해 다각적인 노력을 하고 있다. 그동안 자살에 관한 연구는 자살원인을 밝히기 위하여 정신과적요인, 생물학적·의학적 요인, 사회적·환경적 요인 등에 초점이 맞추어져 왔다(배상민, 우종민, 2011). 자살행동의 원인에 대한 선행연구들은 다양한 변인들이 자살과 관련이 있음을 주장해 왔다. 자살행동의 개인적 변인으로서는 우울(Konick LC et al., 2005), 무망감(Goldston DB et al., 2001), 충동성(Turecki G, 2005), 스트레스 대처능력 부족(Rudd MD, 2001), 낮은 자존감(Wilburn VR, 2008) 등이 보고되고 있다. 자살은 매우 다양하고 복잡한 원인으로 나타나지만, 스트레스와 같은 다양한 외적 자극에 의해서도 영향을 받으며(홍영수 외, 2005), 극심한 스트레스에 노출되는 것은 개인의 능력 혹은 자원을 초과하여 개인의 안녕을 위협할 수 있으며, 자살사고와 자살행동의 위험한 전조라고 보고하고 있다(Izadinia N et al., 2001; 윤영숙, 2012). 자살은 단편적으로 이루어지는 것이 아니라 사회·심리적 자원과 스트레스요인이 상호작용하여 우울증이나 알코올중독 등의 질환이 발생하고(Alloy LB,

1988), 이들 질환이 시간을 두고 점차 진행되면서 자살에 이르는 일련의 과정을 거치게 된다(Tea min Song et al., 2014). 우울증이 있는 사람은 주변의 자살사건과 같은 외적 자극에 노출되었을 때 모방 자살이 증가한다는 연구결과가 있다(Peruzzi N, 1999). 특히, 청소년의 자살원인은 매우 다양하다. 우리나라는 고등학생들의 경우 빈번하게 바뀌는 입시정책의 혼란과 학업에 대한 스트레스가 참기 어려운 긴장을 야기하고, 번어날 수 있는 수단으로써 자살행위를 선택하고 있다(윤영숙, 2012; 서신자, 2013). 자살을 선택하는 청소년의 대부분은 현실에 대한 인지적 숙지보다는 정서적, 감정적 동요에 의해 일시적인 충동감, 분노감, 자기조절능력 저하 등에 의해 일어날 수 있다(김민경, 2012). 스포츠 참여와 같은 사회적 네트워크가 학교생활의 소속감을 증가시켜 청소년의 자살행동을 감소시킬 수 있다(Brown DR et al., 2012). 자아존중감은 청소년 자살의 대표적인 보호요인으로 자아존중감이 낮을수록 자살위험이 높아지는 것으로 보고되고 있다(Heather LR et al., 2010). 청소년 자살에 영향을 미치는 가족요인으로는 부모의 별거와 이혼 및 부모자녀간의 낮은 유대관계가 스트레스로 작용하여 자살의 위험요인이라고 보고하고 있다(Beautrais AL, 2003; Meltzer H 2001). 친구관계에서 발생하는 폭력 피해경험이 우울 등의 부정적 감정을 매개하여 자살생각에 영향을 미친다(Kaufman JM, 2009). 외모 및 용돈 스트레스는 자살충동에 영향을 미치는 것으로 나타났다(김경미 외, 2013). 가족원 중 자살한 사람이 있다는 것은 청소년 자살위험성을 높이는 중요한 요인이 될 수 있다(Bridge JA et al., 1997). 부모의 관심은 자녀들이 자살생각을 하지 않게 하는 긍정적인 효과가 있다(Fergusson DM et al., 1995). 또래 친구와의 상호간계가 부족한 청소년은 우울을 유발하여 자살생각에 이르는 것으로 나타났다(Kendel DB et al., 1991). 자기통제가 부족할수록 자

살의 가능성이 높은 경향이 있다(김현주, 2008). 낮은 가계소득으로 인해 지속적으로 스트레스를 받은 청소년은 자살생각을 많이 하는 경향이 있다(Toprak S et al., 2011; 김선아, 2009). 신문과 방송, 그리고 인터넷에서 구체적으로 자살경위 및 방식을 묘사하는 경우, 이를 그대로 재현하는 모방자살이 뒤따를 수 있다(Kim BS et al., 2011; Stack S et al., 2005; Tousignant M et al., 2005).

한편, 우리나라는 최근 스마트기기의 보급이 확산됨에 따라 모바일 인터넷과 SNS 이용은 급속히 증가하고 있다. 2013년 7월 현재 우리나라의 만 3세 이상 인구의 인터넷 이용률은 82.1%이며 이중 만 6세 이상 인구의 55.1%가 1년 이내 SNS를 이용하고 있는 것으로 나타났다(미래창조과학부, 2013). 사회관계망 서비스인 SNS는 실시간성과 가속성이라는 특성으로 사회전반의 문제에 대한 이슈가 SNS를 통해 확산된다. 특히, SNS는 청소년들이 일상생활 속에서 갖는 우울한 감정이나 스트레스, 고민을 들을 수 있고 행태를 이해할 수 있는 공간으로 SNS 상에서 나타나는 자살에 대한 감정표현이나 심리적 위기 행태들을 분석하게 되면 위험징후와 유의미한 패턴을 감지하여 자살을 예방하는데 긍정적 효과를 발휘할 수 있다(한국정보화진흥원, 2012). 이러한 SNS를 통하여 전송되는 데이터의 양이 기하급수적으로 증가하여 데이터가 경제적 자산으로서의 가치를 인정받기 시작하면서 빅데이터의 활용과 분석이 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하며 다양한 부문에서 빅 데이터의 적극적인 활용을 시도하고 있다. 우리나라는 이미 수많은 빅데이터를 정부 및 공공기관이나 민간기관의 검색포털이나 SNS에서 관리·저장하고 있으나 정보접근과 분석방법의 어려움으로 빅데이터의 활용과 분석은 미흡한 실정에 있다(송태민 외, 2013). 특히, 청소년 자살의 원인과 관련 요인을 구명하기 위하여 기존에 실시하던 횡단적 조사나

종단적 조사 등을 대상으로 한 연구는 정해진 변인들에 대한 개인과 집단의 관계를 보는 데에는 유용하나 사이버 상에서 언급된 개인별 버즈(buzz: 입소문)가 사회적 현상들과 얼마나 어떻게 연관되어 있는지 밝히고 원인을 파악하는 데는 한계가 있다. 이러한 점에서 소셜 빅데이터를 활용한 데이터마이닝의 의사결정나무 분석은 특별한 통계적 가정 없이 결정규칙에 따라 새로운 상관관계나 패턴 등을 발견함으로써 자살과 같은 인간행동의 복잡하고 역동적인 현상에서 발생하는 다양한 원인들의 상호작용 관계를 효과적으로 분석하는데 유용한 도구라고 할 수 있다. 이에 본 연구는 우리나라 온라인 뉴스사이트, 블로그, 카페, SNS, 게시판 등에서 수집된 소셜 빅데이터를 바탕으로 청소년 자살의 위험을 예측할 수 있는 모형과 연관 규칙을 제시하고자 한다. 본 연구의 목적의 소셜 빅데이터를 활용하여 데이터마이닝의 의사결정나무 분석을 통해 한국의 청소년 자살 위험 예측모형을 제시하는 것으로 구체적인 목적은 다음과 같다. 첫째, 청소년 자살 위험에 영향을 미치는 요인을 파악한다. 둘째, 청소년 자살 위험요인을 예측할 수 있는 의사결정나무를 개발한다.

2. 연구 방법

가. 연구대상

본 연구는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 대상으로 하였다. 본 분석에서는 215개의 온라인 뉴스사이트, 4개의 블로그(네이버, 네이버, 다음, 티스토리), 3개의 카페(네이버, 다음, 뽐뿌), 2개의 SNS(트위터, 미투데이), 4개의 게시판(네이버지식인, 네이버지식, 네이버톡, 네이

트판) 등 총 228개의 온라인 채널을 통해 수집 가능한 텍스트 기반의 웹 문서(버즈)를 소셜 빅데이터로 정의하였다. 2011. 1. 1~2013. 3. 31(821일)까지 해당 채널에서 청소년 자살 관련 토픽 221,691건을 수집하여 분석하였다. 본 연구를 위한 소셜 빅 데이터의 수집은 크롤러(Crawler)를 사용하였고 토픽의 분류는 주제분석(text mining) 기법을 사용하였다. 청소년 자살 관련 토픽은 ‘여학생 자살, 남중생 자살, 남고생 자살, 남학생 자살, 여중생 자살, 중학생 자살, 여고생 자살, 고등학생 자살, 고딩 자살, 중딩 자살, 초딩 자살, 초등학생 자살, 중학교 자살, 고등학교 자살, 초등학교 자살’의 15개의 유사어를 사용하였으며, ‘자살골, NLL, 싸이, 김장훈, 문재인, 노무현, 안철수, 공방, 민정수석, 대통령, 김주익, 악어새, 의혹, 박정희, 박근혜, 김대중, 후보, 민심장악, 신경전, 무소속, 대선후보, 여당, 야당, 아랑, 아랑사또전, 강문영, 신민아, 무영, 무연, 이준기, 광해, 이병헌, 베르테르, 영화광해, 알랭드보통, 장편소설, 소설, 영화, 혼자살기, 혼자살지, 이명박, 혼자살게, 돌고래쇼, 혼자살, 동물쇼, 돌고래’의 46개의 불용어를 사용하여 수집하였다. 그리고 청소년 자살 토픽의 수집은 요일별, 주말, 휴일을 고려하지 않고 매 시간단위로 수집하였다. 본 연구의 실제 분석대상은 전체 청소년 자살 버즈(221,691건)에서 청소년 자살의 원인을 언급한 83,657건의 버즈를 대상으로 하였다.

나. 연구도구

청소년 자살과 관련하여 수집된 버즈는 주제분석(text mining)과 감성분석(opinion mining)의 과정을 거쳐 다음과 같이 정형화 데이터로 코드화 하여 사용하였다.

1) 자살 원인

청소년 자살 원인의 분류는 ‘사망, 수능, 우울증, 성폭행, 고통, 충격, 성적, 걱정, 스트레스, 해고, 왕따, 폭력, 생활고, 이혼, 열등감, 얼굴, 질병, 학교폭력’ 18개로 대상이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

2) 자살 방법

자살 방법의 분류는 ‘투신자살, 분신자살, 동반자살, 전깃줄, 번개탄, 음독자살, 자해, 할복자살, 수면제, 연탄, 밧줄, 넥타이, 농약, 가스, 고카페인, 철사, 본드, 질식’의 18개로 방법이 있는 경우는 ‘1’, 없는 경우는 ‘0’으로 코드화 하였다.

3) 자살 감성

자살 감성은 감성분석(opinion mining)을 통하여 자살을 긍정적으로 인식하는 표현(예: ‘자살 할거다’, ‘자살 선택하다’, ‘자살 좋다’, ‘자살 쉽다’ 등)이 담긴 버즈, 자살을 부정적으로 인식하는 표현(예: ‘자살 안타깝다’, ‘자살 나쁘다’, ‘자살 심각하다’, ‘자살 충격적이다’ 등)이 담긴 버즈, 자살을 보통으로 인식하는 표현(긍정적 표현과 부정적 표현이 혼합된 문서)이 담긴 버즈를 코드화(1: 긍정, 2: 보통, 3: 부정) 하였다. 본 연구의 최종 자살감성에 사용된 변인은 자살위험(1: 자살을 긍정과 보통으로 인식하는 문서)과 자살보호(0: 자살을 부정적으로 인식하는 문서)로 코드화 하였다.

다. 분석 방법

본 연구에서는 한국의 청소년 자살 위험요인을 설명하는 가장 효율적인 예측모형을 구축하기 위해 특별한 통계적 가정이 필요하지 않은 데이터마이닝의 의사결정나무 분석방법을 사용하였다. 데이터마이닝의 의사결정나무 분석은 방대한 자료 속에서 종속변인을 가장 잘 설명하는 예측모형을 자동적으로 산출해 줌으로써 각기 다른 원인을 가진 청소년 자살에 대한 위험요인을 쉽게 파악할 수 있다. 본 연구의 의사결정나무 형성을 위한 분석 알고리즘은 CHAID(Chi-squared Automatic Interaction Detection), Exhaustive CHAID, CRT(Classification and Regression Tree), QUEST(Quick, Unbiased, Efficient Statistical Tree) 확장방법(growing method) 중 모형의 예측율이 가장 높은 Exhaustive CHAID를 사용하였다. Exhaustive CHAID는 이산형인 종속변수의 분리기준으로 카이제곱(χ^2 -검정)을 사용하며, 모든 가능한 조합을 탐색하여 최적분리를 찾는다. 정지규칙(stopping rule)로 상위노드(부모마디)의 최소케이스 수는 100로 하위노드(자식마디)의 최소 케이스수는 50으로 설정하였고, 나무깊이는 3수준으로 정하였다. 그리고 데이터 분할에 의한 타당성평가를 위해 훈련표본(training data)과 검정표본(test data)의 비율은 70:30으로 설정하였다. 본 연구의 기술분석, 다중로지스틱 회귀분석, 의사결정나무분석은 SPSS 20.0을 사용하였다.

라. 연구의 윤리적 고려

연구의 대한 윤리적 고려를 위하여 한국보건사회연구원 생명윤리위원회(IRB)의 승인(No. 2014-1)을 얻은 후 연구를 진행하였다. 연구대상 자료는 한국보건사회연구원과 SKT가 2013년 5월에 수집한 2차 자료를 활

용하였으며, 수집된 소셜 빅데이터는 개인정보를 인식할 수 없는 데이터로 대상자의 익명성과 기밀성이 보장되는 연구이다.

3. 연구 결과

가. 주요 변인들의 기술통계

자살 토픽을 언급한 전체 버즈는 221,691건이며, 이중 자살의 원인을 언급한 버즈는 37.7%(83,657건)로 나타났다. 청소년 자살검색의 원인은 학교폭력, 우울, 성적, 외모, 사망, 열등감, 충격, 생활고 순으로 나타났다. 청소년 자살 방법은 투신, 목을 매거나 압박을 가하여 질식(넥타이, 철사), 살충제에 의한 자의중독 및 노출(농약), 나머지 수단(연탄, 본드, 분신, 카페인) 순으로 나타났다(표 6-19 참조).

〈표 6-19〉 Descriptive Statistics of Factors

Cause of suicide				Method of suicide			
Cause	Buzz(%)	Cause*	Buzz(%)	Method	Buzz(%)	Method†	Buzz(%)
Dying	10309 (6.4)	School Violence	28665 (25.3)	Jumping	3404(38.2)	Bond	1017 (12.0)
SAT	2768 (1.7)			Posse	172(1.9)		
Melancholia	8518 (5.3)	Dying	13809 (12.2)	Double suicide	810(9.1)	Coal briquette	1139 (13.5)
Sexual Assault	2977 (1.8)			Cords	199(2.2)		
Pain	9195 (5.7)	Inferiority	9726 (8.6)	Bongetan	56(0.6)	Pesticide	1031 (12.2)
Impact	9057 (5.6)			Overdose	121(1.4)		
Test scores	16234 (10.0)	Melancholy	17921 (15.8)	Self-harm	1130(12.7)	Wire	293 (3.5)

Cause of suicide				Method of suicide			
Cause	Buzz(%)	Cause*	Buzz(%)	Method	Buzz(%)	Method†	Buzz(%)
Worry	8548 (5.3)			Seppuku	62 (0.7)		
Stress	10833 (6.7)	Appearance	15213 (13.4)	Sleeping pills	668 (7.5)	Jumping	3577 (42.3)
Firing	643 (0.4)			Coal briquette	345 (3.9)		
Bullying	15282 (9.4)			Tether	126 (1.4)		
Violence	28665 (17.7)	Economy	1275 (1.1)	Tie	111 (1.2)	Posse	172 (2.0)
Economy	742 (0.5)			Pesticide	300 (3.4)		
Divorce	6249 (3.9)	SAT	17791 (15.7)	Gas	841 (9.4)	Tie	1227 (14.5)
Inferiority	1234 (0.8)			Caffeine	7 (0.1)		
Appearance	8393 (5.2)	Impact	9057 (8.0)	Wire	45 (0.5)	Caffeine	7 (0.1)
Disease	2142 (1.3)			Bond	247 (2.8)		
School Violence	20127 (12.4)			Suffocation	256 (2.9)		

주: * Cause analysis (unique value of 1 or greater) was conducted on 18 causes and cause factors were reduced to 8.

† Cause analysis (unique value of 1 or greater) was conducted on 18 methods and method factors were reduced to 8.

나. 청소년 자살 위험에 미치는 영향 요인

사망요인을 제외하고 모든 요인이 자살보호 보다 자살위험에 더 많이 영향을 주는 것으로 나타났다. 자살위험에 영향을 미치는 요인으로는 외모요인, 열등감요인, 우울요인, 충격요인, 학교폭력요인, 생활고요인, 성적요인의 순으로 나타났다(표 6-20 참조).

〈표 6-20〉 Multinomial Logistic of Suicide Causes

Variable	B	S.E.	P	OR(95%CI) [†]
Intercept	-1.219	0.014	.001	
School Violence	0.179	0.017	.001	1.196(1.16~1.24)
Dying	-0.240	0.022	.001	0.786(0.75~0.82)
Inferiority	0.272	0.023	.001	1.313(1.25~1.37)
Melancholy	0.255	0.019	.001	1.290(1.24~1.37)
Appearance	0.514	0.019	.001	1.673(1.61~1.74)
Economy	0.159	0.062	.01	1.172(1.04~1.32)
SAT	0.128	0.019	.001	1.136(1.09~1.18)
Impact	0.204	0.024	.001	1.227(1.17~1.29)

주: [†] Adjusted odds ratio(95% Confidence interval)

다. 청소년 자살위험 요인 예측모형

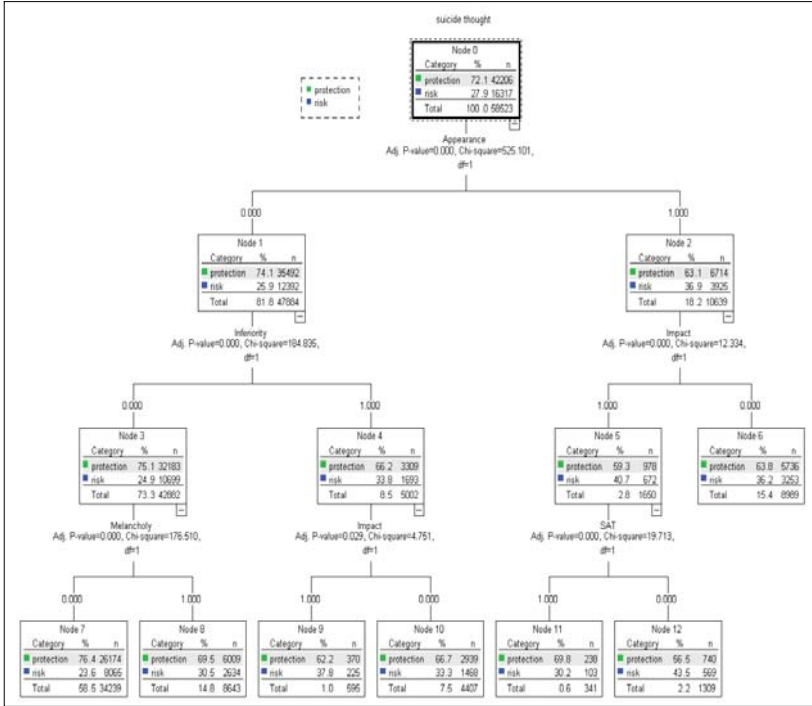
노드분리 기준을 이용하여 나무형 분류모형에 따른 모형의 예측율(정분류율)을 검증하여 예측력이 가장 높은 모형을 선택하였다. 트리의 분리 정확도를 나타내는 정분류율을 비교 분석한 결과 Exhaustive CHAID와 CRT 알고리즘의 훈련표본과 검정표본의 정분류율이 721.%과 72.0%로 가장 높았으나, CRT 알고리즘은 2차 데이터마이닝에서 훈련표본의 정확도가 Exhaustive CHAID 보다 낮아, 훈련표본과 검정표본의 정분류율이 높게 나타난 Exhaustive CHAID 알고리즘을 선택하였다(표 6-21 참조).

〈표 6-21〉 Predictive Performance according to Modeling Methods

Modeling methods	Training data		Test data	
	Correct(%)	Wrong(%)	Correct(%)	Wrong(%)
CHAID	72.1	27.9	71.9	28.1
Exhaustive CHAID	72.1	27.9	72.0	28.0
CRT	72.1	27.9	72.0	28.0
QUEST	72.1	27.9	71.7	28.3

한국의 청소년 자살위험 예측모형에 대한 의사결정나무 분석결과 는 [그림 6-22]와 같다. 훈련표본의 나무구조의 최상위에 있는 네모는 루트노드로서, 예측변수(독립변수)가 투입되지 않은 종속변 수(청소년 자살위험)의 빈도를 나타낸다. 루트노드에서 청소년의 자 살위험은 27.9%(16,317건), 자살보호는 72.1%(42,206건)으로 나 타났다. 루트노드의 하단의 가장 상위에 위치하는 요인이 청소년 자 살위험 예측에 가장 영향력이 높은(관련성이 깊은)요인으로 ‘외모 요인’의 영향력이 가장 큰 것으로 나타났다. 즉, ‘외모요인’의 위험이 높은 경우 청소년 자살위험이 이전의 27.9%에서 36.9%로 증가한 반면, 청소년 자살보호는 이전의 72.1%에서 63.1%로 낮아졌다. ‘외 모요인’이 높더라도 ‘충격요인’이 높으면 청소년 자살위험이 이전의 36.9%에서 40.7%로 증가한 반면, 청소년 자살보호는 이전의 63.1%에서 59.3%로 낮아졌다. ‘충격요인’이 높더라도 ‘성적요인’이 높으면 청소년 자살위험이 이전의 40.7%에서 30.2%로 감소한 반 면, 청소년 자살보호는 이전의 59.3%에서 69.8%로 높아졌다. 루트 노드 하단의 ‘외모요인’의 위험이 낮은 경우 청소년 자살위험이 이전 의 27.9%에서 25.9%로 낮아진 반면, 청소년 자살보호는 이전의 72.1%에서 74.1%로 높아졌다. ‘외모요인’이 낮더라도 ‘열등감요인’ 이 높으면 청소년 자살위험이 이전의 25.9%에서 33.8%로 증가한 반면, 청소년 자살보호는 이전의 74.1%에서 66.2%로 낮아졌다. ‘열 등감요인’이 높더라도 ‘충격요인’이 높으면 청소년 자살위험이 이전 의 33.8%에서 37.8%로 높아진 반면, 청소년 자살보호는 이전의 66.2%에서 62.2%로 낮아졌다.

[그림 6-22] Predictive model of risk factors for Suicide



라. 청소년 자살 위험요인 예측모형에 대한 이익도표

본 연구에서 청소년 자살위험이 가장 높은 경우는 ‘외모요인’의 위험이 높으면서 ‘충격요인’의 위험이 높고 ‘성적요인’의 위험이 낮은 조합으로 나타났다. 즉, 12번 노드의 지수(index)가 155.9%로 뿌리마디와 비교했을 때 12번 노드의 조건을 가진 집단의 청소년 자살위험이 약 1.56배로 나타났다. 청소년 자살보호가 가장 높은 경우는 ‘외모요인’의 위험이 낮으면서 ‘열등감요인’의 위험이 낮고 ‘우울감요인’의 위험이 낮은 조합으로 나타났다. 즉, 7번 노드의 지수(index)가 106.0%로 뿌리마디와 비교

했을 때 7번 노드의 조건을 가진 집단의 청소년 자살보호가 1.06배로 나타났다(표 6-22 참조). 본 연구의 데이터 분할에 의한 타당성 평가를 위해 훈련표본과 검정표본을 비교한 결과 훈련표본의 위험 추정치(risk estimate)는 0.279(standard error: 0.002), 검정표본의 위험추정치는 0.280(standard error: 0.003)으로 본 청소년 자살 위험요인 예측모델의 일반화에 무리가 없는 것으로 나타났다.

〈표 6-22〉 Profit chart of Predictive models of Suicide

Type	Node	Profit index				Cumulative index			
		node: n	node: %	gain(%)	index(%)	node: n	node: %	gain(%)	index(%)
Risk	12	1309	2.2	3.5	155.9	1309	2.2	3.5	155.9
	9	595	1.0	1.4	135.6	1904	3.3	4.9	149.6
	6	8989	15.4	19.9	129.8	10893	18.6	24.8	133.3
	10	4407	7.5	9.0	119.5	15300	26.1	33.8	129.3
	8	8643	14.8	16.1	109.3	23943	40.9	49.9	122.1
	11	341	0.6	0.6	108.3	24284	41.5	50.6	121.9
	7	34239	58.5	49.4	84.5	58523	100.0	100.0	100.0
Protection	7	34239	58.5	62.0	106.0	34239	58.5	62.0	106.0
	11	341	0.6	0.6	96.8	34580	59.1	62.6	105.9
	8	8643	14.8	14.2	96.4	43223	73.9	76.8	104.0
	10	4407	7.5	7.0	92.5	47630	81.4	83.8	102.9
	6	8989	15.4	13.6	88.5	56619	96.7	97.4	100.6
	9	595	1.0	0.9	86.2	57214	97.8	98.2	100.5
	12	1309	2.2	1.8	78.4	58523	100.0	100.0	100.0

4. 논의

본 연구는 소셜 빅데이터를 활용하여 데이터마이닝의 의사결정나무 분석을 통한 한국의 청소년 자살위험요인에 대한 예측모형을 개발하였다. 본 연구에서 사용된 의사결정나무 분석모형은 특별한 통계적 가정 없이 결정규칙에 따라 나무 구조로 도표화하여 분류와 예측을 수행하는 방법으로 소셜 빅데이터에서 여러 개의 독립변수 중 종속변수에 대한 영향력이 높은 변수의 패턴이나 관계를 찾아내는데 유용하다. 본 연구의 결과를 중심으로 논의하면 다음과 같다. 첫째, SNS상의 청소년 자살 원인은 학교폭력, 우울, 성적, 외모, 사망, 열등감, 충격, 생활고 등으로 나타났다. 이는 2012년 청소년 통계에서 조사된 자살 원인인 성적, 경제적 어려움, 외로움, 가정불화, 직장문제, 기타(이성, 질환, 친구불화 등)의 내용과 다르게 나타났다. 따라서 SNS상에 나타나는 자살에 대한 감정이 실제 오프라인 조사와 다르다는 것을 감안할 때 청소년 자살생각의 조사항목에 대한 재분류와 함께 세분화가 필요할 것으로 본다. 둘째, SNS상의 청소년 자살 방법은 투신(42.3%), 질식(18.0%), 살충제(12.2%), 본드(12.0%) 기타의 순으로 나타났다. 이는 2012년 통계청 사망원인통계에서 19세 이하 청소년의 실제 자살방법인 투신(57.1%), 질식(27.5%), 본드(7.4%), 살충제(2.1%), 기타 순과 비슷하게 나타났다. SNS상에서 검색하는 청소년의 자살방법이 실제 청소년 자살 사망자의 자살방법과 비슷한 것은 언론에서 구체적으로 온라인상의 자살 경위 및 자살방법을 묘사하는 경우 모방자살을 유발할 수 있다는 연구(Kim BS et al., 2011; Stack S., 2005; Tousignant M et al., 2013)를 지지하는 것으로 자살 검색을 언론의 자세한 자살보도는 자율적으로 규제되어야 할 것으로 본다. 셋째, 청소년 자살위험 요인은 외모요인, 열등감요인, 우울요인, 충격요인, 학

교폭력요인, 생활고요인, 성적요인의 순으로 나타났다. 이는 외모(김경미 외, 2013), 자아존중감(Heather LR et al., 2010), 우울(Peruzzi N et al., 1999), 충동(김민경, 2012), 학교폭력(Kaufman JM., 2009), 생활고(Toprak Set al., 2011; 김선아, 2009), 학업스트레스(윤영숙, 2012; 서신자, 2013)가 자살충동에 영향을 미친다는 연구들을 지지하는 것으로 나타났다. 넷째, 청소년의 자살위험 요인을 예측하는 가장 중요한 변인은 외모요인으로 나타났다. 외모요인의 위험이 높을 경우 자살위험은 증가하였고 자살보호는 감소한 것으로 나타나 외모 스트레스는 자살 충동에 영향을 미친다는 기존의 연구(김경미, 2013)를 지지하는 것으로 나타났다. 외모요인의 위험이 높고 충동요인의 위험이 높을 경우 자살위험은 증가하였고 자살보호는 감소한 것으로 나타나 자살을 선택하는 청소년의 대부분은 현실에 대한 인지적 숙지보다는 충동감, 분노감, 자기조절능력 저하 등에 영향을 받는다는 기존의 연구(김민경, 2012)를 지지하는 것으로 나타났다. 외모요인의 위험이 높고 충동요인의 위험이 높고 수능요인의 위험이 높을 경우 자살위험은 감소하였고 자살보호는 증가한 것으로 나타나 우리나라 청소년들이 빈번하게 바뀌는 입시정책의 혼란과 학업에 대한 스트레스가 자살에 영향을 준다는 기존의 연구(윤영숙, 2012; 서신자 외, 2013)를 지지하는 것으로 나타났다. 다섯째, 외모요인의 위험이 낮더라도 열등감 위험이 높으면 자살위험은 증가하였고 자살보호는 감소한 것으로 나타나 자아존중감이 청소년의 자살의 대표적인 보호요인이라는 기존의 연구(Heather LR et al., 2010)를 지지하는 것으로 나타났다. 외모요인과 열등감 위험이 낮더라도 우울감 위험이 높으면 자살위험은 증가하였고 자살보호는 감소한 것으로 나타나 우울증이 있는 사람이 주변의 자살사건에 노출이 되었을 경우 모방자살이 증가한다는 연구결과(Peruzzi N, 1999)를 지지하는 것으로 나타났다. 마지막으로, 청소년 자

살위험이 가장 높은 경우는 외모요인의 위험이 높으면서 충격요인의 위험이 높고, 성적요인의 위험이 낮은 조합으로 이 집단의 자살위험은 뿌리마디와 비교했을 경우 약 1.56배 증가하였다. 본 연구를 근거로 우리나라의 자살예방과 관련하여 다음과 같은 정책적 함의를 도출할 수 있다. 첫째, 청소년 자살위험 요인은 외모요인, 충격요인, 성적요인, 우울요인, 열등감 요인 등이 복합적으로 영향을 미치는 것으로 나타나, 정부의 학생 자살예방 및 위기관리 프로그램에 이러한 복합적인 요인을 예방하고 치료할 수 있는 방안이 마련되어야 할 것이다. 둘째, 청소년은 온라인상에서 자살과 관련한 담론을 주고 받으며 이러한 언급이 실제적인 자살과 관련된 심리적·행동적 특성으로 노출이 될 수 있기 때문에 본 연구의 자살 예측모형에 따른 위험 징후가 예측되면 실시간으로 개입할 수 있는 애플리케이션의 개발이 필요할 것이다(송태민, 2013). 셋째, SNS에서 주고받는 자살관련 소셜 담론(Buzz)은 개인의 일상생활에서 갖는 우울한 감정이나 고민이 기록되는 온라인 심리적 부검보고서라고 할 수 있다. 따라서 핀란드가 국가적 차원의 오프라인 보고서를 바탕으로 자살을 줄였다면 우리나라는 소셜 빅데이터 분석으로 국가차원의 자살예방 대책이 마련되어야 할 것이다(송태민, 2014). 본 연구는 개개인의 특성을 가지고 분석한 것이 아니고 그 구성원이 속한 전체 집단의 자료를 대상으로 분석하였기 때문에 이를 개인에게 적용하였을 경우 생태학적 오류(ecological fallacy)가 발생할 수 있다. 또한, 본 연구에서 정의된 자살 원인 관련요인(용어)은 버즈 내에서 발생된 단어의 빈도로 정의되었기 때문에 기존의 조사 등을 통한 이론적 모형에서의 자살 원인과 의미가 다를 수 있어 후속 연구에서의 검증이 필요할 것으로 본다. 그러나 이러한 제한점에도 불구하고 본 연구는 소셜 빅데이터에서 수집된 자살 버즈를 주제분석과 감성분석을 통하여 자살 원인과 방법 등을 분류하였고, 데이터

마이닝 분석을 통하여 청소년 자살위험 요인의 예측모형을 제시한 점에서 분석방법론적으로 의의가 있다. 또한, 실제적인 내용을 빠르게 효과적으로 파악하여 사회통계가 지닌 한계를 보완할 수 있는 새로운 조사방법으로서의 빅데이터의 가치를 확인하였다는 점에서 조사방법론적 의의를 가진다고 할 수 있다.



제7장

보건복지 빅데이터 활성화 전략 및 발전방안

제1절 보건복지 빅데이터 활성화를 위한 전략분석

제2절 보건복지 빅데이터 발전방안

7

보건복지 빅데이터 활성화 << 전략 및 발전방안

제1절 보건복지 빅데이터 활성화를 위한 전략분석

보건복지 빅데이터 활성화를 위한 정책을 도출하기 위하여 2장 빅데이터 이론적 배경, 3장 국내외 공공빅데이터 공개현황, 4장 보건복지 이슈 분석 및 통계제공 현황 등의 분석결과를 살펴보면 다음과 같다. 첫째, 보건복지 빅데이터를 통합적으로 관리하기 위한 범부처 차원의 조직이 필요할 것이다. 둘째, 보건복지 빅데이터의 선도기술(저장기술, 처리기술 등)의 활용이 우선되어야 할 것이다. 셋째, 빅데이터의 수용과정은 기술 수용 목적에 따라 다르기 때문에 수요자와 공급자 중심의 빅데이터 활용 방안의 모색이 필요할 것이다. 넷째, 소셜 미디어에서 공개된 개인정보는 위변조 오남용이 쉽고 상업적 목적의 정보수집 등에 노출될 수 있기 때문에 보건복지 빅데이터 프라이버시 보호방안의 제도적 마련이 필요할 것이다. 다섯째, 키워드 분석을 통한 보건복지 이슈와 분류 체계의 구축이 필요할 것으로 본다. 여섯째, 소셜 미디어에서 정보를 추출하고, 정부와 공공기관이 가진 정형데이터와의 연계를 위한 소셜 빅데이터 분석방안이 필요하다. 마지막으로 오프라인 보건복지 수요조사와 함께 소셜 빅데이터를 활용한 보건복지 수요예측을 실시하기 위한 방안 마련이 필요하다.

한편, 본 연구의 목적은 국내외의 보건복지 관련 빅데이터 현황과 기술을 분석하고 수요공급의 예측과 분석사례를 통하여 수요자 중심의 맞춤형 서비스를 위한 보건복지 빅데이터의 효율적 관리방안을 제시하는 것 위함이며, 이러한 목표를 달성하기 위하여 <표 7-1>과 같이 SWOT분

석을 실시하였다. SWOT분석은 빅데이터 산업에 내재하는 주요 요인들과 외부 환경요인들을 분석하여 이를 정의하고 외부의 기회(Opportunity)와 위협(Threat), 내부의 강점(Strength)과 약점(Weakness)을 연계하여 사업의 전략적 대안을 도출한다. 그리고 내부의 강점 및 약점은 현재나 미래의 전략적 의사결정을 수행하는 데 있어 영향을 주며 약간의 통제 및 조정이 가능하나 외부의 기회와 위협은 흔히 발생하고 대상조직이나 업무에 중대한 영향을 주며 통제 또는 조정이 어렵다. 이 결과로 기회를 이용하여 강점으로 활용하는 전략과 위협을 피하고 약점을 최소화하는 각각의 전략을 수립할 수 있다.

〈표 7-1〉 보건복지 빅데이터 서비스 활성화를 위한 SWOT 분석

외부요인	내 부 요 인	강점(Strength)	약점(Weakness)
		• 보건복지 공공 빅데이터(건보, 연금, 심평원 등) 보유 • 보건복지 공공 정보시스템(사통망, 지역보건소정보시스템 등) 운영 • 스마트미디어 보급확산에 따른 소셜 빅데이터 양적 증가 • 정보통신(ICT)과 보건복지 기술의 융합 기술능력 보유 • 공공기관의 보건복지 빅데이터 핵심과제 추진	• 보건복지 빅데이터 상호운용성 확보를 위한 표준화 미비 • 부처별 빅데이터 사업에 대한 통합연계 체계 부재 • 공공과 민간 빅데이터 공동 활용을 위한 협력체계 부재 • 빅데이터 분석력 미흡 및 전문 인력 부족 • 보건복지 빅데이터 프라이버시 보호를 위한 법/제도 미비
기회(Opportunity)		강점·기회전략	약점·기회전략
• 정부3.0의 추진동력으로 빅데이터 선정 • 사회 전반적인 불확실성과 위험 관리 대응방안으로 빅데이터 활용 • 보건복지 욕구 증가로 빅데이터 기반 개인맞춤형 서비스 수요 증가 • 보건의로 패러다임이 치료중심에서 예방관리 중심으로 전환 • 빅데이터 융합기술에 대한 필요성 및 수요 증대 • 수요자와 공급자의 빅데이터에 대한 관심 증가		• 정부 및 공공기관 보유 보건복지 빅데이터의 최우선 활용 • 공공 정보시스템의 연계를 통한 보건복지 빅데이터 서비스 체계 구축 • 소셜 빅데이터 분석 기반 보건복지 정책 개발 • 정형 빅데이터와 비정형 빅데이터의 연계 체계 구축 • 빅데이터 분석 기술과 ICT를 결합한 보건복지 맞춤형 서비스 체계 구축 • 보건복지부 3.0에 보건복지 빅데이터 분석을 위한 플랫폼 조기 구축 및 활용	• 보건복지 빅데이터 분석 선도 기술 개발 유도 및 지원 • 보건복지 빅데이터 분석 기술 표준화 체계 구축 • 범정부 보건복지 빅데이터 거버넌스 구축 • 민간기관과 협조체계를 위한 보건복지 빅데이터 포럼 조직 및 운영 • 보건복지 빅데이터 데이터 사이언티스트 인재 양성 • 보건복지 빅데이터 프라이버시 보호를 위한 법/제도 정비 • 보건복지 빅데이터 연계체계 구축
위협(Threat)		강점·위협전략	약점·위협전략
• 해외 선진기업의 국내 보건복지 빅데이터 시장 및 기술 선점 • 해외 주요 선진국의 보건복지 빅데이터 R&D 강력 추진 • 개인건강정보 유출로 인한 사생활침해 확대 • 빅데이터 비용·효과에 대한 분석 미비 • 공공과 민간의 니즈에 부합되는 보건복지 R&D 사업 개발 요구 증대 • 부처별/자체별 보건복지 빅데이터 사업 별도 추진		• 글로벌 보건복지 빅데이터 활용 기술 지원 • 보건복지 빅데이터 확산을 위한 비즈니스 모델 개발 • 보건복지 빅데이터 원천기술 및 플랫폼 구축을 위한 공공/민간 R&D 투자 확대 • 산·학·연 협력을 통한 보건복지 빅데이터의 기술 개발 활성화 • 정부 차원의 보건복지 빅데이터 컨트롤타워 조직 및 운영	• 글로벌 보건복지 빅데이터 기술 및 표준화 체계 마련 • 보건복지 빅데이터 서비스 성공 사례 발굴 및 확산 • 산·학·연 협력을 통한 보건복지 데이터 사이언티스트 인력 양성 • 보건복지 빅데이터 비용효과 분석연구 • 보건복지 빅데이터 활성화를 위한 중장기 계획 수립 • 보건복지 빅데이터 프라이버시 보호를 위한 관련 법 지속 추진

자료: '송태민 외(2011). u-Health 현황과 정책과제. p129'와 'ETRI(2014. 2). 보건의로 Big Data R&D 사업기획. p45'에서 수정 보완

제2절 보건복지 빅데이터 발전방안

1. 보건복지 빅데이터의 효율적 활용 방안²⁵⁾

보건복지 분야 빅데이터를 효율적으로 활용하기 위해서는 다음과 같은 전략이 필요할 것이다.

첫째, 보건복지 빅데이터를 통합적으로 관리하기 위한 범부처 차원의 (가칭)보건복지 빅데이터 관리 위원회의 운영이 필요하다. 현재 보건복지 빅데이터는 보건복지부, 고용노동부, 지식경제부(현 산업통상자원부, 미래창조과학부), 식품의약품안전처, 통계청 등 많은 정부부처와 국민건강보험공단, 건강보험심사평가원, 국책연구기관 등 많은 공공기관에서 관리·운영되고 있어 각 기관에서 운영 중인 정보의 연계와 공유를 위해서는 범정부 차원의 조직이 필요할 것이다. 둘째, 비정형화된 보건복지 빅데이터를 관리하고 있는 민간 기관과의 협조체제가 마련되어야 할 것이다. 비정형화된 보건복지 빅데이터는 민간 기관의 검색포털이나 SNS를 통해서 생산·저장 되고 있어 민간기관과의 긴밀한 협조체제(가칭: 보건복지 빅데이터 포럼)가 구축되어야 할 것이다. 셋째, 국가 차원의 오픈 API(Open Application Programming Interface)의 제공이 필요하다. 보건복지의 대부분의 빅데이터는 공공부문에서 독점하고 있다. 정보를 수집/분석하고 수집과 동시에 정보를 실시간으로 웹상에 공개하는 것도 중요하지만 보건복지 빅데이터를 효과적이고 효율적으로 활용하기 위해서는 정부 차원의 API 공개를 적극적으로 검토할 필요가 있다. 우리나라의 공공데이터 제공은 공공데이터 포털에서 제공하고 있으며 2014년 6월 30일 현

25) 본 절은 '송태민 (2013. 9). 우리나라 보건복지 빅데이터 동향 및 활용방안, 과학기술정책, 192, 과학기술정책연구원.'의 내용을 수정·보완함.

재 698개의 기관에서 9,378건의 데이터 셋을 제공하고 있다. 공공데이터를 제공하고 있는 서비스 유형으로 대부분 공공데이터 포털에서 다운로드나 링크서비스로 제공되고 있으며, 제공중인 빅데이터 유형도 문서파일이나 통계표 위주의 2차 자료 제공으로 정부가 요구하는 민간의 창의성 및 혁신적 아이디어를 결합하기 위한 공공데이터로는 미흡한 실정에 있다. 즉, 익명화된 원시정보(1차 자료)가 제공되어야 민간이 가진 소셜 빅데이터와의 연계가 가능하나, 2차 자료로는 새로운 비즈니스를 창출하기 어려울 것이다. 공공데이터 포털에서 제공하고 있는 오픈API는 총 1,471건으로 이중 보건복지분야 공개 API는 100건이다. 대부분의 보건의로 오픈API는 지자체의 병원/의원/약국 조회서비스를 제공하고 있고, 식품의약품안전 오픈 API는 모범음식점, 화학물질, 식품정보 등을 포털을 통해 서비스 하고 있다. 복지 분야는 복지시설 관련 정보(주소, 서비스 내용) 등을 제공해 주고 있다. 공공데이터 활용 사례는 311건으로 이중 보건복지 분야 활용사례는 50건이 제공되고 있다. 보건 분야는 의료기관의 위치정보와 진료과목, 예방접종 등에 대한 어플리케이션을 제공하고 있으며, 복지 분야는 사회복지시설 정보, 봉사활동 종합정보, 노인일자리 정보 등을 제공하고 있다. 현재 제공 중인 공공데이터는 공공기관의 포털에서 제공 중인 서비스가 대부분을 차지하며, 행정자료로 자동 취합되는 원시자료의 실시간 제공은 매우 미흡한 실정이다. 따라서 보건복지 빅데이터를 활용하기 위해서는 현재 공공기관의 포털에서 제공되는 정보의 원시정보를 실시간으로 제공할 수 있는 방안이 마련되어야 할 것으로 본다. 끝으로 정부와 공공기관이 보유·관리하고 있는 빅데이터는 통합하는 방안보다는 각각의 빅데이터의 집단별 특성을 분석하여 위험(또는 수요)집단 간 연계를 통한 예측(위험예측 또는 질병예측 등) 서비스가 제공되어야 할 것이다. 즉, 빅데이터 분석을 통한 개인별 맞춤형 서비스는 프

라이버시를 침해할 수 있기 때문에 위험(또는 수요) 집단별 맞춤형 서비스가 제공되어야 할 것으로 본다. 넷째, 보건복지 빅데이터를 분석 처리할 수 있는 관련 기술의 개발이 필요하다. 스마트 시대에는 비관계형, 비정형 데이터의 저장과 분석, 클라우드 서비스의 확산, 시멘틱 검색 서비스, 추론에 기반한 상황인식 서비스 등의 기술이 핵심이 될 것이다. 따라서 관련부처와 협력하여 보건복지 분야 빅데이터를 ‘수집→저장→분석→추론’ 할 수 있는 기술개발은 물론 기술 표준화가 우선적으로 추진되어야 할 것이다. 다섯째, 구조화되지 않은 대규모 데이터 속에서 숨겨진 정보를 찾아내는 데이터 사이언티스트(Data Scientist)의 인재 양성이 필요할 것이다. 빅데이터 시대에는 데이터를 관리하고 분석할 수 있는 인력이 매우 중요하다. 이미 글로벌 IT업체에서는 데이터 사이언티스트에 대한 인재 확보와 역량 강화에 많은 노력을 경주하고 있다. 따라서 교육부와 협력하여 보건복지 분야 데이터 사이언티스트를 양성할 수 있는 전략이 마련되어야 할 것이다.

끝으로 보건복지 빅데이터의 개인정보와 기밀정보에 대한 보안정책이 마련되어야 할 것이다. 보건복지 빅데이터는 개인에 대한 거의 모든 정보가 저장되어 있지만 아직 법·제도는 미비한 상황이며 논의조차 되지 못하고 있다. 빅데이터의 활용도 중요하지만 과도한 개인정보의 유출은 프라이버시 침해는 물론 사이버 인권침해나 범죄에 악용될 수 있다. 빅데이터로부터 개인을 보호하기 위해 가장 중요한 것은 특정 개인을 식별하지 못하도록 하는 익명화와 정보접근 및 정보처리에 대한 통제이다. 그러나 정보접근 및 정보처리에 대한 통제를 강하게 하면 정보 활용이 활성화되지 않기 때문에 보건복지 빅데이터 ‘활용과 보호의 균형’에 대한 효과적인 정책이 마련되어야 할 것이다(송태민, 2013).

2. 보건의료 및 공공분야 빅데이터 선도기술 활용방안

앞서 공공과 보건의료 빅데이터 수요공급 예측 연구를 위하여 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 데이터마이닝의 의사결정나무 분석 기법을 적용하여 분석하였다. 본 연구의 결과를 요약하면 다음과 같다.

첫째, 보건의료와 공공 분야 모두 빅데이터 기술은 저장기술(데이터베이스, 데이터웨어하우스, 클라우드컴퓨팅, 디스크, NoSQL, San), 분석 기술('NDAP, 기계학습, 네트워크, 마이닝, 시맨틱, 온톨로지, 패턴인식, EDW, 텍스트마이닝, 음성인식, 영상인식, AR, AI, Mahout, ETL, 알고리즘, Pregel), 처리기술('가상화, 맵리듀스, 스케일아웃, 어플라이언스, 하둡, Hbase, MR, Appliance, Dremel, Percolator, 검색, 인덱싱, 스톱, HDFS, 인메모리), 활용기술(BI, 인포그래픽스, SaaS, Paas, BaaS, Daas) 등의 순으로 나타났다. 둘째, 보건의료 분야의 빅데이터 분석기술은 고객관리, 생산관리, 통계, 인사관리, 컨설팅 등의 순으로 활용되고 있으며, 공공 분야의 빅데이터 분석기술은 고객관리, 생산관리, 컨설팅, 의사결정 등에 활용되고 있는 것으로 나타났다. 셋째, 보건의료분야는 혁신요인과 우려요인에 정적(+) 영향을 미치는 것으로 나타나, 혁신과 우려로 인해 공급자가 빅데이터를 수용할 확률이 높은 것으로 나타났다. 공공분야는 혁신요인과 성장요인에 정적(+) 영향을 미치는 것으로 나타나, 혁신과 성장으로 인해 수요자 보다 공급자가 빅데이터를 수용할 확률이 높은 것으로 나타났다. 넷째, 보건의료분야의 감성요인이 빅데이터 수요공급 예측모형에 미치는 영향은 '혁신요인'이 가장 큰 것으로 나타났으며, 보건의료분야의 기반요인이 빅데이터 수요공급 예측모형에 미치는 영향은 '품질요인'이 가장 큰 것으로 나타났다. 다섯째, 공공분야의 감성요인이

빅데이터 수요공급 예측모형에 미치는 영향은 ‘혁신요인’이 가장 큰 것으로 나타났으며, 공공분야의 기반요인이 빅데이터 수요공급 예측모형에 미치는 영향은 ‘프라이버시요인’이 가장 큰 것으로 나타났다.

본 연구를 근거로 국내의 보건의료 및 공공 분야 빅데이터 수요공급의 예측과 관련하여 다음과 같은 정책적 함의를 도출할 수 있다. 첫째, 국내의 보건의료 및 공공 분야에서 빅데이터 관련 기술은 주로 고객관리에 활용되고 있는 것으로 나타났다. 이는 정부 3.0과 빅데이터는 유기적인 연관관계가 있으며, 국민중심 서비스를 위해 빅데이터 분석 결과를 기초로 수요자 맞춤형 서비스를 제공함으로써 신뢰성(Veracity)있는 새로운 가치(Value)를 창출하기 위해 빅데이터 기술을 적극 활용하고 있는 것으로 보여진다. 둘째, 보건의료 분야의 공급자는 혁신을 위하여 빅데이터를 활용하지만 수요자는 성장을 위해 빅데이터를 활용하는 것으로 나타났다. 이는 공급자는 혁신을 통하여 조직을 위기에서 벗어나게 하고 성공시키기 위해 빅데이터를 도입하지만 수요자는 경쟁력을 강화시키고 성장하기 위하여 빅데이터를 활용하는 것으로 보여진다. 셋째, 보건의료 분야에서는 혁신요인과 품질요인이 빅데이터 수요공급 예측에 미치는 영향력이 가장 큰 것으로 나타났다. 이는 조직을 성공시키고 최고로 만들기 위해 빅데이터를 활용하지만 서비스 품질을 우선적으로 고려해야 한다는 것을 나타낸다. 넷째, 공공 분야에서는 혁신요인과 프라이버시요인이 빅데이터 수요공급 예측에 미치는 영향력이 가장 큰 것으로 나타났다. 이는 조직을 혁신하고 성공시키기 위해 빅데이터를 활용하지만 프라이버시를 우선적으로 고려해야 한다는 것을 나타낸다.

3. 수요자와 공급자 중심의 빅데이터 기술 도입방안

새로운 기술의 출현은 사회를 변화시킨다. 빅데이터 기술은 이제껏 관리하기가 힘들었던 다양한 종류의 데이터와 이를 처리하고 분석하여 가치를 만들어 내는 일련의 과정을 가능케 하고 있다. 앞선 빅데이터 수요요인 분석 연구를 통해 살펴본 기업의 빅데이터 기술수용과정은 기술의 수용목적에 따라 매우 다른 구조를 가짐이 확인되었다. 이러한 분석결과를 바탕으로 성공적인 기업의 빅데이터 기술도입 방안은 다음과 같다. 첫째, 빅데이터 기술에 대한 막연한 기대감으로 기술 도입을 서두르고 있지 않는 지 기업 스스로 확인이 필요하다. 본 연구를 통해 확인된 ‘주관적 규범’요인의 영향력은 이러한 점을 환기하고 있다. 빅데이터 기술을 활용하여 어떠한 가치를 만들어 낼 것인가에 따라 적용되는 기술 및 전체적 아키텍처의 설계가 모두 달라질 수 있다. 기업 입장에서는 때로는 관리되고 있지 않던 데이터 차제를 수집하는 것이 필요할 때도 있고 금융이나 통신처럼 너무 방대한 데이터에 대한 처리구조를 효율화해야 할 필요성도 있다. 이에 따른 적용기술과 검토사항은 매우 차별화된다. 둘째, 빅데이터 기술의 지속적 확산을 위해서는 일반 이용자들도 쉽게 기술기반의 가치를 체험할 수 있는 보완기술과 서비스들이 확장되어야 할 것이다. 특히 많은 기업들이 빅데이터 수집, 처리, 분석 플랫폼에 대한 자체 구축을 검토하는데 반드시 자체 구축만이 가장 바람직한 것은 아니다. 보안에 관한 부분이 잘 관리될 수 있다면 ASP나 SaaS기반의 플랫폼을 활용하는 부분도 조기 기술적용을 통한 투자 대비 성과측면의 실패를 범하지 않을 수 있는 좋은 접근법이다.

끝으로, 빅데이터는 데이터와 기술 자체이지만 이를 활용하기 위해서는 조직의 의사결정 프로세스 및 문화적인 부분이 변화되어야 할 필요성

도 있다. 특히 본 연구에서 살펴 본 ‘기술이용자’ 집단과 ‘기술활용자’ 집단의 경우 기업 내의 조직과 업무 프로세스가 차별적인 부분을 감안하여 이들 집단에 최적화된 인적 역량과 문화적 역량이 갖추어져야 할 것이다.

4. 보건복지 빅데이터 프라이버시 보호방안

현재 어느 나라를 막론하여 빅데이터의 활용에 있어 가장 큰 과제는 개인의 사생활 비밀보호 및 개인정보보호이다. 앞에서 이미 언급한 바와 같이 개인정보보호에 중점을 두면 빅데이터의 활용을 저해하게 될 우려가 있다. 개인정보보호법 등 관련 법률을 자의적으로 해석하여 수집된 개인정보를 공공의 목적으로 활용하기 위해 제삼자에게 제공하는데 소극적인 기관이 적지 않다. 개인정보보호법의 목적이 ‘개인정보의 수집·유출·오용·남용으로부터 사생활의 비밀 등을 보호함으로써 국민의 권리와 이익을 증진하고’로 되어있지만 개인정보와 비개인정보를 명확히 구분하기가 곤란하고 비즈니스에 있어 자동적으로 수집되는 데이터가 비개인정보라고 할지라도 프라이버시를 침해할 가능성이 있다. 특히 소셜미디어에 공개된 개인정보는 위변조 오남용이 쉽고 상업적 이용을 위한 정보수집 등에 노출이 될 수 있기 때문에 프라이버시 침해 등의 문제가 발생할 가능성이 매우 높다. 개인정보의 흐름이 이미 국경을 넘어선지가 오래이다. 구글, 트위터, 유튜브 등의 글로벌 기업의 서비스에 국내법 상에 문제가 있는 부분이 있더라도 이를 국내에서 국내법의 규정에 따라 제재를 가할 수는 없다. 따라서 이제는 국제협력을 도모하면서 선진국 수준에 부응하는 프라이버시보호와 개인정보보호 제도를 정착시켜야 할 시점에 와있다. 해외에서 이미 법적인 관점에서의 개인정보보호 보다는 사회적인 관

점에서의 프라이버시 보호가 중시되는 경향으로 흐르고 있다. 방송통신위원회는 2013년 12월 18일 ‘빅데이터 개인정보보호 토론회’와 2014년 3월 19일 ‘온라인 개인정보보호 세미나’를 통해 가이드라인에 대한 의견을 수렴하고 다음과 같은 가이드라인을 제시하고 있다. 현재 빅데이터 활용에 있어 개인정보보호법의 핵심적 화두는 정보주체의 사전 동의를 받는 것이 곤란할 경우를 법적으로 어떻게 조치할 지의 문제이다. 특히, 우리나라의 개인정보 보호법제는 개인정보의 수집·이용·제공 등과 관련하여 엄격한 사전 동의 방식(Opt-in)을 채택하고 있어, 〈표 7-2〉의 가이드라인 안은 이러한 동의의 엄격한 요건을 상당부분 완화시키는 내용을 담고 있다(국회입법조사처, 2014). 본 가이드라인에 대한 시민단체의 비판핵심은 다음과 같다(국회입법조사처, 2014). 첫째 정보통신서비스 제공자는 공개된 개인정보 및 이용내역정보를 정보주체 동의 없이 수집·이용할 수 있어 문제가 있다. 이는 현행 법률상 개인정보의 수집 및 활용은 일부 예외적인 경우를 제외하고 원칙적으로 그것이 어떠한 정보이든 정보주체의 동의가 필요하다는 점을 전제로 하고 있다. 둘째, 공개된 개인정보 및 이용내역정보를 활용하여 정보주체의 동의없이 새로운 정보를 생성할 수 있는 위험성이 있다. 셋째, 공개된 개인정보를 동의없이 제 3자에게 제공할 수 있도록 하고 있어 사생활 침해의 가능성을 높인다.

현재 제시되고 있는 가이드라인안의 내용을 구현하기 위해서는 다음과 같은 관련 입법(법률)에 대한 권한을 가진 입법자들의 의사결정이 요구된다(국회입법조사처, 2014). 첫째, 동의요건의 문제는 헌법상 기본권으로서 보장되는 개인정보자기결정권의 실현 수단으로서의 성격을 가지는 것이기 때문에 이를 완화시키기 위해서는 사회적 차원의 합의가 요구된다. 둘째, 가이드라인안은 동의요건 완화를 위해 기존 개인정보 보호법제에서는 개념이 정비되어 있지 않은 ‘공개된 개인정보’, ‘이용내역정보’, ‘생

성된 개인정보’의 개념을 새롭게 제시하고 있는데, 과연 이러한 개인정보 개념이 필요한지 만일 필요하다면 어떻게 법률적 차원에서 수용해야 하는지에 대한 국회 차원의 논의가 필요하다.

〈표 7-2〉 ‘빅데이터 개인정보보호 가이드라인’의 주요내용

구분	내용
공개된 개인정보(제3조)	정보통신서비스제공자가 공개된 개인정보를 수집하고자 하는 경우에는 별도의 정보주체의 동의를 얻지 않아도 됨.
이용내역정보의 수집(제4조)	정보통신서비스 제공자는 정보주체의 거부 의사에 있지 않을 경우 서비스 계약체결과 이행을 위하여 필요한 이용내역 정보를 수집하여 조합, 분석 또는 처리하는 경우 별도로 정보주체의 동의를 얻지 않아도 됨.
새로운 개인정보의 생성(제5조)	정보통신서비스제공자는 정보주체의 거부 의사에 있지 않을 경우 정당하고 합리적인 범위 내에서 정보주체의 별도 동의없이 공개된 개인정보 및 이용내역정보 등을 활용하여 새로운 개인정보를 생성할 수 있음.
비식별화(제6조)	- 정보주체의 공익을 받거나 법령상 허용되는 경우가 아닌 한, 공개된 개인정보, 이용내역정보, 생성된 개인정보(공개된 개인정보 등)는 비식별화 조치를 취한 후 조합, 분석 또는 처리하여야 함. - 비식별화 정보는 재식별화 되지 않도록 하여야 함. - 공개된 개인정보 등의 처리과정에서 임시적으로 생성된 개인정보는 목적을 달성한 경우 지체 없이 파기하거나 비식별화되어야 함.
민감정보의 생성금지(제8조)	정보주체의 동의를 받거나 법령상 허용되는 경우가 아닌 한, 사상·신념·건강 등 정보주체의 사생활을 현저히 침해할 우려가 있는 정보의 생성을 목적으로 개인정보 등을 조합, 분석 또는 처리해서는 안됨 .
공개된 개인정보 등의 제3자 제공(제11조)	(공개된 개인정보와 이용내역정보를 활용하여) ‘새로이 생성된 정보’ 그리고 ‘이용내역정보’를 제 3자에게 제공하기 위해서는 제공받는 자, 이용목적, 항목, 보유 및 이용기간 등을 알리고 정보주체의 동의를 받아야 함 - 공개된 개인정보는 정보주체의 동의없이 제3자에게 제공 가능함.

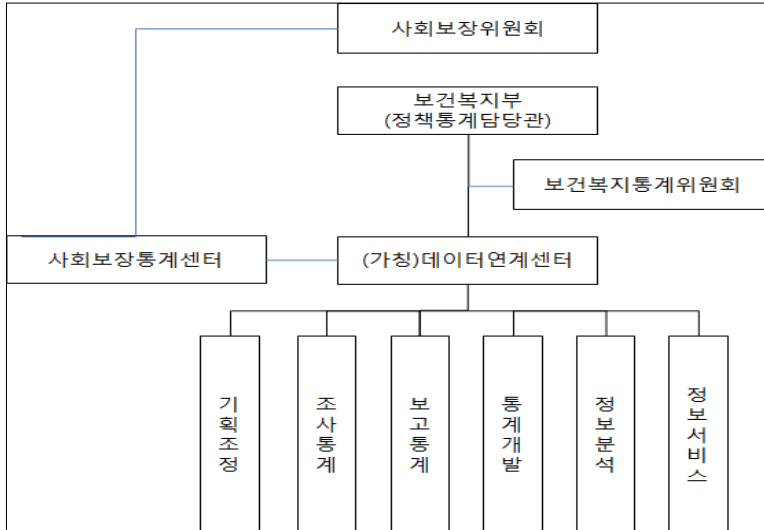
자료: 심우민, 빅데이터 개인정보보호 가이드 라인과 입법과제, 국회입법조사처, 이슈와 논점, 2014. 6. 12.

5. 보건복지 통계 효율적 관리방안

정부 3.0 및 창조경제의 추진과 실현을 위하여 현 정부의 주요 정책과 제의 지원과 함께 다양한 분야에 보건복지 통계의 활용가치가 강조되고 있다. 우리나라 보건복지 통계는 여러 기관에서 분산적으로 통계를 생산하여 큰 틀에서 체계적인 기획 및 관리가 어려운 상황이다. 보건복지 분야는 다른 분야 보다 통계정보와 수요자간의 밀접한 관계가 요구된다. 수요자 중심의 보건복지 통계를 생산하기 위해서는 보건복지부 내 관련부서, 타부처 및 지자체, 관련기관간의 효율적인 보건복지 통계생산을 위한 조정 및 네트워크 구축을 위한 컨트롤타워의 구축이 필요하다. 그리고 보건복지 통계의 계획, 생산, 관리 및 배포 등 제반 활동을 통합적이고 일관적인 체계로 관리할 수 있는 전담조직 신설 및 협업 네트워크 구축이 필요하다(그림 7-1 참조). 또한 서로 다른 기관의 데이터를 연계하여 제공해 줌으로써 데이터의 활용도를 높일 수 있는 데이터연계센터²⁶⁾가 설치되어야 할 것이다.

26) 호주의 보건사회연구원 보건복지통계센터(Australian Institute of Health and Welfare, AIHW)에서는 인구보건과 임상자료의 연계를 통해 보건정책 관련 연구의 질을 제고하고 연계된 데이터 분석을 통해 각 분야(임상, 보건, 환경 등)의 협력관계를 구축함.

[그림 7-1] 보건복지통계 관리 전담조직(가칭 데이터연계센터)



자료: 2015년 보건복지 정책과제(보건복지부-한국보건사회연구원 보건복지 정책포럼). 자료집.

6. SNA 분석을 통한 보건의료 이슈 도출 방안

앞선 보건복지이슈 분석에서는 국내 보건의료 분야 연구의 내적 지식 구조를 파악하기 위해 총 18개의 한국연구재단등재학술지로부터 2009년부터 2014년 6월까지 총 3,457편의 논문을 수집한 후, 중복 및 유사 키워드를 확인하여 분류한 후, 최종적으로 총 13,465개의 키워드를 분석하였다.

6개년도 키워드를 연결정도 중심성을 기준으로 상위 50개를 추출한 후, 이를 대상으로 네트워크를 구축한 결과, 'Social Support(사회적 지지)', 'Self esteem(자존감)', 'Quality of Life(삶의 질)', 'Health Status(건강상태)', 'Attitude(태도)', 'Anxiety(불안)', 'Female(여성)', 'Health behavior(건강 행위)', 'Health(건강)', 'Middle aged(청장년

또는 중년)'의 순으로 나타났다. 또한 이들 군집의 구조적 특성을 파악한 결과, 'Social support(사회적지지)', 'Self esteem(자존감)', 'Attitude(사고방식)', 'Self-efficacy(자기효능감)', 'Depression(우울)', 'Quality of Life(삶의 질)' 등이 중심이 되어 네트워크를 형성하고 있는 것으로 나타났다.

상위 키워드의 연도별 변화를 살펴보면, 우선, 정신·사회적 측면을 볼 때, 'Social support(사회적 지지)'와 'Social-esteem(자존감)'의 경우 최근까지도 연결정도 중심성이 높게 유지되었다. 그러나 'Self-efficacy(자기효능감)'의 경우에는 2012년도를 제외하고는 그렇지 못한 양상을 보였다. 'Anxiety(불안)'의 경우에는 2010년을 제외하고 모두 상위 키워드로 출현하고 있음을 볼 수 있었다. 'Suicide(자살)'의 경우에는 2009년과 2010년 연결정도 중심성이 높았으나 그 이후에는 그렇지 못한 반면, 2013년과 2014년 'Suicidal ideation(자살사고)'가 상위 키워드로 출현한 것을 볼 수 있었다. 이는 국내 보건의료 분야의 연구가 한국의 사회적 현상을 잘 반영하고 있는 것이라고 해석할 수 있겠다. 또한 'Quality of life(삶의 질)'의 경우, 2011년을 제외하고 모두 상위 키워드로 등장하였으며, 'Life style(삶의 방식)'과 함께 최근 활발히 연구되고 있었다.

다음으로 특정 대상과 관련한 키워드 가운데 'Mother(엄마)'의 경우에는 2011년을 제외하고는 상위 키워드로 출현하지 않았으며, 'Nurses(간호사)'의 경우 2013년 이후 상위 키워드로 유지되고 있었다. 'Female(여성)'의 경우에는 2009년과 2014년에는 상위 키워드로 출현하지 않았으나 그 간 지속적으로 연구되어 온 것으로 나타났다. 더불어 여성과 관련이 있는 'Pregnancy(임신)'의 경우에도 지속적이지는 않으나 간헐적으로 상위 키워드로 출현하고 있었다. 이는 곧 남성에 비해 사회적으로 약자에 해당하는 여성에 대한 연구가 관심을 받고 있다는 것을 입증하고 있

다고 해석할 수 있겠다.

2011년 이후 'Immigrants(이민자)'가 상위 키워드로 출현한 것은 국내 보건의로 분야의 연구가 다문화사회로 진입하고 있는 한국의 사회현상을 잘 반영하고 있다는 것으로 해석할 수 있겠다. 'Middle aged(중년층)'에서 2013년 'Elders(노인)'가 상위 키워드로 등장하면서 생애주기별 연구의 관점에서 볼 때, 연구대상이 확대되고 있음을 알 수 있었으며, 이 또한 고령화 사회로 급속하게 변화하고 있는 한국의 사회현상을 잘 반영하고 있는 것이라 할 수 있겠다.

'Health(건강)'와 'Health status(건강상태)'의 경우, 지속적으로 상위 키워드로 출현하고 있었으며, 이는 'Health behavior(건강행위)'로 그 범주가 확장되고 있음을 볼 수 있었다. 이는 2010년을 제외하고는 모두 상위 키워드로 등장하였으며, 'Smoking(흡연)', 'Smoking cessation(금연)', 'Exercise(운동)', 'Physical activity(신체활동)' 등의 관련 키워드와 함께 지속적으로 연구되고 있는 것으로 나타났다.

특정 질환 관련 키워드는 다양한 질환이 지속적이지는 않으나 간헐적으로 상위 키워드로 출현하고 있었다. 이와 관련한 키워드는 2009년과 2010년 'Osteoarthritis(골관절염)', 2010년 'Cancer(암)', 2013년 'Stroke(뇌졸중)', 'Hypertension(고혈압)', 2014년 'Breast cancer(유방암)' 등이다.

마지막으로 상위 키워드의 연도별 응집 구조를 종합적으로 해석하면, 'Adolescents(청소년)'와 'Social support(사회적 지지)', 'Self-esteem(자존감)', 'Self-efficacy(자기효능감)', 'Smoking(흡연)', 'Suicide(자살)' 등이 서로 연결되어 연구되고 있었으며, 'Suicide(자살)'는 다시 'Depression(우울)', 'Stress(스트레스)', 'Social support(사회적 지지)', 'Self-esteem(자존감)' 등과 함께 연결되어 연구되고 있었다.

‘Nurses(간호사)’는 ‘Job stress(직무 스트레스)’와 함께 연구되고 있으며, ‘Elders(노인)’, ‘Female(여성)’, ‘Children(어린이)’, ‘Immigrants(이민자)’ 등이 새로운 연구대상으로 떠오르고 있었다. 따라서 청소년을 대상으로 한 다양한 방면에서의 연구가 지속적으로 수행되어 오고 있으며, 최근 대두되고 있는 노화, 다문화, 사회적 약자에 대한 연구 등 국내 보건의료 분야의 연구가 한국의 사회적 변화를 잘 반영하고 있다고 해석할 수 있겠다.

본 연구를 통한 SNA 분석 시 고려사항은 다음과 같다. 첫째, 키워드 분석시 MeSH 용어를 사용하지 않고 저자가 부여한 키워드를 사용하는 경우, 색인자 효과가 발생한다(Lee, Kim, Park, & Lee, 2006; Kim & Lee, 2007; Mika, 2007). 또한, 동일한 의미의 용어가 다르게 표기된 경우 등 키워드의 색인에 표준화가 되어있지 않아 연구자가 논문의 키워드를 1차로 색인한 후 2차로 자동 색인기를 사용하여 색인된 저자 키워드를 표준화시키는 과정을 거쳐야 한다(Wang, Wang, Ma, & Hu, 2007). 그러나 본 연구는 이러한 과정을 거치는 대신 저자 키워드를 추출하여 그 키워드를 정제하는 과정에서 다양한 변인을 고려하기 위하여 보건의료 분야 전문가의 검토를 거쳤다(Barrat, Barthelemy, Pastor-Satorras, & Vespignani, 2004).

둘째, SNA 분석을 위해 개발된 연결 중심성은 노드 간 연결에 대한 가중치를 반영하지 않는 단점이 있다(Osphal, Agneessens, & Skvoretz, 2010). 본 연구는 연구주제의 중심성을 파악하기 위해 연결된 키워드 수 뿐만 아니라 연결된 횟수에 따라 가중치를 반영하여 가중치가 높을수록 연결정도 중심성이 상승하도록 하였으며(Adamic LA, Lukose RM, Puniyani AR, Huberman BA., 2001; Barrat, Barthelemy, Pastor-Satorras, & Vespignani, 2004). 하지만 이러한 가중치 부여가

연구주제의 질적 수준을 전적으로 반영할 수 있는 것은 아니므로 학술지의 영향력 지수를 반영하는 등 다양한 방법에 의한 상대적인 가중치를 반영하는 등의 대안 마련이 필요할 것이다.

셋째, 자료수집 단계에서 추출된 모든 키워드를 네트워크상에 가시화시킬 수 없으므로 본 연구는 동시 출현 빈도 상위 키워드만을 분석 대상으로 하였다. 따라서 분석 대상에 포함되는 키워드는 전체에 비해 제한적이라 할 수 있다. 따라서 네트워크상에서 배제된 키워드는 연구결과 해석에서 제외될 수밖에 없다는 분석의 한계를 갖는다(Jang, Lee, & An, 2012).

향후 SNA 연구를 위한 제언을 정리하면 다음과 같다. 첫째, 보건의료 분야의 축적되는 다양한 형태의 자료들을 지속적으로 관리함으로써 이러한 분석을 주기적으로 진행해야 할 것이다. 둘째, SNA 분석 결과의 검증 위하여 해당 분야의 자문위원을 구성함으로써 전문가의 심층적인 검토를 거쳐야 할 것이다. 셋째, 보건의료 분야의 연구 영역을 확장하기 위하여 키워드 이외의 연구자 및 그들의 소속 기관 등의 추가 정보를 확장시켜 연결함으로써 연구 분야 뿐 아니라 연구자 네트워크를 체계적으로 구성해 나가야 할 것이다.

7. 소셜 빅데이터 분석방안

빅데이터 분야에서 ‘소셜 애널리틱스(Social Analytics)’가 페이스북, 트위터, SNS 등에서 수집되는 비정형 데이터를 신속하게 분석을 한다. 소셜미디어에서 정보를 뽑아내고 분석하는 방법은 크게 3가지로 나눌 수 있다(송태민, 2013). 첫째, 텍스트마이닝(Text Mining)은 인간이 언어로 쓰인 비정형 텍스트에서 자연어처리기술을 이용하여 유용한 정보를

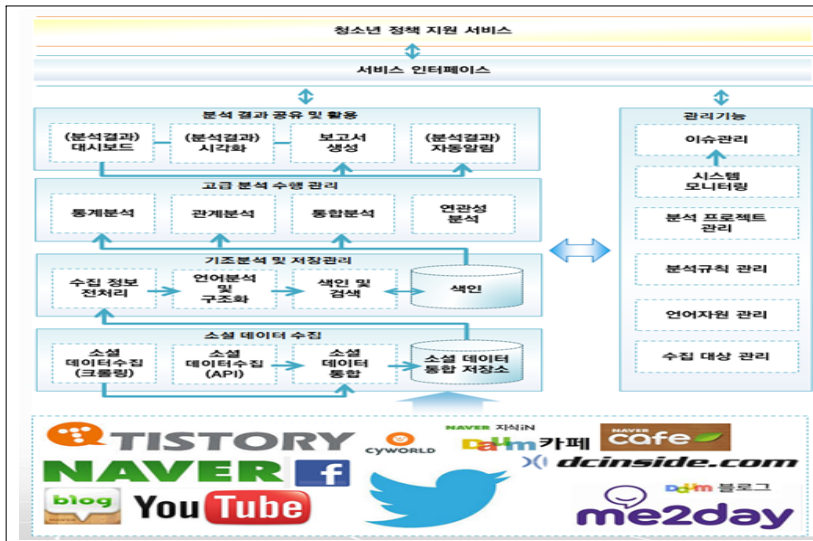
추출하거나, 연계성을 파악, 분류 혹은 군집화, 요약 등 빅데이터의 숨겨진 의미 있는 정보를 발견하는 것이다. 둘째, 오피니언마이닝(Opinion Mining)은 소셜미디어의 텍스트 문장을 대상으로 자연어처리기술과 감성분석 기술을 적용하여 사용자의 의견을 분석하는 것으로 마케팅에서는 버즈(Buzz; 입소문)분석이라고도 한다. 셋째, 네트워크분석(Network Analytics)은 네트워크 연결구조와 연결강도를 분석하여 어떤 메시지가 어떤 경로를 통해 전파가 되는지, 누구에게 영향을 미칠 수 있는지를 파악하는 것이다.

소셜데이터의 수집은 [그림 7-2]와 같이 일반적인 웹환경(HTTP, RSS)에서 수집가능 정보들을 웹크롤러를 통하여 수집하고, 연계정보는 각 출처에서 제공하는 Open API를 이용하여 필요정보를 수집한다(권정은·정지선, 2012)²⁷⁾. 수집데이터는 데이터의 중복성 및 품질을 검사하여 유의미한 정보만을 선별한 후 분석과정에서 쉽게 처리하기 위해 다양한 수집형식을 고려하여 일관된 방식을 저장 및 관리를 해야 한다. 기초분석을 위해 형태소분석을 통하여 추출된 텍스트 데이터의 언어적 형태를 분석하여 구성요소들(명사, 동사, 형용사, 전치사, 조사 등)을 식별 및 분류한다. 구문분석을 통하여 언어적 구성요소들의 배치나 구조적 특성을 분석하여 의미적 연관관계를 유추하고, 감성분석을 통해 내용에 언급된 감성들의 표현들을 선별하여 감성 표현의 대상을 식별하고, 감성의 종류를 구분한다. 분석 기초자료인 언어분석을 위한 각종용어사전, 개체명사전, 이형태어사전, 감성어사전, 분류체계, 분류규칙, 분류학습 데이터 등을 지속적으로 수정·관리해야 한다. 분석결과와 공유를 위해서는 대쉬보드 및 상황판을 구현하여 분석된 결과를 도형이나 표를 이용하여 쉽게 파악

27) 권정은, 정지선(2012). 청소년 위기극복을 위한 빅데이터 기반 정책시나리오. 한국정보화진흥원.

할 수 있는 환경을 조성하고 분석된 결과를 실무보고서에 신속히 적용할 수 있도록 보고서에서 활용가능한 형태의 시각화 결과와 엑셀 등 구조화된 형태로 분석결과를 제공해야 한다.

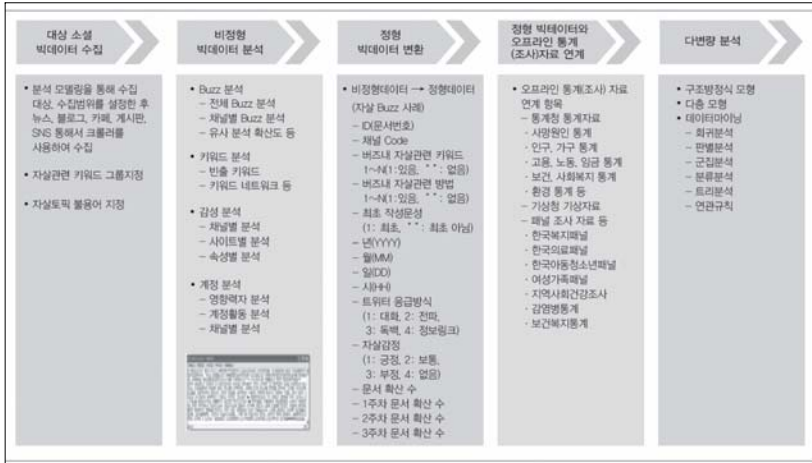
[그림 7-2] 청소년 자살징후 포착을 위한 소셜 분석 시스템 구성도(안)



소셜 빅데이터 분석 절차 및 방법은 [그림 7-3]과 같다(송태민·송주영, 2013). 첫째, 해당 주제와 관련한 문서를 (자살)분석 모델링을 통해 수집대상과 수집범위를 설정한 후, 대상채널(뉴스·블로그·카페·게시판·SNS 등)에서 크롤러 등 수집엔진(로봇)을 이용하여 수집한다. 이때 불용어를 지정하여 수집의 오류를 방지하고 자살 관련 연관 키워드 그룹(원인·유형·대상·성별·장소·지역·방법 등)과 지정한다. 둘째, 수집한 비정형 데이터를 분석한다. 비정형 데이터 분석은 버즈 분석, 키워드분석, 감성분석, 계정분석 등으로 진행한다. 청소년 자살버즈 수집 사례와 같이 비정형 데이터를 연구자가 수집한 원상태로 분석하기에는 어려움이 있

다. 따라서 수집한 비정형 데이터를 텍스트마이닝(text mining), 오피니언마이닝(opinion mining), 네트워크 분석(network analysis)을 통하여 분류하고 정제하는 절차가 필요하다. 셋째, 비정형 빅데이터를 정형 빅데이터로 변환해야 한다. 자살 관련 주제 분석 시의 문서분석 사례를 살펴보면, 자살버즈 각각의 문서는 ID로 코드화하여야 하고, 버즈 내 키워드나 방법 등도 모두 코드화하여야 한다. 넷째, 사회현상과 연계해 분석하기 위하여 정형화된 빅데이터를 오프라인 통계(조사) 자료와 연계해야 한다. 오프라인 통계(조사) 자료는 대부분 정부나 공공기관에서 유료 또는 무료로 제공하기 때문에, 연계대상 자료와 함께 연계 가능한 ID(일별·월별·연별·지역별)를 확인한 후 오프라인 자료를 수집하여 연계(merge)할 수 있다. 다섯째, 오프라인 통계(조사) 자료와 연계된 정형화된 빅데이터의 분석은 요인 간의 인과관계나 시간별 변화궤적을 분석할 수 있는 구조방정식모형이나 일별(월별·연별), 지역별 사회현상과 관련된 요인과의 관계를 분석할 수 있는 다층모형, 그리고 수집된 키워드의 분류 과정을 통해 새로운 현상을 발견할 수 있는 데이터마이닝 분석을 실시할 수 있다.

[그림 7-3] 소셜 빅데이터 분석 절차 및 방법(자살버즈 분석 사례)



빅데이터 연계[big data linkage(matching)] 방법으로는 정확매칭(exact matching)과 통계적매칭(statistical matching)이 있다. 정확매칭은 고유식별 정보가 존재하여 동일한 개체를 연결함으로써 상호 매칭이 되며 주로 행정데이터 간의 연계나 행정데이터와 조사데이터 간 연계를 할 경우 사용된다. 통계적매칭은 고유식별 정보가 존재하지 않기 때문에 유사한 개체를 찾아 상호 데이터를 결합시키는 방법으로 이질적인 조사데이터와 조사데이터 연계나 이질적인 행정데이터와 조사데이터 연계, 그리고 정형데이터와 비정형데이터 연계를 할 경우 사용된다. 소셜 빅데이터와 공공기관 빅데이터[예: 청소년 전화(1388) 상담 빅데이터]의 연계 방법은 [그림 7-4]와 같이 일별 날짜를 고유식별 정보로 활용하는 정확매칭 방법을 활용할 수 있다.

[그림 7-4] 소셜 빅데이터와 정형 빅데이터 매칭

소셜 빅데이터	1388 정형 빅데이터	최종 연계 빅데이터
일자	일자	일자
성적	가출	성적
성	가정폭력	성
우울	학업중단	우울
경제	학교폭력	경제
질병	성	질병
외모	흡연음주	외모
자살위험	자살	자살위험
	인터넷중독	가출
	온둔형	가정폭력
	비행	학업중단
		학교폭력
		성
		흡연음주
		자살
		인터넷중독
		온둔형
		비행

8. 소셜 빅데이터 분석 기반 보건복지 정책 수립

앞선 보건복지 정책 수요 예측 연구에서는 국내의 온라인 뉴스 사이트, 블로그, 카페, 소셜 네트워크 서비스, 게시판 등 인터넷을 통해 수집된 소셜 빅데이터를 데이터마이닝의 의사결정나무 분석기법을 적용하여 분석함으로써 한국의 보건복지 정책 수요에 대한 예측모형을 개발하고자 하였다. 본 연구의 결과를 요약하면 다음과 같다. 첫째, 보건복지 정책과 관련한 버즈는 의료민영화, 과다복지, 증세, 양육수당, 원격의료, 연금, 건강보험 등의 순으로 많이 언급되는 것으로 나타났다. 둘째, 청와대와 관련한 보건복지 정책 관련 버즈는 의료민영화, 증세, 연금, 양육수당, 과다복지 등의 순이며, 보건복지부는 의료민영화, 건강보험, 행복온도, 연금, 원격의료, 양육수당 등의 순으로 나타났다. 그리고 시민단체와 관련한 정책 버즈는 의료민영화, 원격의료, 건강보험, 연금, 양육수당, 기초생활 등

의 순으로 나타났다. 셋째, 보건복지 관련 정책은 전세대책, 양육수당, 기초생활, 건강보험, 퇴직연금, 증독법, 원격의료 정책에 대한 찬성의 확률이 높은 것으로 나타났다. 그리고 과다복지, 증세, 반값등록금, 무상정책, 의료민영화, 연금 정책은 반대의 확률이 높은 것으로 나타났다. 넷째, 보건복지 정책 수요 예측에 가장 영향력이 높은 정책은 ‘의료민영화요인’으로 나타났다. 따라서 보건복지 정책의 찬성에 가장 영향력이 높은 경우는 ‘의료민영화요인’이 높고 ‘원격의료요인’이 높은 조합으로 나타났다. 반면, ‘의료민영화요인’이 높고 ‘원격의료요인’이 낮고, ‘증세요인’이 높으면 보건복지 정책의 반대할 확률이 높은 것으로 나타났다. 본 연구를 근거로 한국의 보건복지 정책 수요에 대한 예측과 관련하여 다음과 같은 정책적 함의를 도출할 수 있다.

첫째, 의료민영화(포괄수가제, 의료상업화, 민영화, 의료민영화), 연금(기초연금, 국민연금), 원격의료(법인약국, 의료영리화, 원격의료, 원격의료), 건강보험(국민건강보험, 민간보험, 건강보험)에 대해서 보건복지 정책의 관심이 많은 것으로 나타났다. 이는 최근 정부의 원격의료 추진과 의료법인의 영리 자법인 허용 계획, 그리고 기초연금의 논의와 무관하지 않은 것으로 보여진다. 둘째, 원격의료 정책은 찬성의 확률이 높은 것으로 나타났지만 의료민영화 요인은 반대의 확률이 높은 것으로 나타났다. 따라서 정부는 원격의료 정책이 의료민영화의 시작이라는 인식을 불식시키기 위한 의료계와의 충분한 논의와 합의가 있어야 할 것으로 본다. 셋째, ‘의료민영화요인’이 높고 ‘원격의료요인’이 낮고, ‘증세요인’이 높은 집단의 보건복지 정책에 대한 반대가 높은 것으로 나타났다. 따라서 보건복지 정책의 실행을 위한 정부의 증세정책을 최소화하고 의료민영화가 배제된 원격医료를 실행할 수 있는 계획을 수립해야 할 것으로 본다.

본 연구는 개개인의 특성을 가지고 분석한 것이 아니고 그 구성원이 속

한 전체 집단의 자료를 대상으로 분석하였기 때문에 이를 개인에게 적용하였을 경우 생태학적 오류(ecological fallacy)가 발생할 수 있다(Song et al., 2014). 또한, 본 연구에서 정의된 보건복지 관련 요인(용어)은 버즈 내에서 발생한 단어의 빈도로 정의되었기 때문에 기존의 조사 등을 통한 이론적 모형에서의 의미와 다를 수 있으며, 2개월간의 제한된 소셜 빅데이터를 분석함으로써 보건복지 정책의 수요예측에 한계가 있을 수 있다. 본 연구는 소셜 빅데이터에서 수집된 빅데이터 주제분석과 데이터마이닝 분석을 통하여 한국의 보건복지 정책에 대한 수요 예측모형을 제시한 점에서 정책적·분석방법론적으로 의의가 있다. 또한, 실제적인 내용을 빠르게 효과적으로 파악하여 사회통계가 지닌 한계를 보완할 수 있는 새로운 조사방법으로서의 빅데이터의 가치를 확인하였다는 점에서 조사방법론적 의의를 가진다고 할 수 있다. 끝으로 국민이 필요한 보건복지 욕구를 사전에 파악하여 보건복지 수요를 정확·신속하게 예측하고 계획을 수립하기 위해서는 오프라인 보건복지 수요 조사와 함께 소셜 빅데이터의 지속적인 활용과 분석을 통한 과학적 보건복지 정책수립이 필요할 것으로 본다.

참고문헌 <<

- 강만모·김상락·박상무 (2012). 빅데이터의 분석과 활용. 정보과학회지. 30(6), pp.25~32.
- 강소라·전방지 (2007). 과업기술적합도(TTF)가 그룹지원시스템(GSS)의 사용 및 성과에 미치는 영향: 전유방식동의 정도와 전유충실도의 조절효과를 고려하여, 기술혁신학회지. 10(4), pp.755~788.
- 강원섭·김종우 (2010). 비만의 정신과적 접근, 대한비만학회지. 19(3), pp.71~77
- 건강보험공단 (2012). 보도자료(2012.11.29.); 흡연, 음주, 비만으로 인해 지출되는 진료비 규모 건강보험 전체 진료비 중 14.5% 차지
- 김광재 (2007). DMB의 수용모형 구축에 관한 연구, 한국언론학회 추계학술대회집. pp.43~47.
- 김경미·염유식·박연민 (2013). 한국청소년의 학교폭력 경험과 심리적 안녕: 주관적 행복감과 자살충동을 중심으로, 한국콘텐츠학회논문지. 13(9), pp.236~247.
- 김규동 (2008). 와이브로 서비스 채택요인에 관한 연구; 확장된 TAM적용, 계명대학교 대학원 박사학위 논문.
- 김미선 (2010). IPTV 이용자의 이용인식과 수용결정요인에 관한 연구, 한국언론정보학보. pp.177~202.
- 김민경 (2012). 대학생의 부정적 정서, 가족탄력성 및 자존감이 자살 생각에 미치는 경로 모형 분석, 한국가족복지학. 17(1), pp.61~83.
- 김민규·김주환 (2007). 실감공간기술 잠재적 사용자들의 수용의도에 영향을 미치는 요인에 대한 분석, 한국 HCI학회.
- 김선아 (2009). 초등학교시기의 심리사회적요인과 중학생시기의 자살생각간의 관계: 스트레스, 우울감과 사회적 지지를 중심으로, 한국가족복지학회, 14(3), pp.5~27.
- 김상훈 (2008). 하이테크 마케팅, 서울: 박영사.

- 김성개 (2009). 사용자 환경과 스마트폰 특성요인이 인지된 유용성과 사용용이성 및 수용의도에 미치는 영향에 관한 연구, (석사학위), 홍익대학교 광고홍보대학원.
- 김성일 (2011). 혁신 확산이론에 따른 스마트폰 수용의도에 관한연구: 스마트폰미사용자를 중심으로, (박사학위), 세종대학교대학원.
- 김성희 (2013). 장애노인의 실태와 과제, 보건·복지 Issue & Focus. 208호.
- 김은영·이정훈·서동욱 (2013). 빅데이터 시스템의 수용의도에 영향을 미치는 수용조직의 환경요인에 관한 연구, 한국데이터베이스학회. 20(4), 1~18.
- 김은정 (2013). 소득계층별 출산행태 분석과 시사점. 보건·복지 Issue & Focus. 191호.
- 김정선·권은주·송태민 (2014). 분석지의 확장을 위한 소셜 빅데이터 활용연구 -국내 '빅데이터' 수요공급 예측-. 지식경영연구. 15(3). pp.173~192.
- 김정태·오봉진·박종열 (2013). 빅데이터 핵심기술 및 표준화 동향, 한국전자통신연구원 전자통신동향분석, pp.92~99.
- 김지현·김홍·손광은·송윤수·윤지혜·임희창·정상호·전주홍 (2014). 빅데이터의 의학적활용, 정보과학회지. 32(3), pp.18~26.
- 김해룡·양필석 (2008). 핵심자기평가와 혁신행동과의 관계에서 셀프리더십의 매개효과에 관한 연구, 대한경영학회지. 21(5), pp.1831~1857.
- 김혜련·강영호·곽노성·강은정·김어지나 (2009). 부문간 협력을 통한 비만 예방관리체계 구축 방안: 비만의 역학적 특성 분석과 비만예방관리를 위한 부문간 협력체계의 탐색, 한국보건사회연구원.
- 김현주 (2008). 청소년 자살위험성에 영향을 미치는 요인, 한국아동복지학회, 27, pp.22~26.
- 노진석 (2012). 빅데이터와 소셜 분석: 빅데이터의 바다에서 '의미'를 찾다.
http://www.imaso.co.kr/?doc=bbs/gnuboard.php&bo_table=article&wr_id=40725
- 대한내분비학회·대한비만학회 (2010). 비만 치료지침 2010 권고안, EndocrinolMetab, 25(4), pp.301~304.

- 대한비만학회 (2010). 대한비만학회 Fact Sheet: 비만, 만병의 바로미터. 대한비만학회.
- 민경훈·이태용 (2011). 성별과 연령을 고려한 비만지표 평가, 한국 산학 기술학회 논문지. 12(9), pp.4046~4053.
- 미래창조과학부(2013). 2013년 인터넷 이용 실태조사, 미래창조과학부.
- 박은아 (2007). 뉴미디어 초기수용자와 비수용자의 비교, 한국방송학회 학술대회 논문집. pp.154~161.
- 백상용 (2009). 조절변수 탐색을 위한 기술수용 메타분석, 경영학연구. 38(5), pp.1353~1380.
- 배상민·우종민 (2011). 의학적 측면의 자살방지 대책: 외국사례를 중심으로, 대한의사협회지. 2011. 54(4). pp.386~391.
- 변숙은 (2010). 하이테크 서비스 초기 수용자들의 인식연구, 마케팅관리 연구. 15(2), pp.43~69.
- 서신자·정민숙 (2013). 가족 탄력성이 청소년의 자살 생각에 미치는 영향: 고등학교 3학년 남학생들 대상으로, 한국 콘텐츠 학회논문지. 13(5), pp.262~274.
- 서정한 (2013). 기업용 클라우드 서비스 도입 결정요인에 관한 실증연구, (박사학위), 한양대학교 대학원.
- 손대경 (2011). 대장암 지식지도 구축 및 연구 네트워크 분석, (박사학위), 충북대학교 대학원.
- 송태민 (2013). 일본의 빅데이터 동향. 보건복지포럼. 통권 제 204호.
- 송태민 (2012). 보건복지 빅데이터 효율적 활용방안. 보건복지포럼. 통권 제 193호.
- 송태민 (2014). 일본의 빅데이터 프라이버시 보호방안. 보건복지포럼. 통권 제 210호.
- 송태민·송주영 (2013). 빅데이터 분석방법론, 한나래아카데미.
- 신현웅 (2013). 건강보험 진료비 분석 및 정책방향. 보건·복지 Issue & Focus. 186호.

- 오미애 (2014). 정부 3.0과 빅데이터: 보건복지 분야 사례를 중심으로. 보건·복지 Issue & Focus. 제230호.
- 오태동 (2009). 유비쿼터스 서비스 수용모형에 관한 연구, (박사학위), 단국대학교 대학원.
- 유재현·박철 (2010), 기술수용모델(Technology Acceptance Model) 연구에 대한 종합적 고찰, LG CNS 엔트루 정보기술 연구소, 정보기술 엔트루지. 9(2), pp.31~50.
- 윤미영 (2013). 공공부문의 빅데이터 활용사례 및 정책적 활성화 방안, 부동산포커스.
- 윤명숙 (2012). 대학생의 우울, 취업 스트레스가 자살 생각에 미치는 영향, 한국청소년학회. 19(3), pp.109~137.
- 이경일 (2006). 온톨로지 구축과 의미 메타데이터 관리. 한국정보문화진흥원, 한국제이터베이스진흥센터, 메타데이터표준화포럼(2006.05.25.)
(Retrieved 06/18, 2014), from
(<http://www.dbguide.net/upload/20060614/1150261432107.pdf>)
- 이대기 (2012). 중소기업의 내부역량과 기술혁신성과의 관계 및 외부네트워크와 조직문화의 조절, (박사학위), 계명대학교 대학원.
- 이문선·강영순 (2003). 창의성과 혁신행동의 관계와 집단특성의 조절효과, 인사관리연구. 27(1), pp.251~271.
- 이윤경·염주희·이선희 (2013). 고령화 대응 노인복지서비스 수요전망과 공급체계 개편, 한국보건사회연구원.
- 이장우·장수덕 (2010). 환경의 불확실성, 의사결정과정 그리고 기업성과 간의 관계: 정보시스템의 조절영향을 중심으로, 경영학연구. 39(5), pp. 1363~1387.
- 이현미 (2008). 웨어러블 컴퓨터의 수용과 소비자 세분화에 관한연구; 혁신기술수용모델을 중심으로, (박사학위), 이화여자대학교.
- 임형주 (2008). EMR시스템 도입 특성요인이 지각된 유용성, 편의성 및 수용의도에 미치는 영향에 관한 연구, (박사학위), 창원대학교 대학원.

- 장활식·박광오·한정희 (2008). ERP 구현단계별 활동과 변화관리 활동의 통합 이 ERP 시스템 도입성과에 미치는 영향, 한국산업정보학회논문지. 13(3), pp.63~86.
- 장상현 (2012). 빅데이터와 스마트교육, 한국정보과학회지. 30(6), pp..59~64.
- 전은주 (2013). 임상실무 지침기반 스마트폰 비만관리 어플리케이션 개발, (석사 학위논문), 서울대학교.
- 정지선 (2011). 신가치창출 엔진, 빅데이터의 새로운 가능성과 대응전략, 한국정보화진흥원.
- 조항정·송찬후·강현구·임동원 (2009). 국내 대기업들의 IT 거버넌스: 비교 사례 연구, 기술혁신학회지. 12(2), 335~359.
- 질병관리본부.(2011). 2011 국민건강통계: 국민건강영양조사. 2차년도 (Ed.).
- 최민수 (2011). 개인의 혁신성, 사회적 영향력, 사용자 인터페이스 요인이 스마트폰 수용에 미치는 영향에 관한 연구; 확장된 기술수용모델을 중심으로, (박사학위), 이화여자대학교 대학원.
- 최중명·김춘배 (2011). 비만관리와 과학적근거, 대한의사협회지. 54(3), pp.250~265.
- 탁기천·류규수 (2001). 성인비만의 건강특성에 관한 분석적 연구, 한국보건교육, 건강증진학회지. 18(2), pp. 115~128.
- 테이코산업연구소 (2013). 빅데이터 관련시장 실태와 전망, 테이코, 2013.
- 한국과학기술정보연구원 (2013). 빅데이터 산업의 현황과 전망, KISTI MARKET REPORT.
- 한국보건산업진흥원·특허청(2013). 사회보장부분 빅데이터 R&D 사업기획연구 특허동향.
- 한국정보화진흥원 (2010). Gov 2.0 시대의 공공정보 및 공공서비스 활성화 전략, CIO Report.
- 함유근·채승병 (2012). 빅데이터 경영을 바꾸다, 삼성경제연구소.
- 황승구·최완·허성진·장명길·이미영·박종열·원희선·김달 (2013). 빅데이터 플랫폼 전략, 전자신문사.

홍영수·전선영 (2005). 청소년의 생활스트레스가 자살생각에 미치는 영향과 우울의 매개효과, 한국정신보건사회복지학회. 19, pp.125~149.

통계청 (2013). 청소년 통계, 통계청.

통계청 (2013). 사망원인통계, 통계청

独立行政法人情報処理推進機構 (2012). くらしと経済の基盤としてのITを考え
る研究会報告書, つながるITがもたらす豊かなくらしと経済
日本ネットワークセキュリティ協 (2012). 2012年情報セキュリティインシデント
にする調査報告書-個人情報ぞい編

Adamic LA·Lukose RM·Puniyani AR·Huberman BA (2001). Search in power-law network. *Phys Rev E*.

Adams, D. A.·Nelson, R. R.·Todd, P. A. (1992). Perceived usefulness, ease of use, and usage of information technology: a replication. *MIS quarterly*. pp.227~247.

Agarwal, R.·Prasad, J. (1997). The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies, *Decision Sciences*. 28(3), pp.557~582.

Ajzen, I. (1991). The Theory of Planned Behavior, *Organizational behavior and decision processes*. 50, 179~211.

AlloyLB·HartlageS·AbramsonLY (1998). Testing the cognitive diathesis-stress theories of depression: Issues of research design, conceptualization, and assessment. In Alloy LB(Ed.). Cognitive Processes in depression. New York, NY(NY): *The Guilford Press*. pp.31~73.

Amabile, T. M. (1988), The model of creativity and innovation in organizations, In, B.M. Staw & L.L. Cummings(Eds.), *Research in Organizational Behavior*. 10, pp.123~167.

- An JY·Jang H·Paik J (2014). Social network analysis on consumers' seeking behavior of health information via the Internet and mobile phones. *Journal of Korean Multimedia Society*. 17(8).
- An, J. Y·Hayman, L. L·Panniers, T. L·Carty, B. (2007), Theory development in nursing and healthcare informatics: a model explaining and predicting information and communication technology acceptance by healthcare consumers. *Advances in Nursing Science*. 30(3), pp.37~49.
- Barrat A·Barthelemy M·Pastor-Satorras R·Vespignani A. (2004). The architecture of complex weighted network. *PENS* 101(11). pp.3747~3752.
- Beautrais AL. (2003). Suicide and serious suicide attempts in youth: A Multiple-Group comparison study. *Am J Psychiatry*. 160(6), pp.1093~1099.
- Becerra-Fernandez, I·Sabherwal, R. (2001), Organizational knowledge management: A contingency perspective, *Journal of management information systems*. 18(1), pp.23~56.
- Boyack,K.W.·Klavans,R.·BÖner,K. (2005). Mapping the backbone of science. *Scientometrics*. 64, pp.351~374.
- Bridge JA·Brent DA·Johnson BA·Connolly J. (1997). Familial aggregation of psychiatric disorders in a community sample of adolescents. *J Am Acad child Adolesc Psychiatry*. 36, pp.628~637.
- BrownDR·BlantonCJ (2012). Physical activity, sport participation, and suicidal behavior among college students. *Med Sci Sports Exerc*. 34(7). pp.1087~1096.
- CEBR(2012) *C.entre for Economics and Business Research*, Data equity: Unlocking the value of big data

- Chen C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization PANS 2004; 101 suppl.1, pp.5303~5310.
- Chen, W. L. C·Bagakas, J. G. (2003). Understanding the dimensions of self-exploration in web-based learning environments. *Journal of Research on Technology in Education*. 34(3), pp.364-373.
- Chin, W. W.·Todd, P. A. (1995). On the use, usefulness, and ease of use of structural equation modeling in MIS research: a note of caution. *Management Information Systems Quarterly*. 19, pp.237~237.
- Courtial JP (1994). A co-word analysis for Scientometrics. *Scientometrics*. 31(3). pp.251~260.
- Davis, F. D. (1989), Perceived usefulness, perceived ease of use and user acceptance of information technology, *MIS Quarterly*. 13(3), pp.319~340.
- DeLone, W. H.·McLean, E. R. (1992). Information Systems Success: The Quest for the dependent variables. *Information Systems Research*. 3(1), pp.60-95.
- Dishaw, M. T.·D. M. Strong (1999). Extending the Technology Acceptance Model with Task-Technology Fit construct, *Journal of Mathematical Psycology*. 41, pp.260~274.
- Drucker, P. F. (2007), Innovation and entrepreneurship: *Practice and principles*, Routledge.
- Effken J·Gephart S·Bianchi D·Verran J. (2010). Correlations of nursing communication network metrics with patient outcomes. *American Medical Informatics Association Annual Symposium Process*. pp.197201.
- Estabrooks,C.A.·Winther,C.·Derksen,L. (2004). Mapping the field:

- Abibliometric analysis of the research utilization literature in nursing, *Nursing Research*. 53, pp.293~303.
- Fergusson DM·Lynskey MT. (1995). Childhood circumstances, adolescent adjustment, and suicide attempts in a New Zealand birth cohort. *J Am Acad child Adolesc Psychiatry*. 34(5), pp.612~622.
- Fishbein, M.·Ajzen, I. (1981), *Acceptance yielding and impact: cognitive processes in persuasion*.
- Fishbein, M.·Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*.
- Folz-Murphy N·Partin M·Williams L·Harris CM·Lauer MS (1998). Physician use of an ambulatory medical record system: matching form and function. In A Paradigm Shift in Health Care Information Systems: *Clinical Infrastructures for the 21st Century, Proceedings of the 1998 AMIA Annual Symposium Edited by Chute CG*. pp.260~264.
- Goldston DB·Daniel SS·Reboussin DM·Frazier PH·Harris AE. (2001). Cognitive risk factors and suicide attempts among formerly hospitalized adolescents: A prospective naturalistic study. *J am Acad Child Adolsec Psychiatry*. 40(1), pp.91~99.
- Goodhue, D. L.·Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly*. 19(2), pp.213~236.
- Goodhue,D.L.(1995). Understanding User Evaluations of Information Systems, *Management Science*. 41(12), pp.1827~1844.
- Goodhue, D. L. (1998). Development and Measurment Validity of a task-Technology Fit Instrument for User Evaluations of Information Systems, *Decision Sciences*. 29(1), pp.105~137.
- Goodhue, D. L.·Klein, B.·March, S. (2000). User Evaluation of IS as

- Surrogates for Objective Performance. *Information and Management*. 38(2), pp.87~101.
- Grover, V.·Goslar, M. D. (1993), The initiation, adoption, and implementation of telecommunications technologies in US organizations, *Journal of Management Information Systems*. 10(1), pp.141~163.
- Guan, J.·Yam, R.·Mok, C.·Ma, N. (2006), A study of the relationship between competitiveness and technological innovation capability based on DEA models. *European Journal of Operational Research*. 170(3), pp.971~976.
- Hair, J. F.·Tatham, R. L.·Anderson, R. E.·Black, W. (2006). Multivariate data analysis, 6, Upper Saddle River, NJ: Pearson Prentice Hall.
- He Q. (1999). Knowledge discovery through co-word analysis, *Library Trends*. 8(1), pp.133~159.
- Heather LR·Michael AB·Nishad K. (2010). Youth Net Hamilton. Youth engagement and suicide risk: Testing a mediated model in a Canadian community sample. *J Youth Adolescence*. 39(3). pp.243~258.
- Hendrickson, A. R.·Massey, P. D.·Cronan, T. P. (1993). On the test-retest reliability of perceived usefulness and perceived ease of use scale, *MIS Quarterly*. pp.227~230.
- Hu, L. T.·Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*. 6(1), pp.1~55.
- Huh S. (2010). Medical Databases from Korea and Abroad, *Journal of Korean Medical Association*. 53(8), pp.659~667.
- Igarria, M.·Zinatelli, N.·Cragg, P.·Cavaye, A. L. (1997). Personal

- computing acceptance factors in small firms: a structural equation model. *MIS quarterly*. 21(3).
- Izadinia N·Amiri M·Jahromi RG·Hamidi S. A (2010). study of relationship between suicidal ideas, depression, anxiety, resiliency, daily stresses and mental health among Teheran university students. *Procedia Soc Behav Sci*. 5, pp.515~519.
- Jackson, C. M·Chow, S·Leitch, R. A. (1997). Toward an understanding of the behavioral intention to use an information system, *Decision Sciences*. 28(2), pp.357~389.
- Jang H·An JY (2014). Social network analysis on health literacy and the use of online health information by the aged, *Healthcare Informatics Research*. 20(3).
- Jang H·Lee YS·AnJY (2012). Application of social network analysis to health Care sectors. *Health care Informatics Research*. 18(01), pp.44~56.
- Jang HL·Kang GW·Tak YJ·Lee YS. (2011). Analysis of medical articles published domestic and abroad by Korean researchers from 1960 to 2008. *Journal of KLISS*. 45(3), pp. 259~277.
- Jeong S·Lee SK·Kim HG (2010). Knowledge structure of Korean medical informatics: A social network analysis of articles in journal and proceedings, *Health Inform Res*. 16(1), pp.52~59.
- Jung MS·Jung DJ (2008). Co-author keyword network and their clustering appearance in preventive medicine fields in Korea: Analysis of paper in the Journal of Preventive Medicine and Public Health. *J Prev Med Public Health*. pp.1-9.
- Jung MS (2010). Academic research activities and their co-author and keyword network in epidemiology fields: Analysis of papers in Korean Journal of Epidemiology. *J Korean Epidemiol*. 30(1).

pp.60~67.

Kang JH, Park HY. (2010). Analysis of scientific publication networks among medical schools in Korea, *Health Inform Res.* 16(2). pp.100~119.

Kaufman JM. (2009). Gendered responses to serious strain: The argument for a general strain theory of deviance. *Justice Q.* 26(3), pp.410~444.

Kendel DB·Ravis VH·Davies M. (1991) Suicidal ideation in adolescence: Depression, substance abuse, and other risk factor. *J Youth Adolesc.* 20(2), pp.289-309.

Kim BS·Yun DH (2011). Prevention of suicide infection on part of advance guard in media, *Newspapers and Broadcasting.* 7, pp.22~26.

Kim JI·Park SM·Park HS·Chung CW·Ahn SH (2007). Analysis of published papers and their keywords in the Korean Journal of Women Health Nursing (2003-2006). *Journal of Korean Academy of Women Health Nursing.* 13, pp.51~59.

Kim PJ·Lee JY (2007). Descriptor profiling for research domain analysis. *J Korean Soc Inform Manage.* 24(4), pp.285~302.

Kim SY (2008). From MeSH indexed to Retrieval. 1st ed. Seoul, Korea: *Korean Medical Library Association.*

Kim HY·Park HA·Min YH·Jeon E (2013). Development of an obesity management ontology based on the nursing process for the mobile-device domain. *J Med Internet Res.* 15(6), e130. doi: 10.2196/jmir.2512

Kim J·Lee J (2002). Critical design factors for successful e-commerce systems, *Behaviour and Information Technology.* 21(3), Vol. 21 pp.185~189.

- Konick LC (2005). Gutierrez PM. Testing a model of suicide ideation in college students, *Suicide and Life Threatening Behavior*. 35(2). pp.181~192.
- Kwon AK·Chae YM (2002). The study on subject words of Korean medical informatics by expanded MeSH: Based on journal of the Korean society of medical informatics. *Health Inform Res*. 8(4), pp.91~98.
- Kwon TH·Zmud RW (1987). Unifying the fragmented models of information systems implementation, In Critical issues in information systems research, *John Wiley & Sons, Inc.*, pp.227~251.
- Larsen TJ·Sorebo AM·Sorebo O (2009). The role of task-technology fit as users' motivation to continue information system use, *Computers in Human Behavior*. 25(3), pp.778~784.
- Lawson B·Samson D (2001). Developing innovation capability in organizations: a dynamic capabilities approach. *International Journal of Innovation management*. 5(3), pp.377~400.
- Lee SK·Jeong S·Kim HG·Yum YH (2011). A social network analysis of research topics in Korean nursing science. *J Korean Acad Nurs*. 45(5), pp.623~632.
- Lee WH·Kim YM·Park KR·Lee MH (2006). A study on the emerging technology mapping through co-word analysis. *J Korean Operations Res Manage Sci*. 23(3). pp.77~93.
- Lee YS (2010). *Research network analysis for the national science knowledge map*. Seoul Korea: Korea Research Council of Fundamental Science and Technology.
- Lim JC·Yun JK (1999). The effects of supervisor's transformational and transactional leadership on subordinate's innovativeness: The

- role of self-efficacy as a mediator. *Korean Journal of Management*. 7(1), pp.1~42.
- Malin B·Carley K (2007). A longitudinal social network analysis of the editorial boards of medical informatics and bioinformatics journals, *J Am Med Inform Assoc*. 14, pp.340~348.
- Mane KK·Borne K (2004). Mapping Topics and burst in PANS. *Proc Natl Acad Sci*. 101(1), pp.5287~5290.
- Mathieson, K (1991). Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior, *Information Systems Research*. 2(3), pp.173~191.
- McKinsey Global Institute (2011). *Big data: The next frontier for innovation, competition, and productivity*.
- Meltzer H·Harrington R·Goodman R·Jenkins R (2001). *Children and adolescents who try to harm, hurt or kill themselves*. National Statistics London Office.
- Mika P (2007). *Social Networks and Sementic Web*, SIKSD is sertation Series No. 2007-03. p33-52.
- Moon JW·Kim YG (2001). Extending the TAM for a World-Wide-Web context, *Information & Management*. 38, pp.217~230.
- Moore GC·Benbasat I (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation, *Information Systems Research*. 2(3), pp.192~222.
- National Information Society Agency (2012). *Implications for suicide prevention policy of youth described in the social Analysis*. Seoul, Korea: National Information Society Agency.
- NHLBI. (1998). *Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: the evidence report* NIH Publication No. 98-4083. Bethesda: The National

- Heart, Lung, and Blood Institute.
- NICE. (2006). *Obesity: The Prevention, Identification, Assessment and Management of Overweight and Obesity in Adults and Children NICE clinical guideline 43*. Hoborn, London: National Collaborating Center for Primary Care and the Centre for Public Health Excellence at NICE.
- Osphal T·Agneessens F·Skvoretz J (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Netw.* 32, pp.245~251.
- Park JS·Park CJ (1995). An analysis of nursing research on pain reported in Korea from 1970 to 1994, *Journal of Korean Academy of Nursing*. 25, pp.30~44.
- Park JY·Lim YO·Yoon HS (2010). Suicidal impulse caused by stress in Korea: Focusing on mediation effects of existent spirituality, family support, and depression. *Kor J Soc Welfare Studies*. 41(4). pp.81~105.
- Peruzzi N·Bongar B (1999). Assessing risk for completed suicide in patients with major depression: Psychologists' views of critical factors, *Professional Psychology: Research and Practice*. 30(6), pp.576~580.
- Pete Warden (2011). *Big Data Glossary*, O'Reilly Media.
- Policy Exchange (2012). *The Big Data Opportunity*.
- Price DD (1965). The pattern of bibliographic references indicates the nature of the scientific research front. *Science*. 149, pp.510~515.
- Rogers EM (1983). *Diffusion of Innovations*, New York: The Free Press.
- Rogers EM (2004). Networks, firm size and innovation, *Small Business Economics*. 22(2), pp.141~153.

- Roh KH (2008). A review of the trends of journal research on dementia and nursing interventions for demented elders. *Journal of Korean Academy of Community Health Nursing*. 19, pp.300~309.
- Rubenstein, AH·Chakrabarti AK·O'eefe RD·Souder WE·Young HC (1976). Factors influencing innovation success at the project level, *Research Management*. pp.15~20,
- Rudd MD·Rajab MH (2001). *Treating suicidal behavior: An effective time-limited approach*. New York (NY): Guilford Press.
- Scott SG·Bruce RA (1994). Determinants of innovative behavior: A path model of individual innovation in the workplace. *Academy of management journal*. 37(3), pp.580-607.
- Shah J·Higgins ET (1997). Expectancy× value effects: Regulatory focus as determinant of magnitude and direction, *Journal of Personality and Social Psychology*. 73(3), pp.447~458.
- SongTM· Song J·An JY, Hayman LL·Woo JM (2014). Psychological and Social Factors Affecting Internet Searches on Suicide in Korea: A Big Data Analysis of Google Search Trends. *Yonsei Med J*. 55(1), pp.254~263.
- Song TM·Song JY·An JY·Jin DL (2013). Multivariate Analysis of Factors for Search on Suicide Using Social Big Data. *Korean Journal of Health Education and Promotion*. 30(3), pp.59~73.
- Stack S (2005). Suicide in the media: a quantitative review of studies based on non-fictional stories. *Suicide Life Threat Behav*. 35, pp.121~133.
- Straub DW·Keil M·Brenner W (1997). Testing the technology acceptance model across cultures: a three country study, *Information & Management*. 33, pp.1~11.

- Subramanian GH (1994). A replication of perceived usefulness and perceived ease of use measurement, *Decision Sciences*. 25(5/6), pp.863~874.
- Suh YO·Park JS·Yang JH·Kim HW·Suk MH·Shin HS, et al. (2007). Analysis of research papers published in the Journal of Korean Academy of Nursing. *Journal of Korean Academy of Nursing*. 37, pp.1013~1019.
- Szajna B (1994). Software evaluation and choice: predictive validation of the technology acceptance instrument, *MIS Quarterly*. pp.319~324.
- Taylor S·Todd P (1995). Assessing IT usage: the role of prior experience, *MIS Quarterly*, 19(4), pp.561~570.
- Toprak S·Cetin I·Guven T·Can G·Demircan C (2011). Self-harm, suicidal ideation and suicide attempts among college. *Psychiatry Res*. 87, pp.140~144.
- Tornatzky LG·Fleischer M·Chakrabarti AK (1990). *Processes of technological innovation*.
- Tousignant M·Mishara BL·Caillaud A·Fortin·V·St-Laurent D (2005). The impact of media coverage of the suicide of a well-known Quebec reporter: the case of Gaëtan Girouard. *Soc Sci Med*. 60, pp.1919~1926.
- Turecki G (2005). Dissecting the suicide phenotype: The role of impulsive-aggressive behaviours. *J psychiatry Neuro sci*. 30(6), pp.398~408.
- Venkatesh V·Davis FD (2000). A theoretical extension of the technology acceptance model: four longitudinal field studies, *Management Science*. 46(2), pp.186~204.
- Venkatesh V·Morris MG (2000). Why do not men ever stop to ask for

- directions? Gender, social influence, and their role in technology acceptance and usage behavior, *MIS Quarterly*. 24(1), pp.115~139.
- Venkatesh V·Bala H (2008). Technology acceptance model 3 and a research agenda on interventions, *Decision sciences*. 39(2), pp.273~315.
- Wang X·Wang J·Ma F·Hu C (2007). The small-world characteristic of author co-words network. *IEEE*. pp. 3717-3720.
- Wilburn VR·Smith DE (2005). *Stress, self-esteem, and suicidal ideation in late adolescents*. Adolescence. 40, pp.33~45.
- Yoo B·Donthu N (2001). Developing and validating a multidimensional consumer-based brand equity scale, *Journal of Business Research*. 52, pp.1~14.
- Zhu K·Kraemer KL·Gurbaxani V·Xu SX (2006). Migration to open-standard interorganizational systems: network effects, switching costs, and path dependency, *MIS Quarterly*. pp.515~539.
- Zigurs I (2000). A theory of task-technology fit and group support systems effectiveness, *MIS Quarterly*. 22(3), pp.313-334.
- 세계인권선언: Universal Declaration of Human Rights
<http://www.humanrights.com/what-are-human-rights/universal-declaration-of-human-rights.html>.
- 통신프라이버시: Communicational privacy:
http://itlaw.wikia.com/wiki/Right_of_privacy.
- Gartner (2012). <http://www.gartner.com/newsroom/id/2124315>.
- Gartner (2011). Hype Cycle for Emerging Technologies. 2011:
<http://www.gartner.com/newsroom/id/2124315>

- Crime mapping: <http://www.crimemapping.com/map.aspx>
- IDC (2012). http://bits.blogs.nytimes.com/2012/03/07/idc-sizes-up-the-big-data-market/?_php=true&_type=blogs&module=Search&abReward=relbias%3Ar&r=0
- IDC (2014). <http://www.idckorea.com/product/getdoc.asp?idx=612&field=pressrelease>
- TDWI (2011). Big Data Analytics: <http://www.TDWI.org/>.
- The Economist (2010). <http://www.economist.com/node/15557443/>.
- The right to control information about oneself: http://en.wikipedia.org/wiki/Right_to_privacy.
- The Foresight Horizon Scanning Centre:
<https://www.gov.uk/government/groups/horizon-scanning-centre>
- RAHS: <http://app.nscs.gov.sg/public/home.aspx>
- Wikipedia (2014). Obesity: <http://en.wikipedia.org/wiki/Obesity>

<부록 1>

1. 보건복지 공공데이터 현황

가. 보건분야

1) 건강보험

<부표 1-1> 건강보험분야 공공데이터

번호	기관	데이터명	서비스유형
1	부산광역시 남구	건강기능식품 판매업소 현황	다운로드
2	전라북도 남원시	산모신생아건강관리사 제공 기관 현황	다운로드
3	건강보험심사평가원	건강보험 통계연보	다운로드
4		건강보험 통계자료	LINK
5		고령환자데이터셋	LINK
6		도서정보	LINK
7		병원약국찾기 정보	LINK
8		보도자료 정보	OPENAPI
9		소아청소년환자데이터셋	다운로드
10		심사기준 정보	다운로드
11		요양기관자원정보	LINK
12		요양기관현황정보	LINK
13		의약품 안전성 정보	LINK
14		의약품 표준코드 정보	LINK
15		입원환자데이터셋	다운로드
16		전체환자데이터셋	다운로드

번호	기관	데이터명	서비스유형
17		진료경향모니터링 정보	다운로드
18		질병통계	다운로드
19		청구소프트웨어인증 정보	LINK
20		행위통계 정보	LINK
21	국민건강 보험공단	Web_대표홈페이지 방문현황	다운로드
22		가상계좌 수납 현황	다운로드
23		건강IN 제공정보 현황	다운로드
24		건강보험이의신청 심의 내역	다운로드
25		결산지원(연도비교재무상대표)	LINK
26		노인장기요양보험 급여비 지급 현황	LINK
27		노인장기요양보험 운영센터별 등급판정결과 현황	LINK
28		노인장기요양보험 인정현황	LINK
29		독촉고지(4대보험료)	LINK
30		마이헬스뱅크(My Health Bank) 이용 현황	다운로드
31		만성질환자 건강지원서비스 교육프로그램 일정	LINK
32		보험료부담대비급여비현황	다운로드
33		복지용구 품목별 이용현황	다운로드
34		사회공헌	다운로드
35		심사청구 심의 내역	다운로드
36		임신출산진료비 지원내역 현황	LINK
37		정기고지	다운로드
38		현금급여(요양비) 지급현황자료	다운로드
39		현금급여(자가도뇨소모성재료) 지급 현황	다운로드
40		현금급여(출산비) 지급현황	다운로드
41		현금급여비(가정산소치료) 지급현황	다운로드
42		현금급여비(도단위별) 지급현황	다운로드
43		현금급여비(장애구 보장구) 지급현황	LINK
44		현금급여비(제1형 당뇨환자소모성재료) 지급현황	LINK
45		현금급여비(광역시별) 지급현황	LINK

2) 보건의료

〈부표 1-2〉 보건의료분야 공공데이터

번호	기관	데이터명	서비스유형
1	강원도속초시	약국현황	LINK
2		한의원현황	LINK
3		치과의원현황	LINK
4		병의원현황	LINK
5	강원도원주시	원주시약국정보	LINK
6		원주시치과의원정보	다운로드
7		원주시병원정보	다운로드
8		원주시한의원정보	다운로드
9		원주시의원정보	다운로드 LINK
10		원주시미용업(종합)소정보	다운로드
11		원주시미용업(일반)소정보	다운로드
12		원주시이용업소정보	다운로드
13		원주시네일아트업소정보	다운로드
14		원주시미용업(피부)소정보	다운로드
15		원주시동물병원정보	다운로드
16	강원도홍천군	생활정보/위생정보/식품위생위반공표	LINK
17	강원도횡성군	의약정보	다운로드
18	경기도과천시	과천시한의원정보	LINK
19		과천시치과정보	LINK
20		과천시일반의원정보	LINK
21		과천시의료기기판매업소정보	다운로드 LINK
22		과천시안경업소정보	LINK
23		과천시금연구역정보	다운로드
24	경기도구리시	구리시소독업소현황	LINK
25		구리시치과의원현황	LINK

번호	기관	데이터명	서비스유형
26		사회복지시설현황	LINK
27		구리시한의원현황	LINK
28		구리시병의원현황	LINK
29		공중화장실관리	LINK
30	경기도군포시	군포시이·미용업현황	다운로드
31	경기도남양주시	남양주요양병원현황	다운로드
32		남양주병원현황	다운로드
33		남양주치과의원현황	다운로드
34		남양주의원현황	다운로드
35		남양주연합의원현황	다운로드
36		남양주한의원현황	다운로드
37		남양주응급의료기관현황	다운로드
38		남양주응급면동별의료기관현황	다운로드
39		남양주응급면동별약업소현황	다운로드
40		남양주예방접종현황	다운로드
41		남양주약국현황	다운로드
42		남양주보건지소및진료소현황	다운로드
43		남양주보건소방문자현황	다운로드
44	경기도동두천시	사회복지시설현황	다운로드
45		숙박업현황	다운로드
46	경기도부천시	희귀난치의료비지원사업	다운로드
47		오정구의료기관,약국,의료기기판매업소현황	다운로드
48		오정구산림및녹지대화장실현황	다운로드
49	경기도성남시	남시동물병원현황	그리드 다운로드 OPEN API
50		남시수정구병원현황	
51		남시수정구약국현황	
52		남시분당구의료기관현황	
53		남시종원구의료기관현황	

번호	기관	데이터명	서비스유형
54		남시분당구필수예방접종위탁의료기관현황	
55		남시중원구필수예방접종위탁의료기관현황	
56		남시수정구필수예방접종위탁의료기관현황	
57		남시분당구산후조리원현황	
58		남시중원구산후조리원현황	
59		남시수정구산후조리원현황	
60	경기도수원시	방역소독업체현황	LINK
61		예방접종실적	LINK
62		의료기관현황	LINK
63	경기도시흥시	시흥시응급의료(심장충격기)설치현황	다운로드
64		시흥시안경업소정보	다운로드
65		시흥시의료기관정보	다운로드
66	경기도안산시	단원구안전상비의약품판매업소정보	다운로드
67		상록구약국의약품정보	다운로드
68		상록구안전상비의약품판매업소정보	다운로드
69		상록구병원급의료기관정보	다운로드
70	경기도안양시	안양시예방접종위탁의료기관현황	다운로드
71		안양시소독업체현황	다운로드
72	경기도양평군	양평군의료기관현황2013_11	다운로드
73		양평군소독업체현황2013_11	다운로드
74		양평군사회복지시설2013_11	다운로드
75	경기도여주시	의료기관현황	다운로드
76		여주시소독업체현황	다운로드
77	경기도용인시	처인가한의원정보	다운로드
78		처인가치과의원정보	다운로드
79		처인가병의원정보	다운로드
80		수지구한의원정보	다운로드
81		수지구치과의원정보	다운로드

번호	기관	데이터명	서비스유형
82		수지구암검진기관정보	다운로드
83		수지구병의원정보	다운로드
84		기흥구한의원정보	다운로드
85		기흥구치과의원정보	다운로드
86		기흥구암검진기관정보	다운로드
87		기흥구병의원정보	다운로드
88		기흥구당변약국정보	다운로드
89		안전상비의약품판매업소정보	다운로드
90	경기도파주시	의료기관현황	LINK
91	경기도하남시	하남시건강및감염병정보	LINK
92	경기도화성시	화성시지역사회통합건강증진사업	LINK
93		화성시정신보건시설현황	다운로드
94		화성시정신및알코올상담센터현황	다운로드
95		화성시의료지원정보	다운로드 LINK
96		화성시심내혈관질환예방사업	LINK
97		화성시소독업소현황	다운로드
98		화성시산후조리원현황	다운로드
99		화성시금연클리닉	LINK
100		화성시금연사업전반	LINK
101		화성시금연구역	LINK
102		화성시금연관련법규	LINK
103		화성시동물병원	다운로드
104		2013화성시보건소주요업무보고(주요행사)	다운로드
105		2013화성시보건소주요업무계획	다운로드
106		화성시건강나눔이동버스	LINK
107		화성시보건소기록물관리현황	다운로드
108		화성시방문보건사업	LINK
109	경상남도김해시	예방접종참여의료기관DB	다운로드
110	경상남도남해군	음식점정보데이터	LINK

번호	기관	데이터명	서비스유형
111	경상남도밀양시	국가필수예방접종	다운로드
112	경상남도산청군	예방접종정보	다운로드
113		보건지소및보건진료소정보	다운로드
114		약국정보	다운로드
115		진주시휴일당번약국정보	다운로드
116	경상남도진주시	진주시휴일당번치과정보	다운로드
117		진주시의료기관정보	다운로드
118		진주시의약업소정보	다운로드
119	경상남도창녕군	의료기관현황	다운로드
120	경상남도하동군	동군의료기관안내DB	그리드 다운로드 OPEN API
121	경상남도함안군	휴일당번약국현황	LINK
122	경상북도김천시	김천시식품위생업소현황	다운로드
123	경상북도문경시	문경시병원약국정보	다운로드
124	경상북도영주시	의료기관	다운로드
125		의약업소현황	다운로드
126	경상북도의성군	관내의료기관현황	LINK
127	경상북도청송군	의료기관현황	LINK
128	경상북도포항시	포항시관내의료기관안내(병의원)	다운로드
129	경상북도포항시	포항시관내의료기관안내(약국)	다운로드
130	광주광역시	협동조합등록현황	다운로드
131	광주광역시북구	식품·공중위생관계업소통계	다운로드
132	광주광역시 지방공무원교육원	의무실운영관리매뉴얼	다운로드
133	대구광역시	대구시외국환자와언어소통가능병원	LINK
134		한방유물정보DB	다운로드
135		안전상비의약품판매자현황	LINK
136		대구시의료기관현황	다운로드 LINK
137	대구광역시달서구	소독업소현황	다운로드

344 보건복지 빅데이터 효율적 관리방안 연구

번호	기관	데이터명	서비스유형
138		의료유사업종현황	다운로드
139		의료기기판매소현황	다운로드
140		의료기관현황	다운로드
141		정신보건시설현황	다운로드
142		설명절당직의약업소	다운로드
143		산후조리원현황	다운로드
144		달서구소독업소현황	다운로드
145		공중위생관리최우수업소현황	다운로드
146	대구광역시달성군	응급실설치기관현황	다운로드
147		약국현황	다운로드
148		한의원	다운로드
149		치과의원	다운로드
150		병원의원현황	다운로드
151		필수예방접종참여의료기관	다운로드
152	대구광역시북구	공중위생업소서비스평가	다운로드
153		1366종합상담서비스기관안내	다운로드
154	대구광역시서구	약국현황DB	다운로드
155		의료기관DB	다운로드
156	대구광역시수성구	수성구산후조리원현황	다운로드
157	대구광역시중구	보건(의료기관)	다운로드
158	대구광역시 시설관리공단	명복공원화장정보	LINK
159	대전광역시대덕구	구강보건	LINK
160		의료기관현황	다운로드
161		의약품관리	다운로드
162		감염병관리	다운로드
163	대전광역시서구	의료기관현황	다운로드
164		약국현황DB	다운로드 LINK

번호	기관	데이터명	서비스유형
165		의료기기판매업소DB	다운로드
166		당직의료기관DB	
167		산후조리원현황DB	
168		안경업소현황DB	
169	대전광역시중구	병원및의원현황DB	LINK
170		공중화장실민간위탁시설현황DB	LINK
171	부산광역시	외국인유치진료등록의료기관현황	다운로드
172	부산광역시기장군	치과의원	다운로드
173		의원	다운로드
174		기타업소	다운로드
175		의료기기판매업소	다운로드
176		안경업소	다운로드
177		응급의료기관안내	다운로드
178		약국	다운로드
179		한의원	다운로드
180		병원	다운로드
181	부산광역시남구	의료법인	LINK
182		의료기기판매업현황	LINK
183		의료기관	LINK
184		안경업소	LINK
185		공중위생업소대장	LINK
186	부산광역시북구	소독업소현황	다운로드
187		안경업소현황	다운로드
188		의료기기판매업소현황	다운로드
189		병의원, 약국현황	다운로드
190		보건소설치운영현황	다운로드
191	부산광역시사하구	동물병원현황	LINK
192	부산광역시서구	약국현황	LINK

346 보건복지 빅데이터 효율적 관리방안 연구

번호	기관	데이터명	서비스유형
193	부산광역시수영구	치과기공소현황	다운로드
194		지역보건의료계획	다운로드
195		특수의료장비병상수공동활용현황	다운로드
196		의료기기판매업소현황	다운로드
197		의료기관현황	다운로드
198		아동복지시설현황	다운로드
199		소독업소현황	다운로드
200		공중위생업소현황(미용, 이용, 숙박, 세탁, 위생관리용역업)	다운로드
201	부산광역시연제구	의료기관현황	다운로드
202		약업소현황	다운로드
203		공중위생업소관리현황	LINK
204	부산광역시영도구	특수의료장비현황	다운로드
205		진단용방사선발생장치현황	다운로드
206		약국현황	다운로드
207		한의원현황	다운로드
208		치과현황	다운로드
209		병의원현황	다운로드
210	부산광역시중구	소독업소현황	LINK
211		안경업소현황	LINK
212		의료기기판매업소현황	LINK
213		의료급여대상자현황	다운로드
214	부산광역시 해운대구	세탁업현황	다운로드
215		이용업현황	다운로드
216		미용업현황	다운로드
217		목욕장업현황	다운로드
218		동물병원현황	다운로드
219		소독의무시설현황	다운로드 개방예정
220	서울특별시	서울시복지건강정보	LINK
221	서울특별시강동구	강동구관내외국어지원가능약국	다운로드

번호	기관	데이터명	서비스유형
222		강동구관내외국인진료의료기관	다운로드
223		강동구관내연중무휴약국	다운로드
224		강동구관내산후조리원	다운로드
225		강동구관내병의원약국(강동구홈페이지보건소)	다운로드
226	서울특별시관악구	관악업소(병의원)현황DB	다운로드
227	서울특별시노원구	건강검진기관	다운로드
228		병원현황	다운로드
229	서울특별시서초구	외국어가능약국현황	LINK
230		외국어가능의약업소	LINK
231	서울특별시송파구	야간및공휴일진료의료기관	다운로드
232	세종특별자치시	세종시의원현황	LINK
233		와이팜	LINK
234		세종시의료정보	다운로드
235		병원현황	다운로드
236	울산광역시	의료/공중위생단체정보	LINK
237		병원/의원/약국현황	LINK
238		심폐소생자동제세동기설치현황	LINK
239		공중위생우수업소	LINK
240		공중위생민원안내	LINK
241	인천광역시	정신의료기관및시설현황	OPEN API
242		인천광역시의료기관및약국현황정보	LINK
243		공공보건의료기관현황	LINK
244	인천광역시강화군	강화군의료,스포츠관련업체정보	다운로드
245		강화군병의원약국현황	다운로드
246		강화군강화군병의원약국진료시간	다운로드
247	인천광역시남구	송의보건지소재활치료실장비목록	다운로드
248		남구사회복지시설현황	다운로드
249		남구공중(개방)화장실현황	다운로드

번호	기관	데이터명	서비스유형
250	인천광역시남동구	남동구한의원현황	다운로드
251		남동구병원(의원)현황	다운로드
252	인천광역시동구	의료기관현황	다운로드
253	인천광역시서구	의료기관(의원)현황	다운로드
254		서구약국현황	다운로드
255		병원현황	다운로드
256	인천광역시연수구	의료기기판매업소현황	다운로드
257		의원현황	다운로드
258		약국현황	다운로드
259		안경업소현황	다운로드
260	인천광역시중구	중구공중화장실현황	LINK
261		의료기관현황	LINK
262		약국현황	LINK
263	인천광역시의료원	2010년도의료원사업운영계획	다운로드
264	전라남도	병·의원현황	다운로드
265		약국현황	다운로드
266		응급의료기관현황	다운로드
267		보건소현황	다운로드
268		병원급이상의료기관현황	다운로드
269		다문화가족지원센터현황	다운로드
270	전라남도나주시	의약업소현황	다운로드
271		위생업소현황	다운로드
272		기초생활수급자구분별총괄현황	다운로드
273	전라남도목포시	목포시약국현황	다운로드
274	전라남도목포시	목포시병원현황	다운로드
275	전라남도보성군	약국현황	다운로드
276		보성상수도정수장수질검사결과현황	다운로드
277		병의원현황	다운로드

번호	기관	데이터명	서비스유형
278	전라남도순천시	공중위생서비스평가등급	LINK
279		숙박업소현황	LINK
280		세탁업소현황	LINK
281		휴일당번약국현황	LINK
282	전라남도신안군	신안군보건기관현황	다운로드
283		신안군의약업소현황	다운로드
284	전라남도여수시	여수시병원현황	다운로드
285		여수시의료기관현황	다운로드
286	전라남도영암군	의약업소현황	다운로드
287		사회복지시설	다운로드
288	전라남도장성군	임도시설현황	LINK
289	전라남도진도군	의료기관및약국현황	다운로드
290		미용실현황	다운로드
291	전라남도함평군	감염병발생현황	다운로드
292		함평군출생관련자료	다운로드
293		함평군금연정책	다운로드
294		함평군안경업소현황	다운로드
295		함평군의료기기판매업소	다운로드
296		약국현황	다운로드
297		함평군병원,의원현황	다운로드
298	전라남도화순군	의료기관현황	다운로드
299		약국현황	다운로드
300	전라북도남원시	목욕업소현황	LINK
301		치과기공소현황	다운로드
302		안경업소현황	다운로드
303		의료기관현황	다운로드
304		공중위생관리등급현황	LINK
305	전라북도부안군	진단용방사선발생장치검사정보	LINK
306		의료기관정보	다운로드

번호	기관	데이터명	서비스유형
307		공중위생업소정보	다운로드
308	전라북도장수군	공중위생관리등급현황	LINK
309	전라북도전주시	병원·약국정보	OPEN API
310		개방화장실	OPEN API
311	전라북도정읍시	의료기기	LINK
312		안전상비의약품	LINK
313		의약품한약도매	LINK
314	충청남도계룡시	보건소및보건지소정보	다운로드
315	충청남도논산시	재가장기요양기관현황	다운로드
316		노인요양로시설현황	다운로드
317	충청남도당진시	당진시의료기관현황	다운로드
318	충청남도부여군	부여군의약품판매업소현황	다운로드
319		부여군병원현황	다운로드
320		공중위생업소현황(미용업)	다운로드
321	충청남도서산시	산시소독업체현황	그리드 다운로드 OPEN API
322		산시약국현황	
323		산시의료기관현황	
324		산시공중위생업소(미용업)	
325		산시공중위생업소(미용업)	
326		산시공중위생업소(세탁소)	
327		산시공중위생업소(목욕장업)	
328	충청남도서천군	약국현황	다운로드
329		약국현황	다운로드
330		한의원현황	다운로드
331		치과병원	다운로드
332		병원현황	다운로드
333	충청남도태안군	의약품소현황	
334	충청남도홍성군	홍성군세탁업현황	다운로드
335		홍성군목욕업소현황	다운로드

번호	기관	데이터명	서비스유형
336		홍성미용업소현황	다운로드
337		홍성군이용업현황	다운로드
338		홍성군병의원약국현황	LINK
339	충청북도보은군	의료기기판매업소현황	다운로드
340		의약업소현황	다운로드
341	충청북도영동군	영동군 의약품취급업소현황	다운로드
342		영동군 의료기관(한의원)현황	다운로드
343		영동군 의료기관(치과의원)현황	다운로드
344		영동군 의료기관(일반의원)현황	다운로드
345		영동군 의료기관(병원)정보	다운로드
346	충청북도증평군	당번약국현황	LINK
347		의료기관현황	LINK
348	충청북도청원군	의료기간현황	다운로드
349	충청북도청주시	지역사회건강조사결과	다운로드
350		청주시상당구 의료기관등업소현황	다운로드
351		청주시상당구 야간진료 의료기관현황	다운로드
352		의료기관/약국등업소안내	다운로드
353		복지관현황	다운로드
354		열린화장실등록현황	LINK
355	충청북도충주시	산후조리시설현황	
356	충청북도충주시	병원원관리	다운로드
357	국립암센터	암등록통계자료	LINK
358		국립암센터홈페이지	LINK
359	국립중앙의료원	전국자동심장충격기정보조회서비스(AED)	OPEN API
360		전국공공병원데이터	LINK
361		전국응급의료정보조회서비스	OPEN API
362		전국병·의원찾기서비스	OPEN API
363		전국명절비상진료기관정보	OPEN API
364		전국약국정보	OPEN API

352 보건복지 빅데이터 효율적 관리방안 연구

번호	기관	데이터명	서비스유형
365	국민건강보험공단	Web_노인장기홈페이지방문자현황	다운로드
366		주요수술통계	LINK
367	보건복지부 건강정책국	청구현황DB	LINK
368		보건기관정보DB	LINK
369	보건복지부 국립공주병원	기초정보DB	다운로드
370		원무통계DB	다운로드
371		진료통계DB	다운로드
372		비급여정보	다운로드
373		의료진정보	다운로드
374	보건복지부 국립나주병원	진료비청구현황	다운로드
375		입원재원자수현황	다운로드
376		외래환자수현황	다운로드
377		외래/입원수입금현황	다운로드
378	보건복지부 국립부곡병원	국립부곡병원진료통계	다운로드
379		국립부곡병원약물중독자치료보호실적	다운로드
380		국립부곡병원고객만족도조사	다운로드
381	보건복지부 국립서울병원	비급여수가및의약품목록	다운로드
382		서지정보	LINK
383	보건복지부 국립재활원	의료진(진료과)별내원환자통계정보	다운로드
384		상병별환자통계정보	다운로드
385		장애원인별통계정보	다운로드
386		장애인건강정보	다운로드
387		장애인보조기구품목및분류체계	다운로드
388		진료정보	다운로드
389	보건복지부 국립춘천병원	환자진료정보	다운로드
390		진료비청구현황	다운로드
391		입원재원자수현황	다운로드
392		외래/입원수입금현황	다운로드
393		외래환자수현황	다운로드

번호	기관	데이터명	서비스유형
394	보건복지부 질병관리본부	예방접종지원사업관리지침정보(의료기관용)	LINK
395		예방접종업무위탁의료기관현황정보	LINK
396		연도별·월별연간감염병발생통계	LINK
397		성별·연령별연간감염병발생통계	LINK
398		동물실험윤리위원회위원및동물실험계획DB	다운로드
399		항공검역정보	
400		법정감염병역학조사서정보	다운로드
401		임상연구DB	LINK
402		LIMS/품질문서	다운로드
403		질병별지역별감염병발생통계	LINK
404		질병별성별·연령별감염병발생통계	LINK
405		지역별월별역학적연관성별감염병발생통계	LINK
406		직업별·질병별감염병발생통계	LINK
407		지역별연간감염병발생통계	LINK
408		질병별연간감염병발생통계	OPEN API
409		지역별전년대비주별·월별감염병발생통계	LINK
410		지역별일별·주별·월별감염병발생통계	LINK
411		지역별성별·연령별감염병발생통계	LINK
412		질병별전년대비주별·월별감염병발생통계	LINK
413		질병별일별·주별·월별감염병발생통계	LINK
414		해외감염병발생정보DB	LINK
415	대한적십자사	대한적십자사혈액관리본부홈페이지공개·개방정보	다운로드
416		대한적십자사보건안전정보	다운로드
417		대한적십자사혈액통계정보	다운로드
418	안전행정부	국가기초구역서비스	LINK
419	한국보훈복 지의료공단	위탁병원현황DB	다운로드
420		보훈병원진료정보	다운로드
421	한국산업 안전보건공단	[안전보건]화학물질및물리적인자의노출기준정보	다운로드

354 보건복지 빅데이터 효율적 관리방안 연구

번호	기관	데이터명	서비스유형
422	한국환경공단	석면피해인정통계	다운로드 LINK
423	한국건강증진재단	국민구강건강실태조사	LINK
424		기타자료정보DB	LINK
425		발간자료정보	LINK
426		보건교육자료	LINK
427		연구보고서DB	LINK
428		건강정보DB	LINK
429		각종생활정보DB	LINK
430		건강매거진	LINK
431	한국과학 기술정보연구원	디지털코리언	
432	한국국제 보건의료재단	의약품/의료기기후원및지원목록	다운로드
433		외국인근로자이동클리닉지원목록	다운로드
434		국제보건의료관련연구용역결과	다운로드
435		외국인근로자보건교육자료	다운로드
436	한국보건복지 정보개발원	보건의료정보용어표준	다운로드
437	한국보건 산업진흥원	병원경영분석통계	LINK
438		보건산업통계	
439		고령친화우수제품지정현황	다운로드
440	한국보건의료인 국가시험원	보건의료인국가시험면허교부상황DB	다운로드
441		보건의료인국가시험전환성적발급병원목록DB	다운로드
442		보건의료인국가시험응시자학교목록DB	다운로드
443		보건의료인국가시험요양보호사교육원목록DB	다운로드
444		보건의료인국가시험시험정보DB	다운로드
445		보건의료인국가시험시험장정보DB	다운로드
446		보건의료인국가시험간호조무사학원목록DB	다운로드
447	없음	전국 약국 현황	그리드 다운로드 OPEN API
448		전국 병원현황	
449		전국 동물병원 현황	

3) 식품의약품안전

〈부표 1-3〉 식품의약품안전 공공데이터

번호	기관	데이터명	서비스유형
1	강원도	착한가격업소	다운로드
2	강원도삼척시	삼척시 착한가격 업소 현황	다운로드
3	강원도원주시	원주시 기타식품판매업체 정보	다운로드
4		원주시 단란주점 정보	다운로드
5		원주시 담배소매인 지정 정보	다운로드
6		원주시 식품제조가공업체 정보	다운로드
7		원주시 유흥주점 정보	다운로드
8		원주시 일반음식점 정보	다운로드
9		원주시 제과점 정보	다운로드
10		원주시 즉석판매제조가공업체 정보	다운로드
11	경기도과천시	과천시 공중위생서비스평가 우수업소 정보	다운로드
12		과천시 안전상비의약품 판매업소 정보	LINK
13		과천시 약국 정보	다운로드
14		과천시 주방공개CCTV설치업소 정보	다운로드
15	경기도군포시	군포시 일반음식점 현황	다운로드
16		군포시 휴게음식점 현황	다운로드
17	경기도남양주시	남양주 모범업소 현황	다운로드
18		남양주 식품제조가공업체 현황	다운로드
19		남양주 약취중점관리업소정보	다운로드
20		남양주 위생위반 행정처분업소정보	다운로드
21	경기도동두천시	모범음식점 현황	다운로드
22	경기도부천시	부천시 각종 위생허가 건수	다운로드
23		부천시 건강기능식품일반판매업 현황	다운로드
24		부천시 식품제조가공업 현황	다운로드 LINK
25		부천시 즉석판매제조가공업 현황	다운로드

번호	기관	데이터명	서비스유형
26		식품진흥기금 용자에 관한 정보	다운로드
27		오정구 공중위생업소 현황	다운로드 LINK
28		오정구 모범음식점 지정업소 현황	다운로드
29		오정구 식품위생업소 현황	LINK
30	경기도성남시	식품관련업 현황	다운로드
31		식품접객업 현황	다운로드
32		식품제조_식품소분판매업현황	다운로드
33	경기도성남시	유홍단란주점현황	다운로드
34	경기도연천군	미용업소현황	다운로드
35	경기도연천군	숙박업소현황	다운로드
36	경기도용인시	식품위생 위반업소 정보	LINK
37		식품제조가공업 정보	다운로드
38		위해식품공개 정보	LINK
39		처인구 안전상비의약품 판매업소 정보	다운로드
40	경기도파주시	모범음식점 현황	LINK
41		착한가격업소 현황	LINK
42	경기도화성시	모범음식점현황	다운로드 개방예정
43		화성시 식육포장처리업 영업현황	다운로드
44	경기도교육청	영양교육상담	LINK
45	경상남도남해군	의약업소정보데이터	LINK
46	경상남도의령군	모범음식점정보 DB	LINK
47	경상남도함안군	집단급식소 및 위탁급식 현황	LINK
48	경상북도	으뜸음식점 DB	다운로드
49	경상북도봉화군	식품위생법 위반업소	LINK
50	대구광역시남구	어린이기호식품우수판매업소현황(대구남구)	다운로드
51	대구광역시달서구	안전상비의약품 판매업 현황	다운로드
52	대구광역시서구	공중위생평가DB	LINK

번호	기관	데이터명	서비스유형
53		식품산업DB	LINK
54	대구광역시수성구	수성구 모범음식점 현황	다운로드
55		수성구 식품관련업 현황	다운로드
56		수성구 식품접객업소 현황	다운로드
57		수성구 유흥 단란주점 현황	다운로드
58	대구광역시중구	보건(의약업소)	다운로드
59	부산광역시	유해물질 수거검사 현황	다운로드
60		집단급식소 현황	다운로드
61	부산광역시기장군	건강기능식품일반판매업현황	LINK
62		공중위생서비스평가 등급별 업소 현황	LINK
63		집단급식소 현황	LINK
64	부산광역시남구	모범음식업소현황	LINK
65		식품위생업소 대장	LINK
66		식품위생업소 행정처분	LINK
67	부산광역시동래구	일반음식점	다운로드
68	부산광역시수영구	건강기능식품판매업 현황	다운로드
69		의약품도매상 현황	다운로드
70	부산광역시 해운대구	동물약국 현황	다운로드
71		위탁급식 현황	다운로드
72		유흥단란주점 현황	다운로드
73		제과점 현황	다운로드
74		집단 급식소 현황	다운로드
75	서울특별시	식품안전정보	LINK
76	서울특별시강동구	강동구 모범음식점(강동구 홈페이지 문화포털 내)	다운로드
77	서울특별시구로구	구로구 가축사육업체 현황	LINK
78		구로구 공중위생업소 현황	LINK
79		구로구 동물약국 현황	LINK
80		구로구 동물약품 도매업체 현황	LINK

번호	기관	데이터명	서비스유형
81		구로구 모범음식점 지정현황	LINK
82		구로구 새울행정(위생)	LINK
83		구로구 식품위생업소 소재지별 운영현황(법정동별)	LINK
84		구로구 장난감 정보	LINK
85		구로구 지혜의 등대 도서관정보	LINK
86		구로구 축산물 운반차량 등록현황	LINK
87		구로구 축산물판매업체 현황	LINK
88	서울특별시서대문구	새울행정(서대문구-위생)	LINK
89	서울특별시서초구	모범음식점	다운로드
90	세종특별자치시	치매치료약제	다운로드
91		치매치료약제목록	다운로드
92	울산광역시	착한가격업소 현황	LINK
93	울산광역시동구	착한가격업소	LINK
94	인천광역시남구	모범음식점 현황	다운로드
95	인천광역시동구	건강식품판매업소정보	다운로드 LINK
96		일반음식점정보	다운로드 LINK
97	인천광역시보건환경연구원	농산물 중 잔류농약 검사 결과	LINK
98		유통식품 검사결과	LINK
99	인천광역시연수구	연수구 맛고을길 업소 현황	다운로드
100		연수구 모범음식점	다운로드
101		연수구 유흥주점	다운로드
102		연수구 음식문화시범거리 지정업소	다운로드
103		연수구 지정 맛있는 집 현황	다운로드
104		인천시지정 연수구 맛있는 집	다운로드
105	인천광역시중구	식품위생업소 현황	LINK
106	전라남도	농수축산물 지리적 표시현황	다운로드
107		모범음식점 정보	다운로드
108		식품위생업소현황	다운로드

번호	기관	데이터명	서비스유형
109	전라남도보성군	모범음식점지정현황	다운로드
110		보성녹차군수품질인증현황	다운로드
111		숙박시설 DB	다운로드
112	전라남도순천시	단란주점 현황	LINK
113		부정·불량식품 신고포상금 지급 현황	LINK
114		소분업 현황	LINK
115		식품제조가공업 현황	LINK
116		유흥주점 현황	LINK
117		일반음식점 현황	LINK
118		즉석판매제조가공업 현황	LINK
119		집단급식소 현황	LINK
120		휴게음식점 현황	LINK
121	전라남도여수시	여수시 공중위생업소 현황	다운로드
122	전라남도영암군	식품 위생업소 현황	다운로드
123	전라남도함평군	모범업소 현황	다운로드
124	전라남도화순군	착한가격업소	LINK
125	전라북도남원시	휴게음식점 현황	LINK
126		건강기능식품 판매유통업 현황	LINK
127		세탁업소 현황	LINK
128		식품 제조가공업 현황	LINK
129		이·미용업소 현황	LINK
130		제과점 현황	LINK
131		즉석판매제조업 현황	LINK
132	전라북도정읍시	건강기능식품일반판매	LINK
133		다소비의약품 판매가격조사	LINK
134		식품소분업소	LINK
135		식품운반업소	LINK
136		식품제조가공업	LINK

번호	기관	데이터명	서비스유형
137		유통판매업소	LINK
138		일반음식점	LINK
139		즉석판매제조가공업소	LINK
140	충청남도서산시	서산시 모범음식점	그리도 다운로드 OPEN API
141		서산시 휴게음식점현황	
142	충청남도천안시	모범음식점	다운로드
143		착한가격업소	LINK
144	충청남도태안군	착한가격업소	다운로드
145	충청남도홍성군	홍성군 음식점 현황	다운로드
146	충청북도단양군	숙박업DB	다운로드
147	충청북도단양군	음식점 DB	다운로드
148	충청북도영동군	영동군 음식점	LINK
149	충청북도충주시	식품관련 영업소 현황	다운로드
150	식품의약품안전처	HACCP 지정현황	LINK
151		Paragraph IV 대상 의약품(특허)	LINK
152		건강기능식품	OPEN API
153		검사 부적합 식품	OPEN API
154		국내소송 의약품(특허)	LINK
155		기구 등의 살균소독제 정보	LINK
156		기구 및 용기포장 정보	LINK
157		마약류 및 원료물질 정보	LINK
158		마약류 종합정보	OPEN API
159		수입식품 통계정보	LINK
160		수입식품 허가정보	LINK
161		식·의약 통계 정보	LINK
162		식중독 통계	LINK
163		식중독예방 대국민 홍보자료	LINK
164		식품 및 식품첨가물 생산실적 정보	LINK

번호	기관	데이터명	서비스유형
165		식품 영양성분 데이터베이스	LINK
166		식품 행정처분	OPEN API
167		식품 허위·과대광고 정보	OPEN API
168		식품기준규격	LINK
169		식품첨가물 정보	LINK
170		실험동물 제도 및 실험동물 시설 정보	LINK
171		어린이급식관리	LINK
172		유전자재조합식품(GMO) 정보	LINK
173		의료기기 GMP 지정현황	LINK
174		의료기기 광고 사전심의	LINK
175		의료기기 재평가	LINK
176		의료기기 제조(수입)업허가 정보	OPEN API
177		의료기기 품목허가 정보	OPEN API
178		의료기기 회수·판매중지	LINK
179		의약품 DB	LINK
180		의약품 기업 정보	LINK
181		의약품 업소 허가정보	OPEN API
182		의약품 임상시험	LINK
183		의약품 재심사	LINK
184		의약품 재평가	LINK
185		의약품 제품 허가정보	OPEN API
186		의약품 특허정보(기타)	LINK
187		의약품 특허정보(오렌지북)	LINK
188		의약품 특허정보(제품)	LINK
189		의약품 행정처분	LINK
190		의약품 허가사항 정보	LINK
191		의약품 회수·판매중지	OPEN API
192		인체조직은행 정보	LINK

번호	기관	데이터명	서비스유형
193		임상시험 실시기관 정보	LINK
194		추적관리대상 의료기기 정보	LINK
195		화장품 제조(판매)업 정보	LINK
196		국가출하승인	LINK
197		기능성화장품 보고 품목 정보	LINK
198		기능성화장품 심사 품목 정보	LINK
199		독성유전체 통합정보	LINK
200		독성정보	OPEN API
201		약물유전정보	LINK
202		임상시험 지식 DB	LINK
203		잔류화학물질정보	LINK
204	축산물품질평가원	쇠고기 이력정보	OPEN API
205	한국보건산업진흥원	HACCP지정업소 현황	다운로드
206	없음	전국 으뜸 음식점 현황	그리도 다운로드 OPEN API

나. 사회복지분야

1) 노인·청소년

〈부표 1-4〉 노인·청소년 공공데이터

데이터 명	제공 Data 변수 및 자료설명	제공기관
노인장기 요양보험 (통계)	상반기 시군구별 노인장기요양 신청, 인정현황, 판정등급별 급여실적, 노인장기요양기관 수 및 인력현황	국민건강 보험공단
노인장기 요양보험 (연보)	-	국민건강 보험공단

2) 사회복지 보건의료

〈부표 1-5〉 사회복지 보건의료 공공데이터

데이터 명	제공 Data 변수 및 자료설명	제공기관
진료DB	보훈복지의료공단의 2008~2012년도 까지의 진료과 별 진료DB	한국보훈 복지의료공단
환자 진료내역	연도별, 병원별, 산재병원 환자진료 실적	고용노동부

3) 사회복지 일반

〈부표 1-6〉 사회복지 일반 공공데이터

데이터 명	제공 Data 변수 및 자료설명	제공기관
건강검진 통계연보	국민건강검진 운영에 따른 검진종별 결과를 세부적인 통계로 작성하여 건강검진 및 건강증진 정책수립을 위한 기초자료 지원	국민건강 보험공단
건강보험DB	요양급여 실적, 의료기관별 급여적정성평가현황(약제 평가 항목별 지표)	건강보험 심사평가원
건강보험 주요통계	반기별 건강보험 주요 통계를 제공하는 것으로 보건 의료정책 수립을 위한 기초자료 지원	국민건강 보험공단
건강보험 통계연보	건강보험에 대한 기본통계로 의료보장적용인구, 건강 보험재정현황(건강보험료, 정부지원금), 시군구별 급여 실적, 질병통계	국민건강보험공단
희귀돌연 변이환자	한국인이 발굴된 희귀유전질환 유전자의 돌연변이 환자 정보 제공	보건복지부
요양기관 종합정보	의료기관명, 기호, 대표자, 성별, 주민번호, 자격, 근무, 장비종류, 장비코드	건강보험 심사평가원
인플루엔자 DB	주변 인플루엔자 바이러스 검출결과	보건복지부
장기 이식정보	국내에서 발생한 장기기증 및 이식에 관한 정보	보건복지부

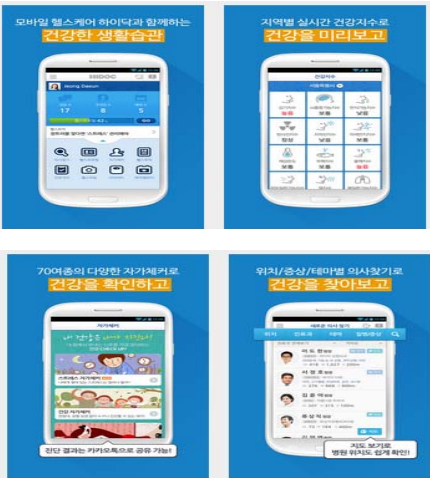
데이터 명	제공 Data 변수 및 자료설명	제공기관
조혈모세포정보	국내에서 발생한 조혈모세포(골수, 제대혈) 기증 및 이식에 관한 정보	보건복지부
지역별 이용통계	지역별 의료이용 특성을 비교분석을 통한 지역 건강 복지 행정 지원	국민건강보험공단
진료경향 모니터링DB	종별, 지역별, 청구건수, 금액, 환자수	건강보험심사평가원
질병통계	상병소분류 통계, 상병세분류 통계	건강보험심사평가원
통계자료DB	건강보험통계연보 현황, 진료비 통계표(건강보험/의료급여), 평가관련통계(약제급여 적정성 평가결과)	건강보험심사평가원
한국인 척추데이터	퇴행성 척추 질환을 갖는 한국인 척추(시신, 환자)에 대한 의료영상, 3차원 영상, 척추 치수 및 물성정보	한국과학기술정보연구원
희귀돌연변이 DB	한국인에서 발굴된 희귀유전질환, 유전자의 돌연변이 정보제공	보건복지부
희귀난치성 질환 DB	916개 희귀난치성질환에 대한 개요, 증상, 원인, 진단, 치료 정보 제공	보건복지부
희귀난치성 질환의료비 지급 DB	희귀난치성질환자의 의료비지원사업을 통해 지원된 의료비지급현황	보건복지부
행위통계 DB	수가, 보험등재약, 치료재료 등의 진료코드에 따른 실적회수 및 금액	건강보험심사평가원

2. 보건복지 공공데이터 활용현황

가. 국내

1) 건강관리

□ 하이닥

서비스 명	Hidoc(하이닥)	제작사	(주)엠서클 모바일팀
서비스 대상	스마트폰을 이용하는 모든사람	등록일	2014.01.06
		개발유형	모바일앱
핵심 서비스	• 내가 머무는 지역의 방사선 지수, 감기지수, 뇌졸중 지수, 미세먼지 지수 등 18종의 건강지수 • 스트레스 지수, 질병 증상의 원인, 아이 성장 발달 상태 등을 체크할 수 있는 셀프진단 • 가족 모두가 다니는 병원 의사를 어플리케이션에 등록하여 관리하는 기능 • 위치 기반의 거리, 증상별 새로운 의사찾기 서비스 • 관심정보에 따른 개인 맞춤형 건강정보 제공		
주요기능화면			
활용API	기타, 생활기상지수, 보건기상지수, 전국 명절 비상 진료기관 정보 조회 서비스		

□ 유투책

서비스 명	유투책	제작사	(주)유투바이오
서비스 대상	의료기관(검진센터,의원,종합병원)	등록일	2014.01.03
		개발유형	기타
핵심 서비스	• 환자정보 입력 시 내부 데이터 조회 후 없는 주소에 대하여 Open API 호출		
활용API	미래창조과학부 우정사업본부, 도로명주소 조회 서비스		

□ 중앙대학교병원

서비스 명	중앙대학교병원	제작사	
서비스 대상	-	등록일	2013.11.20
		개발유형	웹사이트
핵심 서비스			
활용API	미래창조과학부 우정사업본부, 도로명주소 조회 서비스		

2) 공공보건 의료분야

□ 착한병원 착한의사

서비스 명	착한병원 착한의사	제작사	조성호
서비스 대상	스마트폰을 이용하는 모든 국민	등록일	2014.07.01
		개발유형	모바일앱
핵심서비스	• 내주변 진료과목별 병원찾기 • 병원명으로 병원찾기 • 긴급상황에서 응급실 등이 있는 가장 가까운 병원 찾기 • 관절, 수지접합, 척추 등의 전문병원 찾기 • 내 주변약국 찾기 • 건강검진 정보 확인하기		
활용API	전국 병·의원 찾기 서비스, 전구 응급의료정보조회 서비스		

□ 아이체크

서비스 명	아이체크	제작사	네오비엔에스
서비스 대상	스마트폰을 이용하는 모든 국민	등록일	2014.06.19
		개발유형	모바일앱
핵심서비스	• 우리 아이와 같은 연령대의 아이들의 발육 상태 비교 • 주기적인 성장 기록 입력으로 아이의 성장 그래프를 만들어 자녀에게 자신의 성장과정을 남겨 줄 수 있음 • 자녀의 예방접종 시기를 알 수 있고 접종일정을 기록		
활용API	전국 병·의원 찾기 서비스		

□ 병원 약국 찾기

서비스 명	병원 약국 찾기	제작사	이상협
서비스 대상	스마트폰을 이용하는 모든 국민	등록일	2014.03.06
		개발유형	모바일앱
핵심서비스	• 스마트 폰에서 전국 병원, 약국, 응급의료기관 정보를 한눈에 확인 할 수 있도록 제공		
활용API	전국 병·의원 찾기 서비스, 전국 약국 조회 서비스, 전국 병·의원 찾기 서비스		

□ 강동경희대학교병원 홈페이지

서비스 명	강동경희대학교병원	제작사	-
서비스 대상	-	등록일	2013.12.18
		개발유형	웹사이트
활용API	미래창조과학부 우정사업본부, 도로명주소 조회 서비스		

□ 다음지도

서비스 명	다음지도	제작사	-
서비스 대상	홈페이지를 이용하는 모든 사람	등록일	2013.12.18
		개발유형	웹사이트
핵심서비스	• 병·의원의 주소, 명칭, 좌표, 전화번호 등을 연동반야 서비스 제공 • 일반적인 길 찾기 기능 서비스 제공		
활용API	국립중앙의료원, 전국 병·의원 찾기 서비스		

□ 수의사 처방관리 시스템

서비스 명	수의사 처방관리 시스템	제작사	(주)대한수의사회
서비스 대상	홈페이지를 이용하는 모든 사람	등록일	2013.08.01
		개발유형	웹사이트
활용API	미래창조과학부 우정사업본부, 도로명주소 조회 서비스		

□ 질병관리본부 예방접종도우미

서비스 명	질병관리본부 예방접종도우미	제작사	질병관리본부
서비스 대상	스마트폰을 사용하는 모든 국민	등록일	2013.01.06
		개발유형	모바일앱
핵심 서비스	• 국가 예방접종 기록 및 DB에 등록된 예방접종 기록 조회(모바일 아기수첩 기능) • 감염병 의학정보 치료 및 예방법 안내 등		
주요기능화면			
활용API	-		

3) 보건산업

□ 메리라떼

서비스 명	메리라떼	제작사	에이디벤처스
서비스 대상	스마트폰을 이용하는 모든 국민	등록일	2014.03.11
		개발유형	모바일앱
핵심서비스	• 할 수 있음 (영업시간, 주소, 전화번호, 카톡, 시술사진, 약력 등) • 6만 8천개의 병원 DB를 활용하여 환자 맞춤형 병원정보를 제공하는 어플리케이션		
활용API	국립중앙의료원, 전국 병·의원 찾기 서비스		

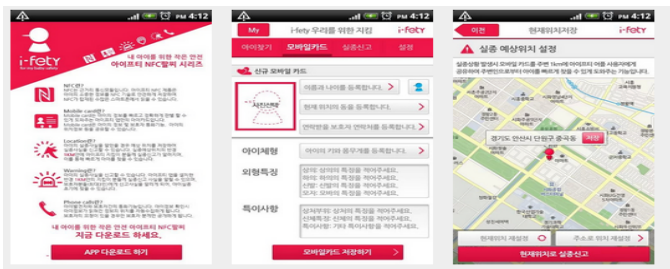
4) 장애인복지

□ 서울장애인 자립 생활 정보

서비스 명	서울시 장애인 자립정보 앱	제작사	able forum
서비스 대상		개발유형	모바일앱
핵심서비스	• 서울에서 거주하는 장애인이 생활을 자립하는데 필요한 정보들을 제공하는 어플리케이션		
주요기능화면			
활용API	기타, 장애인 주차장, 화장실, 경사로, 엘리베이터, 장애인 관련 기관 연락처		

4) 아동·청소년 복지

□ 실종아동 예방 NFC 솔루션 아이프티

서비스 명	실종아동 예방 NFC 솔루션 아이프티	제작사	바오미르
서비스 대상		개발유형	모바일앱
핵심서비스	• 아이 미아방지 시스템		
주요기능화면			
활용API	초록우산재단 실종아동정보, 경찰서 위치정보, 안행부 주민정보		

□ 나누다

서비스 명	나누다	제작사	한성택
서비스 대상		개발유형	모바일앱
핵심서비스	• 봉사활동 종합정보 서비스		
주요기능화면			
활용API	안전행정부 코드조회서비스, 봉사배움정보서비스, 봉사참여정보서비스		

□ 월드비전

서비스 명	월드비전	제작사	김명현
서비스 대상	홈페이지를 이용하는 모든사람	개발유형	웹사이트
핵심서비스	• 국제구호배발 NGO 단체		
활용API	미래창조과학부 우정사업본부, 도로명주소 조회 서비스		

□ 농어촌 청소년 장학재단

서비스 명	농어촌 청소년 장학재단	제작사	농어촌청소년육성재단
서비스 대상	홈페이지를 이용하는 모든사람	개발유형	웹사이트
핵심서비스	• 농어촌 출신 청소년 대상 장학금 지원 재단		
활용API	미래창조과학부 우정사업본부, 도로명주소 조회 서비스		

□ 오산시청

서비스 명	오산시청	제작사	경기도 오산시
서비스 대상	스마트폰을 이용하는 모든 국민	개발유형	모바일앱
핵심서비스	• 오산시 관련정보 및 자원봉사 정보 서비스 제공		
활용API	안전행정부 코드조치서비스		

□ 스마트 창원/창원시 모바일 웹

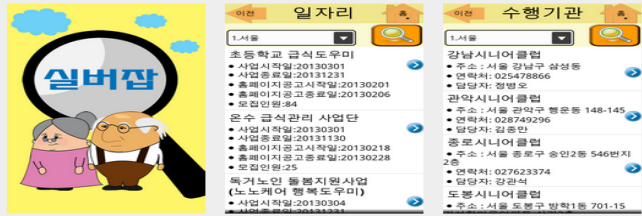
서비스 명	스마트 창원	제작사	경상남도 창원시
서비스 대상	홈페이지를 이용하는 모든사람	개발유형	모바일웹, 웹사이트
핵심서비스	• 창원시 하이브리드 어플리케이션, 주요시설 및 실생활과 밀접한 분야에 대한 위치기반서비스		
활용API	안전행정부 코드조치서비스		

□ 복지로 모바일 웹페이지 및 앱

서비스 명	복지로 모바일 웹페이지 및 앱	제작사	한국보건복지정보개발원
서비스 대상	홈페이지를 이용하는 모든 사람	개발유형	웹사이트
핵심서비스	• 복지관련 학술정보 서비스를 제공하는 대국민 서비스		
활용API	안전행정부 기부관련단체정보서비스, 안전행정부 코드조회 서비스		

5) 노인복지

□ 실버잡

서비스 명	실버잡	제작사	서용훈
서비스 대상		개발유형	모바일앱
핵심서비스	• 노인 일자리 정보 제공, 전화, 지도 등 편리한 기능 제공		
주요 기능 화면			
활용API	한국노인인력개발원 노인일자리 모집 공고 서비스		

□ SK텔레콤 행복창업지원센터

서비스 명	SK 테렐콤 행복창업지원센터	제작사	양타규
서비스 대상		개발유형	웹사이트
핵심서비스			
활용API	미래창조과학부 우정사업본부, 도로명주소 조회 서비스		

6) 화장품 위해정보

□ 화해(화장품을 해석하다)

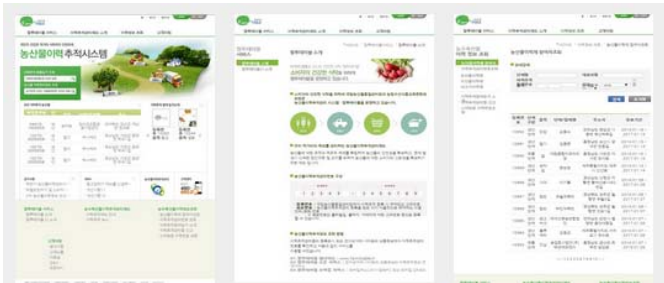
서비스 명	화해	제작사	화해
서비스 대상		개발유형	모바일웹
핵심서비스	• 화장품 성분들을 쉽고 빠르게 검색		
주요 기능 화면			
활용API	식품의약품안전처 화장품 원료 및 성분 데이터		

□ 피부랑

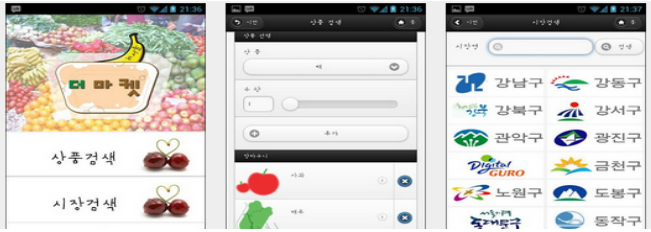
서비스 명	피부랑	제작사	송태관
서비스 대상		개발유형	웹사이트
핵심서비스	• 피부 관리 정보 제공		
주요 기능 화면			
활용API	미래창조과학부 우정사업본부, 지번주소조회 서비스		

7) 유통정보

□ 팜투데이블

서비스 명	팜투데이블	제작사	농림수산식품교육문화정보원
서비스 대상		개발유형	웹사이트
핵심서비스	• 먹거리 족보 관리를 통한 농산물에 대한 소비자 신뢰성을 확보하기 위한 사이트		
주요 기능 화면			
활용API	농림수산식품교육문화정보원, 농산물 이력 관리 정보 조회 서비스		

□ 더 마켓 인 서울

서비스 명	팜투데이블	제작사	박병준
서비스 대상		개발유형	모바일앱
핵심서비스	• 서울시내 존재하는 재래시장, 슈퍼마켓 등에서 판매하는 16가지 품목들에 대하여 가장 저렴하게 판매하는 시장을 찾아 주는 어플리케이션		
주요 기능 화면			
활용API	농림수산식품교육문화정보원, 농산물 이력 관리 정보 조회 서비스		

나. 국외

1) 건강관리

□ 23andMe

서비스 명	강동경희대학교병원	제작사	23andMe Inc
서비스 대상	-	개발유형	웹사이트
핵심서비스	- 유전자 정보를 추적·활용·분석하여 질병성향을 알고 예방하도록 함 - DNA를 공유하는 사람들의 특성을 차아 가게도 서비스를 이용하여 데이터 구축 - 유전자 정보를 분석하고 고객 질환의 위험 등 245개 항목에 대한 판정		
활용API	DB에 있는 유전자 정보는 API로 제공됨		

□ UK Dentists

서비스 명	UK Dentists	제작사	Elbatrop Ltd.
서비스 대상		개발유형	웹사이트
핵심서비스	- 영국의 보건 및 사회복지 센터가 제공하는 데이터를 기반으로 영국에서 등록된 모든 NHS치과의 99.4%를 커피하고 있는 치과의사 정보를 제공하는 어플리케이션 - 현재 지점에서의 검색, 지명, 우편번호를 통한 검색이 가능 함 - 위치를 지도상에 표시하는 서비스 제공		
활용API	data.gov.uk에 등록된 Open data 사용		

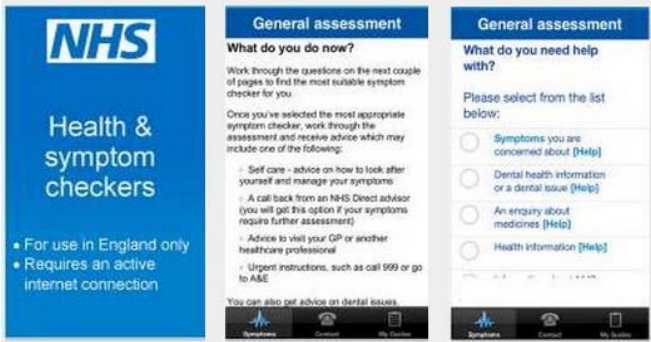
□ Nothing haminformation prescriptions

서비스 명	Nothing haminformation prescriptions	제작사	NHS Nottinghamshire Contry
서비스 대상		개발유형	웹사이트
핵심서비스	- 장기간 치료를 필요로 하는 병에 관하여 개인화된 정보를 제공 - 온라인 상의 정보를 열람/다운로드 할 시 주치의로부터 연락받음 - 처방에 대한 서비스를 제공 함 - NHS 공인 정보 제공 함		
활용API	-		

□ Obesity Data

서비스 명	Obesity Data	제작사	National Obesity Observatory
서비스 대상		개발유형	웹사이트
핵심서비스	- 비만관련 데이터 감시 및 분석정보, 비만 해결에 관한 지침 정보를 제공 - 비만의 결정적 요인에 대한 분석		
활용API	잉글랜드 건강 조사(HSE) 사용		

□ NHS Dircetory

서비스 명	NHS Dircetory	제작사	NHS Dircet
서비스 대상		개발유형	모바일앱
핵심서비스	- 몸 상태가 나쁠 때 증상을 자가체크 할 수 있는 어플리케이션 - 지금 현재의 몸 상태, 대처방안에 대한 조언을 얻을 수 있음 - 증상 확인 후 건강 상담과 자가치료 후 전화 문의 또는 기타 서비스를 받을 수 있음		
주요기능화면			
활용API	-		

2) 보건산업

□ Reference NHS

서비스 명	Reference NHS	제작사	Michael Bevan
서비스 대상		개발유형	모바일앱
핵심서비스	• 올인원 NHS정보 어플리케이션 • 임상적, 임상코드, 정보 분석가, NHS 매니저 및 재무 스텝용 • ICD-10 진단 코드 등 제공		
주요기능화면			
활용API	-		

□ Patient Opinion

서비스 명	Patient Opinion	제작사	Patient Opinion
서비스 대상		개발유형	웹사이트
핵심서비스	• 환자들을 대상으로 헬스케어 서비스에 대한 평가를 받음 • 독립적인 비영리 피드백 플랫폼		
활용API	-		

□ Nursing Homes Directory

서비스 명	Nursing Homes Directory	제작사	Chic Computer Consultants Ltd.
서비스 대상		개발유형	웹사이트
핵심서비스	• 보살핌, 간호 서비스, 케어 에이전시에 대한 정보를 제공		
활용API	-		

□ NRSA Action UK

서비스 명	NRSA Action UK	제작사	MRSA Action UK
서비스 대상		개발유형	웹사이트
핵심서비스	• 감염 정보 및 감염 현황 정보 제공		
활용API	-		

□ HealthyApps-MRSA & CD.iff

서비스 명	HealthyApps-MRSA & CD.iff	제작사	HealthyApps Ltd.
서비스 대상		개발유형	모바일앱
핵심서비스	• 세균 감염자 및 병원에 대한 정보를 제공		
주요기능화면			
활용API	-		

□ Find Pharmacies

서비스 명	Find Pharmacies	제작사	Elbatrop Ltd.
서비스 대상		개발유형	모바일앱
핵심서비스	• 영국의 보건 및 사회복지 센터가 제공하는 데이터를 기반으로 영국에서 등록된 모든 NHS치과의 99.4%를 커버하고 있으며 주변의 약국정보를 검색		
주요기능화면			
활용API	data.gov.uk에 등록된 Open data 사용		

□ Find GPS

서비스 명	Find GPS	제작사	Elbatrop Ltd.
서비스 대상		개발유형	모바일앱
핵심서비스	- 영국의 보건 및 사회복지 센터가 제공하는 데이터를 기반으로 영국에서 등록된 모든 NHS치과의 97.9%를 커버하고 있으며 근처의 병원정보를 제공하고 있음 - 지도를 제공하여 이용자의 이해도를 높임		
주요기능화면			
활용API	data.gov.uk에 등록된 Open data 사용		

□ PrescriptionSearch

서비스 명	PrescriptionSearch	제작사	Leo
서비스 대상		개발유형	웹사이트
핵심서비스	- 영국에서 발행된 처방전 정보를 제공하며 GSP에 의해 발행한 처방전을 지도상에 보여줌 2011년에 영국의 일반의사가 발행한 모든 처방전 정보를 볼 수 있으며 - 우편번호에 따라서 그룹화가 되어 있으며, 발행된 처방전 현황을 알 수 있음		
활용API	data.gov.uk에 등록된 Open data 사용		

3) 가족·다문화

□ StreetBank

서비스 명	SK 테렐콤 행복창업지원센터	제작사	Sam, Ryan, Nic and Alice
서비스 대상		개발유형	웹사이트
핵심서비스	• 근처 이웃들을 사귀거나 좋은 관계를 맺도록 하는 목적에서 물건을 공유하거나 대여가능 • 우편번호를 통한 가입 서비스, 국가별로 서비스 제공		

□ Scottish Index of Multiple Deprivation

서비스 명	Scottish Index of Multiple Deprivation	제작사	The Scottish Government
서비스 대상		개발유형	웹사이트
핵심서비스	• 스코틀랜드 빈곤지역에 살고 있는 사람들의 환경과 결과에 대한 이해를 개선하는 것을 돕기 위한 정보를 제공 • 스코틀랜드 빈곤 상황을 복수의 지표로 지도에 표시됨 • 연도별 데이터 비교 가능 • 분석과 배경 지식, 방법론, 데이터 소스에 관한 정보를 제공		

□ OpenCharities

서비스 명	Open Charities	제작사	Open Charities
서비스 대상		개발유형	웹사이트
핵심서비스	• 영국과 웨일스의 등록된 자선기부에 관한 정보 제공		

□ Compare care Homes

서비스 명	Compare care Homes	제작사	Peolsford Ltd.
서비스 대상		개발유형	웹사이트
핵심서비스	• 윤리적인 서비스 제공을 목표로 하는 케어홈 서비스 • 이용자가 자신의 경험에 대한 정보를 제출하고 각 요양소의 서비스를 평가하여 데이터 누적		

□ City of SanFfancisco REcreation and Parks official Mobile

서비스 명	City of SanFfancisco REcreation a	제작사	appallicious LLC
서비스 대상		개발유형	모바일앱
핵심서비스	• 도시 내 각종 편의시설에 대한 정보를 제공		
주요 기능 화면			

□ Carehome.co.kr

서비스 명	carehome.co.kr	제작사	Tomorrow's Guides Ltd
서비스 대상		개발유형	모바일앱
핵심서비스	• 장기간에 걸친 케어정보를 제공, 케어를 통한 필요한 제품을 사이트를 통해 판매		
주요 기능 화면			

4) 장애인복지

□ care Home Map

서비스 명	Carehome map	제작사	carehomeamp
서비스 대상		개발유형	웹사이트
핵심서비스	• 영국내 보살핌 서비스에 대한 정보		
활용API	케어품질위원회(Care quality commission)가 가진 간병시설 데이터 이용		

5) 노인복지

□ Beat care Home

서비스 명	Beatcare Home	제작사	utilities Knowledge management
서비스 대상		개발유형	웹사이트
핵심서비스	• 18,500가구의 품질 지수, 지역, 가격, 노년층의 서비스 요구, 당뇨병교육, 정신건강 검색 • 케어 희망 집을 모니터링하고, CQC 검사 보고서가 게시될 때 이메일을 통해 제공		

6) 식품 안전

□ Scores on the Doors

서비스 명	Scores on the Doors	제작사	Transparency Data Ltd
서비스 대상		개발유형	모바일앱
핵심서비스	• 음식점이나 식료품점 등 식품을 취급하고 있는 가게의 안전검사 등급 정보 • 식품을 안전하게 하는 등의 여러 가지 정보를 제공 함 • 영국 전역 439,544 곳에 이르는 정보를 제공 함		
활용API	식품 표준 기관에 의해 공급되고 있는 지방 자치 단체에 의하여 정보 제공 받음		

□ Ratemyplace

서비스 명	Ratemyplace	제작사	스태퍼드셔주 지방당국
서비스 대상		개발유형	웹사이트
핵심서비스	• 식품안전에 대한 검사 정보 제공, 스지도상에 표시를 통하여 검사 등급에 대한 이해를 높임		
활용API	스텐퍼드셔주 지방당국		

〈부록 2〉 빅데이터 기술수용모형 조사표

1. 조직 혁신성에 관한 질문

1. 조직혁신성에 관한 질문	매우 그렇지 않음-----매우 그러함						
	①	②	③	④	⑤	⑥	⑦
1-1. 나는 혁신적 아이디어를 활용하기 위해 노력하는 편이다							
1-2. 나는 업무를 해결하기 위해 독창적 아이디어를 고민하는 편이다							
1-3. 나는 혁신적인 아이디어에 대한 지원을 획득하기위해 노력하는 편이다							
1-4. 내가 속한 조직은 구성원의 가치와 믿음을 중요하게 생각한다							
1-5. 내가 속한 조직은 개인의 성과를 자랑스럽게 생각한다							
1-6. 나의 리더는 혁신적 아이디어를 수용하는 편이다							
1-7 나의 리더는 업무에 있어 구성원에게 충분한 권한을 부여한다							
1-8 나의 리더는 구성원의 학습을 지원하는 편이다							

2. 빅데이터 기술과 관련한 주관적 규범에 대한 질문

2. 빅데이터 기술과 관련한 주관적 규범에 대한 질문	매우 그렇지 않음-----매우 그러함						
	①	②	③	④	⑤	⑥	⑦
2-1. 빅데이터를 이용하는 사람들은 그렇지 않은 사람들보다 우월해 보이는 것 같다							
2-2. 빅데이터를 이용하는 사람들은 대체적으로 업무평가에 좋은 점수를 받고 있는 것 같다							
2-3. 빅데이터 기술,서비스를 이용하면 다른 사람들에게 모범이 될 수 있다고 생각한다							
2-4. 나의 주변사람들은 빅데이터 기술이 나의 업무 영역에 도입이 필요하다고 생각할 것이다							

3. 빅데이터 기술과 관련한 과업적합성에 대한 질문

3. 빅데이터 기술과 관련한 과업적합성에 대한 질문	매우 그렇지 않음-----매우 그러함						
	①	②	③	④	⑤	⑥	⑦
3-1. 내가 사용하거나 사용하고 싶은 빅데이터 관련 기술, 서비스는 내 목적에 충분히 적합하다							
3-2. 기업 또는 부서의 데이터 활용 시 빅데이터관련 기술, 서비스는 데이터를 활용하기에 그 구조나 내용이 나에게 적합하다							
3-3. 빅데이터 기술,서비스를 이용하면 내가 필요한 데이터가 읽기 및 이해가 가능한 형식으로 제공 된다							
3-4. 빅데이터 기술, 서비스를 이용하면 정확히 내 업무를 실행할 때 필요하다							

4. 빅데이터 기술과 관련한 유용성에 대한 질문

4. 빅데이터 기술과 관련한 유용성에 대한 질문	매우 그렇지 않음-----매우 그러함						
	①	②	③	④	⑤	⑥	⑦
4-1. 빅데이터 기술, 서비스를 이용하면 보다 효과적인 결과를 기대할 수 있다							
4-2. 빅데이터 기술, 서비스를 이용하면 보다 빠르게 결과를 기대할 수 있다							
4-3. 빅데이터 기술, 서비스를 이용하면 보다 정확한 결과를 기대할 수 있다.							
4-4. 빅데이터 기술, 서비스를 이용하면 보다 신뢰할 수 있는 결과를 제공 받는다							
4-5. 빅데이터 기술, 서비스를 이용하면 보다 높은 생산성을 기대할 수 있다							

5. 빅데이터 기술과 관련한 용이성에 대한 질문

5. 빅데이터 기술과 관련한 용이성에 대한 질문	매우 그렇지 않음-----매우 그러함						
	①	②	③	④	⑤	⑥	⑦
5-1. 빅데이터 기술, 서비스를 이용하면 보다 쉽게 결과를 기대할 수 있다							
5-2. 빅데이터 기술, 서비스를 이용 시 방법을 익히는 것은 쉽다.							
5-3. 빅데이터 기술, 서비스를 이용 시 많은 노력이 필요하지 않다.							
5-4. 빅데이터 기술, 서비스를 이용 시 능숙하게 하는 부분은 어렵지 않다							
5-5. 빅데이터 기술, 서비스를 이용하면 원하는 결과를 쉽게 만들어 낼 수 있다							

6. 빅데이터 기술과 관련한 사용태도에 대한 질문

6. 빅데이터 기술과 관련한 사용태도에 대한 질문	매우 그렇지 않음-----매우 그러함						
	①	②	③	④	⑤	⑥	⑦
6-1. 모든 면을 고려해 볼 때, 빅데이터 기술, 서비스를 이용하는 것이 좋다고 생각된다.							
6-2. 모든 면을 고려해 볼 때, 빅데이터 기술, 서비스를 이용하여 업무처리를 하는 것이 현명하다고 생각된다.							
6-3. 모든 면을 고려해 볼 때, 빅데이터 기술, 서비스를 이용하여 업무를 하는 것은 긍정적이라 할 수 있다.							
6-4. 모든 면을 고려해 볼 때, 빅데이터 기술, 서비스를 이용하여 업무를 하는 것은 경제적이라고 할 수 있다.							
6-5. 모든 면을 고려해 볼 때, 빅데이터 기술, 서비스를 이용하여 업무를 하는 것은 시간을 절약할 수 있다.							

7. 빅데이터 기술과 이용한 업무를 통해 느낀 사용의도에 대한 질문

7. 빅데이터 기술과 이용한 업무를 통해 느낀 사용의도에 대한 질문	매우 그렇지 않음-----매우 그러함						
	①	②	③	④	⑤	⑥	⑦
7-1. 나는 앞으로 빅데이터 기술, 서비스를 업무에 적용할 것이다							
7-2. 나는 앞으로 빅데이터 기술, 서비스를 다른 사람에게 적극 추천할 것이다							
7-3. 나는 앞으로 빅데이터 기술, 서비스를 도입할 것으로 예상된다							
7-4. 나는 앞으로 빅데이터 기술, 서비스를 사용할 의향이 있다							

8. 빅데이터 기술과 서비스 이용을 통해 느낀 전반적 평가와
지속사용에 대한 질문

8. 빅데이터 기술과 서비스 이용을 통해 느낀 전반적 평가와 지속사용에 대한 질문	매우 그렇지 않음-----매우 그러함						
	①	②	③	④	⑤	⑥	⑦
8-1. 나는 앞으로 빅데이터 기술, 서비스를 언제나 나의 첫 번째 선택으로 고려할 것이다.							
8-2. 나는 앞으로 빅데이터 기술, 서비스를 지속적으로 이용할 것이다							

〈부록 3〉 소셜빅데이터 비만 담론 분석 속성별 키워드 분류

주 키워드	연관 키워드
원인	원발성비만, 단순비만, 단순성비만, 이차성비만, 증후성비만, 2차성비만
체질량지수	과체중, overweight, 위험체중, 1단계비만, 비만1단계, 2단계비만, 비만2단계, 3단계비만, 비만3단계, 비만위험군, 경도비만, 중등도비만, 고도비만, 초고도비만
성별	여성, 여자, 남성, 남자
연령별	태아, 저체중아, 낮은출생체중, 소아, 아동, 어린이, 학령전기, 학령기, 유아, 아이, 청소년, 10대, 학생, 초등학생, 중학생, 고등학생, 대학생, 여대생, 성인, 청년, 중년, 장년, 노년, 노인, 20대, 30대, 40대, 50대, 60대, 직장인, 회사원, 임신, 산후, 폐경기, 임신부, 산모
세팅별	임상, clinicalsetting, 공공보건, publichealth, 비임상, 지역사회, 보육시설, 어린이집, 유치원, 학교, 직장
부위	전신비만, 부분비만, 상체비만, 얼굴, 팔뚝, 허리, 하체비만, 둔부형비만, 하비, 허벅지, 종아리, 엉덩이, 복부비만, 복부형비만, 뱃살, 내장비만
식사습관	고지방식이, 고칼로리식이, 가공식품, 섭식장애, 지방, 기름기, 기름부위가많은고기, 튀김, 과다한지방섭취, 포화지방섭취증가, 트랜스지방섭취, 비스킷, 고지방, 고칼로리, 기름, 고기, 삼겹살, 포화지방, 트랜스지방, 패스트푸드, 인스턴트식품, 스낵, 탄산음료, 정크푸드, 햄버거, 콜라, 과자, 사탕, 초콜릿, 라면, 치킨, 피자, 빵, 밀가루, 식사횟수감소, 빠른식사시간, 과식, 다이어트, 야식증후군, 식탐, 야식
생활습관	활동부족, 좌식생활, 운동부족, 비활동성생활습관, TV시청, 컴퓨터게임, 대중교통, 지하철, 버스, 놀이공간, 수면부족, 불충분한수면, 금연, 금단, 흡연, 음주, 알코올중독, 음주문화, 잦은회식, 술, 회식, 안주
사회경제적요인	사회, 경제, 지위, 가난, 빈곤, 환경, 사회결정, 사회환경
심리적요인	스트레스, 디스트레스, 압박감, 불안, 외로움, 사랑결핍, 관심결핍
유전적요인	유전, 가족력
유전	비만유발유전자, 비만유전자, 유전자, ob, db, POMC, MC4R유전자, ob유전자, db유전자, POMC유전자, MC4R
인종	인종,흑인,황인,백인

주 키워드	연관 키워드
투약	항정신병약물, Thioridazine, Olanzapine, Clozapine, Quetiapine, Risperidone, 알파-2길항제, Imipramine, Mirtazapine, 선택적세로토닌재흡수억제제, SSRI, Paroxetine, 항전간제, Valproate, Carbamazepine, Gabapentin, 당뇨병치료제, 인슐린, 설fonyosozje, 글리나이드제제, 티아졸리디네돈, 세로토닌길항제, Pizotifen, 항히스타민제, Cyproheptadine, 베타차단제, Propranolol, 알파차단제, Terazosin, 스테로이드제제, 경구피임제, 당질코르티코이드제제, 티오리다진, 치오리다진, 올란제핀, 클로자핀, 쿼에티아핀, 리스페리돈, 이미프라민, 미트라자핀, 미르타스, 미르타자핀, 파록세틴, 항간질제, 벨프로에이트, 발프로에이트, 카르바마제핀, 항간질약, 가바펜틴, 가바페딘, 사이프로헵타딘, 프로프라놀롤
신체계측	BMI, Bodymassindex, 체질량지수, 신체질량지수, 신체비만지수, 케틀레지수, Queteletindex, 허리둘레, waistcircumference, 허리-엉덩이둘레비율, waist-hipratio, 허리엉덩이둘레비, WaistHipRatio, WHR, 신장, 키, 체중, 몸무게, 비만도, 상대체중법, 체지방량, 체지방률, 체내지방량, 근육량, 피부주름두께, 복부내장지방량, 체형
검사	지질성분검사, lipidprofile, 혈당, 공복시혈당, 간기능, 내분비기능, 혈압
기기이용평가	체성분분석기, 생체전기저항분석, 이중에너지X-선흡수법, DXA, 임피던스, 브로카변법, Broca, 복부CT, 복부MRI, X선, CT, MRI
위험요소사정(기타)	생활양식, 생활습관, 식이, 식사, 식이습관, 신체적활동, 운동, 활동, 유전, 가족력, 비만, 과체중가족력, 동반질환가족력, 동기화, 변화의지, 모티베이션, 정신적스트레스, 디스트레스
신체적증상	무증상, 트림, 소화불량, 목의통증, 호흡곤란, 갈증, 소변량증가, 잦은소변, 현기증, 요통, 두통, 불규칙한월경, 발진, 시력장애, 청력감소
정신적증상	스트레스, 디스트레스, 자살충동, 고민, 불안, 공포, 불면, 대인기피, 우울, 우울증, 신경성식욕부진, 신경성폭식증, 자존감저하, 자아존중감저하, 대인관계위축, 차별
근골격계질환	골관절염, 퇴행성관절염, 요통, 통풍, 골다공증, 성조숙증
내분비계질환	당뇨병, 제2형당뇨병, NIDDM, 대사증후군, 인슐린저항성
생식/호르몬계질환	성기능장애, 호르몬장애, 불임, 다낭성난소증후군
동반질환	비만관련선천성장애, 프라더-윌리증후군, 로렌스-문-비들증후군, 알스트롬증후군, 코헨증후군, 카펜터증후군, 신경및내분비계질환, 시상하부성비만, 외상, 종양, 감염성질환, 수술, 뇌압상승, 쿠싱증후군, 갑상선기능저하증, 인슐린종, 다낭난소증후군, Polycysticovarysyndrome, 성인성장호르몬결핍증, 정신질환, 과식장애, 계절성정서장애, 프라더윌리증후군, 프래더윌리증후군, 프레더윌리증후군, 로렌스문비들증후군, SAD

주 키워드	연관 키워드
소화기계질환	담낭질환, 담석증, 지방간
심혈관계질환	관상동맥질환, 심근경색증, 협심증, 부정맥, 율혈성심부전, 허혈성심질환, 비후성심근증, 심장발작, 돌연사, 고혈압, 고지혈증, 이상지질혈증, 고콜레스테롤혈증
신경및내분비계질환	시상하부성비만, 외상, 종양, 감염성질환, 수술, 뇌압상승, 쿠싱증후군, 갑상선기능저하증, 인슐린종, 다낭난소증후군, Polycysticovarysyndrome, 성인성장호르몬결핍증
호흡기계질환	수면무호흡증, 폐쇄성수면무호흡증, 수면무호흡, 호흡저하, 저환기장애, 코골이, 숨가쁨, breathlessness
정신질환	과식장애, 계절성정서장애, SAD
암	유방암, 대장암, 담낭암, 췌장암, 신장암, 방광암, 자궁경부암, 전립선암, 자궁내막암
합병증(기타)	마취, 정신장애, 모성비만, 태아, 뇌졸중, 고요산혈증, 요요현상, 정신장애
식사치료	식이요법, 다이어트, 단일식품다이어트, 칼로리제한다이어트, 저탄수화물다이어트, 저지방다이어트, 저인슐린다이어트, 채식다이어트, 고탄수화물다이어트, 지중해식다이어트, 군인다이어트, 마시는다이어트, 액체다이어트, Zone다이어트, 마요네즈다이어트, 원푸드다이어트, 채식, 덴마크다이어트, 황제다이어트, 식이요법교육, 영양교육, 제한식품, 제한영양소, 열량, 지방, 당질, 음주, 허용식품, 허용영양소, 단백질, 무기질, 저칼로리식사, 저열량, 저칼로리, 열량식품제한, 저지방식사, 기름기없는고기, 기름을넣지않고튀긴팝콘, 저지방식다이어트, 저지방우유, 저지방, 고단백식사, 고단백식이, 고단백다이어트, 단백질파우더, 계란찜, 삶은계란, 생선류, 흰살생선, 등푸른생선, 살코기, 장조림, 사태찜, 섬유질, 야채, 채소, 오이, 상추, 녹황색채소, 과일, 샐러드, 잡곡류, 잡곡밥, 현미, 콩, 강냉이, 비타민, 시금치, 당근, 콩나물, 배추, 무, 연근, 김, 미역, 김치, 버섯, 깻잎, 가지, 깍두기, 고사리, 단무지, 양배추, 쪽갓, 호박, 도라지, 피망, 풋고추, 양파, 토마토, 다시마, 건포도, 꿀, 수박, 자두, 딸기, 포도, 레몬, 참외, 단감, 사과, 건대추, 복숭아, 배, 과일주스, 칼슘, 보통우유, 전지분유, 요플레, 요구르트, 베이포, 잔멸치, 철, 달걀, 쇠고기, 조개, 인, 우유, 치즈, 파, 땅콩, 요오드, 파래, 나트륨, 소금
약물치료	비만치료제, 식욕억제제, 한약, 건강기능식품, 시부트라민, sibutramine, 오리스타트, orlistat, 올리스태트, 올리스타트,펜터민, phentermine, 허벌라이프, 알로에겔,쉐이크

주 키워드	연관 키워드
행동치료	자극조절기법, 자극조절법, 자극통제, 식사행동조절, 보상, 양성강화물, 강화물, 자기관찰, 영양교육, 신체활동조절, 문제해결, 대체행동기법, 인지적재구조화, 인지적재구조화기법, 인지재구성, 사회적지지프로그램, 사회적지지, 지지, 프로그램, 자기관찰기법, 일기, 식사행동일기, 활동일기, 감정일기, 수첩, 다이어트다이어리, 다이어트일기, 산보, 영화, 전화, 꽃, 목욕, 음악, 드라이브, 독서, 공원산책, 세차, 편지, 운동, 양치질, 집안청소, 쇼핑, 앨범정리, 동기부여, 건강한생활방식유지, 건강행태개입, 자가감시, self-monitoring, 자가모니터링, 스트레스관리
운동치료	저항성운동, 바벨, 덤벨, 운동밴드, 짐볼, 근력운동, 유산소운동, 자전거타기, 실내, 에어로빅댄스, 조깅, 벤치스텝에어로빅, 농구, 달리기, 하이킹, 축구, 걷기, 인라인스케이팅, 댄싱, 실내조정기구운동, 줄넘기, 라켓스포츠, 계단오르기, 수영, 배구, 엘립티컬운동, 수중에어로빅, 핸드볼, 에어로빅라이딩, 스쿼시, 줄넘기, 자전거, 실내자전거, 좌식싸이클, 에어로빅, 트레킹, 노젓기운동, 로잉머신, 배드민턴, 테니스, 계단, 싸이클런, 줄넘기, 복합성운동, 유연성운동, 운동처방, 스트레칭, 요가, 신체활동안내, 신체활동교육
수술	수술, 비만대사수술, 고도비만수술, 시술, 복강경수술, 조절형위밴드설치술, 루와이우회술, RY우회술, 지방흡입, 주사, 컷복시, 메조테라피, 보톡스, 지방이식, PPC주사, HPL
체중유지	체중감량목표, 신체활동유지, 행동치료병행, 지속적인체중감량, 식사조절, 스트레스관리, 정기적인체중측정, 정기적추적관리, 휴대폰단축메시지서비스제공, 방문서비스, 문자, 카카오톡
인식고취	교육, 캠페인, 홍보, 식사, 운동, 신체적활동, 행동치료, 생활습관
중재(전문가)	식사, 운동, 신체적활동, 행동치료, 다학제팀, 진료기관연계
중재(지역사회)	식사, 운동, 신체적활동, 행동치료, 지지, 우편, 전화, 인터넷, 자조모임, 지역기관, 지역환경개선, 정책개선
중재(직장)	식사, 운동, 신체적활동, 행동수정, 직장행동수정프로그램, 인센티브, incentive, 건강식제공

