



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

박사학위논문

빅데이터를 활용한 건강관리 경향 분석

Trend Analysis of Health Care Using Big Data

한국체육대학교 대학원

체 육 학 과

김 지 은

지도교수 김 현 태

2021년 2월

빅데이터를 활용한 건강관리 경향 분석

Trend Analysis of Health Care Using Big Data

한국체육대학교 대학원

체 육 학 과

김 지 은

이 논문을 박사학위 논문으로 제출함

지도교수 김 현 태

2021년 2월

국문요약

빅데이터를 활용한 건강관리 경향 분석

김 지 은

한국체육대학교 대학원

체육학과

이 연구는 빅데이터를 활용하여 시기별 건강관리 경향을 파악하는 것이 목적이다. 연구의 목적을 달성하기 위해 지난 30년간 빅카인즈(www.bigkinds.or.kr)에 수집된 뉴스 기사를 대상으로 python 3 프로그램을 이용하여 키워드 및 토픽 분석을 실시하였다. 먼저 1시기(1990년대), 2시기(2000년대), 3시기(2010년대)로 구분하여 시기별 건강관리 유형에 따른 주요 키워드를 추출하고 워드클라우드로 시각화하였다. 두 번째 키워드 빈도분석을 통해 상위20위 주요 키워드의 빈도, TF-IDF(단어빈도-역문서빈도) 값을 산출하여 빈도수가 큰 순서대로 나열하였다. 세 번째는 키워드 출현빈도 분석에서 나타난 시기별 공통/차별 키워드의 순위 변화를 확인하기 위해 카이제곱 검정을 실시하였다. 네 번째 시간흐름에 따라 공통 키워드 트렌드가 유의미하게 증감하였는지 확인하기 위해 Hot_Cold 분석과 회귀분석을 실시하였다. 다섯 번째 연구내용은 토픽모델링 분석을 통해 토픽 내 키워드의 비중을 확인함으로써 각 토픽마다 어떤 키워드가 주요하게 다뤄지고 있는지 확인하였다. 여섯 번째 시간흐름에 따라 각 토픽의 트렌드가 유의미하게 증감하였는지 확인하기 위해 Hot_Cold 분석과 회귀분석을 실시하였다. 연구 결과, <운동>1시기에 ‘선수’, ‘체력’, 2시기는 ‘유산소’, ‘스트레칭’, 3시기는 ‘피트니스’, ‘시간’ 키워드가 높은 비중을 차지하였다. <식이>는 전체시기에서 ‘식이요법’이 압도적으로 큰 비중을 차지하였으며 2시기는 ‘식이섭유’, 3시기에 ‘식이조절’ 키워드가 높은 비중을 차지하였다. <영양제>는 전체시기에서 ‘비타민’이 핵심 키워드로 나타났고 1시기에서 ‘영양’, 2시기에 ‘의약품’, 3시기는 ‘영양소’가 주요 키워

드로 도출되었다. 이 외에도 ‘유산균’, ‘셀레늄’, ‘오메가’, ‘코엔자임’ 등 다양한 키워드가 출현하였다. <음주>는 전체시기에서 ‘관계자’가 핵심키워드로 나타났으며 1시기, 2시기에 ‘가능성’, 3시기에 ‘절주’가 주요 키워드로 나타났다. <흡연>은 전반적으로 내용이 비슷하게 나타났는데 ‘담배’, ‘흡연’ 관련 키워드를 중심으로 1시기는 ‘금연운동’, 2시기, 3시기에서 ‘보건소’가 주요 키워드로 도출되었다. <휴식>1시기, 2시기에 ‘재충전’, ‘시간’이 주요 키워드로 나타났고 3시기에는 ‘힐링’이 가장 큰 비중을 차지하였다. <수면>에서도 ‘시간’이 주요 키워드로 나타나면서 1시기, 2시기에서 ‘불면증’, 3시기는 ‘관계자’가 주요키워드로 도출되었다. <스트레스>1시기는 ‘해소’, 2시기, 3시기에서 ‘시간’이 주요 키워드로 나타났으며 스트레스를 받는 대상자 관련 키워드가 다수 출현하였다. 이와 같이 건강관리 유형에 따른 주요 키워드는 다르게 나타났으며 공통/차별 키워드는 순위에 유의미한 차이가 있는 것을 확인하였다($p<.001$). 회귀분석과 Hot_Cold분석을 이용하여 공통 키워드의 연도별 트렌드 추이를 살펴본 결과 운동, 영양제, 음주, 흡연에서 Hot(정적상관)이, 식이, 휴식에서 Cold(부적상관)이 유의미하게 나타났으며($p<.05$) 수면, 스트레스에서는 유의미한 차이가 나타나지 않았다. LDA 토픽모델링 분석을 위해 건강관리 유형별로 토픽 수($k=7$)를 지정하고 비중이 큰 순으로 토픽 번호를 부여하였다. 그 결과 <운동>토픽1:피트니스, 토픽2:유산소운동, 토픽3:체력, <식이>토픽1:식이요법, 토픽2:운동과 식이요법, 토픽3:환자 치료식, <영양제>토픽1:건강기능식품, 토픽2:제품, 토픽3:의약품, <음주>토픽1:절주, 토픽2:부작용 및 문제점, 토픽3:치료법, <흡연>토픽1:금연정책, 토픽2:금연구역, 토픽3:금연교육, <휴식>토픽1:힐링, 토픽2:휴가여행, 토픽3:재충전, <수면>토픽1:스마트폰 사용, 토픽2:수면부족, 토픽3:수면시간, <스트레스>토픽1:발병 및 치료, 토픽2:힐링, 토픽3:스트레스 해소로 분류되었다. 회귀분석과 Hot_Cold분석을 이용하여 토픽의 트렌드 추이를 살펴본 결과, 각 항목별 부분적으로 Hot(정적상관), Cold(부적상관)에 유의한 차이가 나타났다($p<.05$). 결론적으로 이 연구는 빅데이터를 활용하여 키워드 및 토픽 분석을 통해 시기별 건강관리 경향을 파악하여 시대적 특성을 반영한 건강관리 정보를 제공함으로써 향후 역량강화를 위한 건강정책 및 과제의 기초자료로 활용될 수 있을 것이다.

목 차

I. 서 론	1
1. 연구의 필요성	1
2. 연구 가설	5
II. 연구 방법	6
1. 연구 절차	6
2. 데이터 수집	7
1) 분석 자료	8
2) 분석 키워드 선정	9
3) 키워드 검색	10
4) 시기 구분	13
3. 데이터 클리닝	14
4. 데이터 분석	14
1) 키워드 분석	14
2) 토픽 분석	15
5. 연구의 타당도	17
6. 자료처리	17
1) 빅카인즈의 분석방법	17
2) python 3 프로그램	17
III. 연구 결과	18
1. 시기별 건강관리 유형에 따른 주요 키워드	18
1) 운동	18

2) 식이	19
3) 영양제	20
4) 음주	21
5) 흡연	22
6) 휴식	23
7) 수면	24
8) 스트레스	25
2. 시기별 건강관리 유형에 따른 키워드 출현빈도	26
3. 시기별 건강관리 유형에 따른 공통·차별 키워드	35
4. 시기별 건강관리 유형에 따른 키워드 트렌드	41
5. 시기별 건강관리 유형에 따른 주요 토픽	42
6. 시기별 건강관리 유형에 따른 토픽 트렌드	51
 IV. 논의	 54
 V. 결론	 61
 참고문헌	 64
 ABSTRACT	 69

표 목차

표 1. 시기별 건강관리 유형에 따른 뉴스기사 수	8
표 2. 시기별 건강관리 유형에 따른 키워드 수	9
표 3. 검색 키워드 선정	10
표 4. 빅카인즈 검색연산자 옵션	12
표 5. 연구에 적용한 검색 키워드 범위	12
표 6. python 3에서 사용한 라이브러리 패키지	17
표 7. 시기별 운동 상위20 키워드 출현빈도	26
표 8. 시기별 식이 상위20 키워드 출현빈도	27
표 9. 시기별 영양제 상위20 키워드 출현빈도	28
표 10. 시기별 음주 상위20 키워드 출현빈도	29
표 11. 시기별 흡연 상위20 키워드 출현빈도	30
표 12. 시기별 휴식 상위20 키워드 출현빈도	31
표 13. 시기별 수면 상위20 키워드 출현빈도	32
표 14. 시기별 스트레스 상위20 키워드 출현빈도	33
표 15. 시기별 운동 공통/차별 키워드	35
표 16. 시기별 식이 공통/차별 키워드	35
표 17. 시기별 영양제 공통/차별 키워드	36
표 18. 시기별 음주 공통/차별 키워드	37
표 19. 시기별 흡연 공통/차별 키워드	38
표 20. 시기별 휴식 공통/차별 키워드	38
표 21. 시기별 수면 공통/차별 키워드	39
표 22. 시기별 스트레스 공통/차별 키워드	40
표 23. 건강관리 유형에 따른 공통 키워드 Hot_Cold	41

표 24. 건강관리 유형에 따른 공통 키워드 트렌드 추이	42
표 25. 운동 토픽모델링 분석결과	43
표 26. 식이 토픽모델링 분석결과	44
표 27. 영양제 토픽모델링 분석결과	45
표 28. 음주 토픽모델링 분석결과	46
표 29. 흡연 토픽모델링 분석결과	47
표 30. 휴식 토픽모델링 분석결과	48
표 31. 수면 토픽모델링 분석결과	49
표 32. 스트레스 토픽모델링 분석결과	50
표 33. 건강관리 유형에 따른 토픽별 Hot_Cold	51
표 34. 건강관리 유형에 따른 토픽 트렌드 추이	52

그림 목차

그림 1. 연구절차	6
그림 2. 빅카인즈 뉴스분석 시스템	8
그림 3. 빅카인즈 키워드 검색창	11
그림 4. 빅카인즈 키워드 상세검색	11
그림 5. 헬스케어 패러다임 변화	13
그림 6. 워드클라우드 생성 코드	14
그림 7. 잠재 디리클레 할당 분석 모델의 개념	16
그림 8. 토픽모델링 생성 코드	16
그림 9. 시기별 운동 워드클라우드	18
그림 10. 시기별 식이 워드클라우드	19
그림 11. 시기별 영양제 워드클라우드	20
그림 12. 시기별 음주 워드클라우드	21
그림 13. 시기별 흡연 워드클라우드	22
그림 14. 시기별 휴식 키워드클라우드	23
그림 15. 시기별 수면 워드클라우드	24
그림 16. 시기별 스트레스 워드클라우드	25
그림 17. 운동 Intertopic Distance Map(IDM)	43
그림 18. 식이 Intertopic Distance Map(IDM)	44
그림 19. 영양제 Intertopic Distance Map(IDM)	45
그림 20. 음주 Intertopic Distance Map(IDM)	46
그림 21. 흡연 Intertopic Distance Map(IDM)	47
그림 22. 휴식 Intertopic Distance Map(IDM)	48
그림 23. 수면 Intertopic Distance Map(IDM)	49

그림 24. 스트레스 Intertopic Distance Map(IDM)	50
--	----

I. 서 론

1. 연구의 필요성

첨단의료기술의 발달로 건강수명에 대한 기대가 높아지면서 건강관리 패러다임이 질병 치료중심에서 벗어나 예방·맞춤형 건강관리 형태로 변화되고 있다. 특히 스마트헬스케어의 등장으로 언제 어디서나 자신의 건강상태를 확인 또는 관리하는 것이 가능해짐으로써 건강에 대한 인식과 관심이 고조되고 있다. 하지만 인터넷과 스마트폰의 과다 사용으로 인한 정신건강 문제, 좌식생활로 인한 비만환자 증가 등 생활습관 변화에 따른 건강상의 문제들이 끊임없이 도출되고 있는 상황이다. 뿐만 아니라 환경오염, 미세먼지와 같은 환경적 건강위해요인이 증가하고 운동부족, 음주, 흡연, 스트레스, 식생활의 변화로 인한 만성질환에 고통 받고 있는 것이 현실이다. 이처럼 급격하게 변화하는 현대사회에서 나타나는 건강 문제와 관련하여 예방적 건강관리의 중요성이 대두됨에 따라 관리방법도 다양해지고 있다.

과거 공중 보건이나 보건교육은 예방접종과 전염병 통제에 대한 캠페인에 한정되었으나 1990년대에 이르러 국민건강증진사업이 시행되면서 개인의 건강증진과 질병예방을 위한 생활습관 개선에 중점을 두었다. 대표적인 예로 금연운동이 있으며 당시 국가적 금연정책 및 캠페인의 영향으로 약 11% 정도의 남성 흡연율이 감소되었다(이종구, 2004). 국민건강증진사업의 영향으로 1991년 국민생활체육회가 설립된 이후 체육시설에 대한 인프라가 확충되면서 국민의 체육활동 참여가 더욱 활성화되었고 질병예방 차원의 신체활동 중요성이 대두되었다(김홍남, 최정웅, 2008). 국민생활체육조사에 따르면 규칙적인 체육활동은 신체적·정신적 건강유지에 효과가 있는 것(85%이상)으로 조사되었다(통계청 공식블로그, 2019). 이와 같이 체력 및 근력, 근지구력 향상, 체지방 감소, 정신건강 등 운동의 긍정적 효과가 입증되면서(Gould et. al., 2014; Solomon et al., 2008; Kokkinos & Papademetriou, 2000) 지금까지도 건강관리의 주요 요인으로 인식되고 있다.

2000년대 후반 웰빙(well-being)이 사회적 이슈로 떠올라 신체적·정신적 건강의 조화를 통해 행복한 삶을 추구하는 유형으로 확장되면서 새로운 웰빙 문화가 형성되었다. 그 중에

하나가 건강기능식품 및 친환경 농산물, 슬로푸드(slow food) 섭취이다. 홍삼을 비롯해 건강 제품에 대한 관심이 증가하면서 다양한 브랜드의 건강식품이 출시되기 시작하였고 현대인의 일상 속 새로운 식문화로 자리매김하였다(박명은, 유소이, 2109). 2000년-2014년까지는 웰빙을 시작으로 행복 소비트렌드가 경제적 환경 변화에 따라 쾌락적 행복론 중심의 개인주의화로 심화되면서 2015년을 기점으로 힐링(healing) 열풍이 불었다. 각박한 사회 속에서 공감과 위로, 치유에 대한 사회적 요구가 급증하면서 힐링 트렌드가 빠르게 확산된 것이다(엄경희, 정보리, 2014; 조경두, 2014). 호텔에서 즐기는 바캉스라는 의미의 ‘호캉스’라는 신조어가 생길만큼 편안하고 안락한 힐링시간을 보내기 위한 호캉스족이 증가했고 릴렉스(relax) 즉 휴식의 의미에서 여행과 스파, 찜질, 마사지, 명상, 요가 등 다양한 방법으로 육체적 건강은 물론 정신적 건강까지 관리하면서 만족감을 높이고 있다. 이처럼 건강관리의 영역이 넓어지면서 개인의 생활패턴을 반영한 맞춤형 건강관리로 변화하는 추세이다.

최근 몇 년 간 혼족, 얼로너(aloner)가 증가하면서 집에서 혼자 하는 운동을 뜻하는 ‘홈트’가 큰 인기를 얻고 있다. 또한 홈과 이코노미의 합성어인 ‘홈코노미’라는 단어가 등장하면서 집이 단순한 주거공간을 넘어 휴식, 문화, 레저를 즐기는 공간으로 확대되어 대부분의 여가와 소비 활동까지 해결하는 현상이 증가하고 있다. 이와 같이 사회문화적 트렌드를 반영한 건강정보 및 제품, 건강서비스는 소비욕구와 연결되어 개인의 라이프스타일에 크게 관여하여(김소연, 김선이, 2012) 결과적으로 삶의 광범위한 영역과 패턴이 건강 관련 행동으로 나타나면서 새로운 건강관리 문화를 형성하기도 한다.

현재의 건강관리 트렌드는 셀프 메디케이션(Self-Medication)이라고 할 수 있다. 셀프 메디케이션은 자신의 건강을 스스로 챙기는 것으로 건강관리에 집중 투자하고 소비를 아끼지 않는 현상을 말한다(WHO, 1998). 통계청 생활시간조사에 따르면 개인의 건강관리, 외모관리 시간은 계속 증가하고 있으며 특히 홈 뷰티, 건강기능식품, 홈 트레이닝 분야에 관심이 집중되면서 셀프케어를 통해 신체적·정신적·감정적 스트레스로부터 자신을 지키려는 것으로 나타났다(통계청 공식블로그, 2020). 이처럼 신체적 건강뿐 아니라 웰니스(Wellness) 측면 건강이 강조되면서 셀프케어와 같이 자기 주도적 건강관리의 필요성이 대두되고 있다.

최근 스마트 헬스케어의 등장과 의료기술 발달로 치료중심에서 예방·관리 중심의 건강 관리 패러다임으로 변화하고 있다. 헬스케어의 패러다임 변화에서 중요한 요소는 데이터의 생산과 공유, 그리고 효율적 활용이다. 실제로 스마트기기의 보급은 빅데이터 생산의 주요 활로가 되었고 정보통신의 발달과 웹 미디어의 보급이 확대되면서 방대한 양의 데이터를 축적하였다(송태민, 2013). International Business Machines(IBM) 보고서에 따르면 전 세계 정보의 80%는 비정형(unstructured) 데이터이며 비정형 데이터의 증가율은 정형 데이터 증가율의 15배에 이른다(Zikopoulos et al., 2012). 이러한 지식정보화 시대 속에서 빅데이터를 활용하여 새로운 지식과 의미를 발견하고 미래를 예측하는 것은 충분한 가치가 있다(안창원, 황승구, 2012; Bertino, 2013).

빅데이터란 데이터의 형식이 다양하고 생성속도가 빨라서 새로운 관리, 분석 방법이 필요한 대용량 데이터를 의미하며(권영욱, 2013) 정형 또는 비정형 데이터로부터 의미 있는 정보를 추출하고 결과를 분석하는 기술을 말한다. 빅데이터는 방대한 양의 데이터를 수집-저장-분석을 통해 특정 사회현상을 파악하고 미래를 예측하는 것이 가능하며(이은별, 전진오, 백지선, 2017) 실시간으로 생성되는 대용량의 텍스트 자료를 효과적으로 처리할 수 있다는 강점이 있어 교육, 행정, 비즈니스 등 다양한 분야에 활용되고 있다(Blei, Lafferty, 2007). 특히 비정형 데이터를 처리하기 위한 분석기법인 텍스트 마이닝(text mining)은 “비정형화된 데이터를 정형화된 데이터로 변환시켜 분석하는 일련의 과정”으로 정해진 구조가 없는 비정형데이터의 개념이나 특성을 추출할 수 있다(이상훈 외, 2015). 그 중에서 키워드 분석(keyword analysis)과 토픽모델링(topic modeling) 분석은 비정형데이터 자료에서 주요 키워드를 추출하고 키워드 간 의미와 구조를 파악할 수 있는 분석기법으로 학술 분야의 새로운 방법론으로 주목받고 있다(유예림, 2017). 이처럼 빅데이터의 활용은 설문이나 인터뷰 자료가 갖는 한계점을 보완할 수 있는 새로운 자료수집 방법으로 소개되면서 4차 산업시대에 핵심 수단으로 국가 경쟁력의 중요한 요소로 대두되고 있다.

그동안의 건강 관련 동향연구(권태연, 2017; 안영창, 2014; 사공진, 2011)는 국민건강보험 표본코호트 데이터베이스, 건강보험제도의 만족도 분석, 재정안정화와 수가정책 현황 분석

등 건강보험 관련 주제에 초점을 맞추고 있는 반면 사회문화 변화를 반영한 건강관리 경향 분석은 미비한 실정이다(이수정, 2020). 이전의 동향분석에서는 구조화된 설문지를 이용하거나 주로 문헌조사와 전문가 평가, 델파이 기법과 같은 질적 연구방법이 동원되었지만 이러한 방법은 대용량의 데이터를 분석하기 위해 많은 시간과 비용이 소모되며 연구자의 주관적 가치가 반영될 수 있다는 지적이 이어왔다(송태민, 2015). 특히 국가를 포함한 공공기관의 사회조사 자료는 장기간의 사회동향 여부를 살펴보는데 효과적일 수 있으나 조사내용의 일관성, 반복성으로 인해 사회의 변화를 따라가지 못하는 경우가 발생할 수 있으며 조사의 주제가 분명하게 고정되어 있는 경우 확장성을 고려한 조사 설계가 어렵다는 한계가 있다(박주섭, 김나랑, 한은정, 2018; 손준석 외, 2012). 따라서 조사를 통한 기존의 정보수집 체계의 한계를 보완하기 위해 데이터 마이닝과 같은 새로운 수집방법을 활용한 동향분석의 필요성이 요구된다(류지혜, 2018; 김현정 외, 2014).

이처럼 빅데이터를 활용한 연구의 요구와 수요가 증가하는 현 시점에서 뉴스기사와 같이 일회성으로 읽고 지나가는 자료를 대상으로 데이터분석을 실시하여 특정한 이슈 및 패턴이나 구조를 살펴보는 것은 의미 있는 일이다. 특히 토픽을 도출하는 것은 일반적인 뉴스 생산과 보도행위의 최종 결과물이 무엇인지 파악하는데 단초를 제공할 뿐 아니라 방대한 양의 데이터가 지니는 의미와 패턴을 파악할 수 있다는 점에서 다양하게 활용되고 있다(박승정 외, 2017). 이처럼 빅데이터 분석의 핵심은 연구대상의 확장과 추리통계만으로 알아낼 수 없는 범주를 찾는 것이 가능할 뿐 아니라 장기간의 데이터를 인식하는데 용이하도록 정리하고 시각화하는 것이다(Manovich, 2011).

따라서 이 연구는 빅데이터를 활용하여 지난 30년간 건강관리 경향이 어떻게 변화하였는지 파악하는 것이 목적이다. 급격하게 변화하는 현대사회에서 그 동안의 건강관리 경향을 살펴보고 더 나아가 보건의료, 국가정책, 과학기술 발달을 반영한 건강관리 전망 및 과제를 모색하고자 한다.

2. 연구 가설

이 연구는 빅데이터를 활용하여 건강관리 경향을 분석하는데 목적이 있다. 이 연구의 목적을 달성하기 위해 시기별, 건강관리 유형별, 분석방법별로 분류하여 건강관리 경향을 파악하고자 한다. 이 연구의 목적을 바탕으로 설정한 연구가설은 다음과 같다.

가설 1. 시기별 건강관리 유형에 따른 주요 키워드는 다를 것이다.

가설 2. 시기별 건강관리 유형에 따른 키워드 출현빈도는 다를 것이다.

가설 3. 시기별 건강관리 유형에 따른 공통/차별 키워드는 다를 것이다.

가설 4. 시기별 건강관리 유형에 따른 키워드 트렌드는 다를 것이다.

가설 5. 시기별 건강관리 유형에 따른 주요 토픽의 구조적 특성은 다를 것이다.

가설 6. 시기별 건강관리 유형에 따른 토픽 트렌드는 다를 것이다.

II. 연구 방법

1. 연구절차

이 연구는 연구절차는 <그림1>과 같다.



그림1. 연구절차

데이터 수집을 위해 빅카인즈(www.bigkinds.or.kr) 뉴스 빅데이터 분석시스템을 이용하여 건강관리 유형별 ‘운동’, ‘식이’, ‘영양제’, ‘음주’, ‘흡연’, ‘휴식’, ‘수면’, ‘스트레스’ 8개 항목으로 나누고 각 항목별 키워드 상세검색을 실시하였다. 검색 결과는 메타데이터 (언론사, 기고자, 제목, 키워드)가 포함된 텍스트 데이터를 .xlsx 파일형식으로 다운로드 하여 데이터 분석 자료로 이용하였다.

데이터 클리닝 단계에서는 빅카인즈 뉴스 데이터 분석시스템과 python 3 프로그램을 이용하였다. 키워드 검색을 통해 수집된 자료에서 빅카인즈 자체 알고리즘에 의해 추출된 명사 형태의 키워드는 python 3 프로그램을 이용하여 숫자, 영어, 한문 등 불용어(stopword)

를 제거하였다. 각 문서마다 명사를 추출하여 문서-단어 행렬(document-term matrix)을 생성하고 확보된 최종 데이터를 이용하여 키워드 및 토픽 분석을 실시하였다.

키워드 분석을 위해 시기별 건강관리 유형에 따른 상위50 주요 키워드를 추출하여 워드 클라우드(word cloud)를 생성하고 이를 통해 주요 키워드 특성을 파악하였다. 또한 상위20 주요 키워드를 이용하여 출현빈도 산출하고 TF, TF-IDF 두 가지 방식을 적용한 주요 키워드의 빈도 순위를 비교하였다. 키워드 빈도분석 결과를 토대로 공통/차별 키워드를 분류하고 연도별 공통 키워드 트렌드 변화를 확인하기 위해 카이제곱 검정과 Hot_Cold 분석을 실시하였다. 토픽 분석에서는 python 3의 gensim을 이용하여 키워드 간 특성을 기준으로 묶어주는 LDA 모델 기반 토픽모델링 분석을 실시하였다. 건강관리 유형별 주요 토픽을 파악하고 토픽 주제 간 유사성, 토픽 내 키워드의 비중을 산출하여 연도별 토픽 트렌드를 시각화하였다. 키워드 및 토픽 분석에서 얻은 결과는 python 3 프로그램을 이용하여 word cloud, visualizing LDA와 같은 방법으로 시각화하였다.

2. 데이터 수집

이 연구의 데이터 분석을 위해 빅카인즈(www.bigkinds.or.kr)에서 제공하는 온라인 뉴스 기사를 활용하였다. 빅카인즈는 1990년 이후 현재까지 6천8백만 건이 넘는 국내 최대 뉴스 데이터베이스(DB)에 빅데이터 분석기술을 접목한 뉴스분석서비스이다. 지상파 방송사 MBC, SBS, KBS, 그리고 뉴스 전문 채널 YTN과 경기·인천지역 민영방송사인 OBS까지 5개 방송사와 <조선일보>, <중앙일보>, <동아일보>의 주요 언론사 뉴스를 포함한 중앙지 11개, 경제지 8개, 지역종합지 28개, 전문지 2개 총 54개 언론사에 대한 뉴스를 제공한다. 빅카인즈의 뉴스추출 방식은 코딩 대신 자연어처리(natural language processing: NLP) 방식을 적용하여 코더 간 신뢰도를 별도로 측정할 필요가 없으며 알고리즘 기반 데이터 자동화 처리를 통해 데이터 수집 과정에서의 오류를 최소화할 수 있다. 이러한 장점으로 최근 많은 연구에서 활용하고 있는 추세이다(박상준, 원도연, 2020; 곽재현, 홍지숙, 2018; 한승희, 2018; 이은별, 전진오, 백지선, 2017). 빅카인즈 뉴스분석 시스템의 데이터 분석 절차는 <그림2>와 같다.

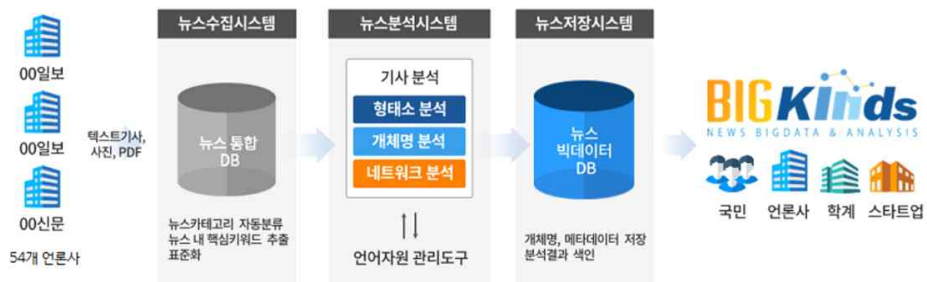


그림2. 빅카인즈 뉴스분석 시스템

1) 분석 자료

이 연구의 분석대상은 1990.01.01.-2019.12.31 기간 동안 빅카인즈 사이트에 수집된 건강 관리 관련 뉴스기사이다. 시간흐름에 따라 데이터양이 급격하게 증가하면서 뉴스기사 수와 키워드 수도 함께 증가하였다. 시기별 건강관리 유형에 따른 데이터양을 정리하면 다음과 같다.

표1. 시기별 건강관리 유형에 따른 뉴스기사 수

구분	1990년대	2000년대	2010년대	총합
운동	4,755	8,495	53,151	66,401
식이	1,358	5,026	14,220	20,604
영양제	227	3,842	11,292	15,361
음주	1,080	4,284	7,584	12,948
흡연	4,369	19,875	48,431	72,675
휴식	684	4,156	25,123	29,963
수면	3,002	3,777	36,961	43,740
스트레스	12,631	61,955	180,822	255,408
합계	28,106	111,410	377,584	517,100

〈표1〉은 이 연구의 데이터 분석을 위해 건강관리 유형별 검색 키워드로 수집한 뉴스기사건수이며 전체시기에서 수집된 뉴스기사는 총 517,100건이다. 시기별 뉴스기사건수가 급격하게 증가하였는데 특히 1990년대에서 2000년대로 전환되면서 뉴스 기사건수가 폭발적으로 증가하였다.

표2. 시기별 건강관리 유형에 따른 키워드 수

구분	1990년대	2000년대	2010년대	총합
운동	131,762	206,355	425,233	763,350
식이	38,915	127,789	333,241	499,945
영양제	6,541	94,225	247,483	348,249
음주	29,053	101,935	164,075	295,063
흡연	122,011	509,425	436,918	1,068,354
휴식	20,428	98,102	433,829	552,359
수면	92,406	342,356	419,427	854,189
스트레스	31,218	475,414	435,848	942,480
합계	472,334	1,955,601	2,896,054	5,323,989

〈표2〉는 이 연구의 데이터 분석에 사용된 키워드 수이며 전체시기에서 수집된 키워드는 총 5,323,989개이다. 뉴스기사건수와 마찬가지로 시간흐름에 따라 키워드 수가 점차적으로 증가하고 있다.

2) 분석 키워드 선정

데이터 분석에 앞서 데이터 수집을 위한 검색 키워드를 선정하였다. 검색 키워드 선정은 건강유지방법 또는 체력유지 통계청 사회조사 자료를 근거로 건강관리 유형(운동, 식이, 영양제, 음주, 흡연, 휴식, 수면, 스트레스)에 대한 키워드 검색 및 분석을 실시하였다.

표3. 검색 키워드 선정

통계청(2019) 건강유지방법	분석데이터 키워드 범위
1. 규칙적인 체육활동	운동
2. 규칙적인 식사	식이
3. 영양 보충	영양제
4. 금주	음주
5. 금연	흡연
6. 충분한 휴식	휴식
7. 수면	수면
8. 스트레스 안 받기	스트레스

이 연구의 분석 키워드 선정은 <표3>과 같다. 검색 키워드 범위는 라이프스타일에 관한 연구(유단비, 2010; 문옥륜 외, 2000)에서 건강생활실천 행동으로 제시한 항목과 질병관리본부(2016)의 건강행태 개선사업 관점에서 제시한 내용을 모두 포함하고 있어 건강관리 유형에 대한 분류는 타당성을 갖는다. 또한 사전 예비분석을 통해 얻은 결과에서 ‘건강관리’ 단일 주제어만으로는 건강관리행동 관련 상세내용을 파악하는데 한계가 있었기 때문에 정밀한 결과도출을 위해 통계청, 문화체육관광부에서 시행한 사회조사의 건강관리행동과 건강유지방법 평가항목을 참고하여 텍스트 상 주로 사용하는 단어를 근거로 자료검색에 부합하는 키워드를 설정하였다.

3) 키워드 검색

빅카인즈(www.bigkinds.or.kr) 웹사이트에 접속하여 아래 <그림3>과 같이 검색창에 건강관리 유형별(운동, 식이, 영양제, 음주, 흡연, 휴식, 수면, 스트레스) 검색 키워드를 입력 후 검색조건 설정에서 기간, 언론사, 통합분류, 사건사고 분류, 상세검색 메뉴를 이용하여 연구 목적에 맞는 뉴스 기사를 검색하였다.



그림3. 빅카인즈 키워드 검색창

키워드 상세검색 단계에서는 다양한 검색 필터를 통해 검색 유형, 검색어 처리, 검색어 범위를 선정하여 해당 키워드의 검색 범위를 한정시킴으로써 결과의 적합성 및 정확성을 확보하였다. 상세검색 메뉴는 아래 <그림4>와 같다.

그림4. 빅카인즈 키워드 상세검색

상세검색을 통해 기간을 설정하고 검색키워드 범위를 검색 연산자를 사용하여 더 정확한 결과를 얻을 수 있도록 하였다. 먼저 검색어 처리에서 형태소 분석을 설정한 다음 상세 범위를 결정하는 과정에서 검색 단어 중 1개 이상 포함, 모두 포함, 정확하게 일치하는 단어, 제외 단어를 AND, OR, NOT 등과 같이 설정하여 복합 검색어를 구성하였다. 이 때 두자 이상의 검색어로 검색할 경우는 쉼표(.)를 넣어 복합검색을 실시하였다. <표4>의 빅카인즈 사용자 매뉴얼에서 제시하는 검색연산자 옵션을 참고하여 <표5>와 같이 이 연구에서 적용한 검색 키워드 범위를 설정하였다.

표4. 빅카인즈 검색연산자 옵션

구분	설명
단어 중 1개 이상 포함	기본 검색어를 포함하면서 입력한 단어 중 하나라도 포함된 뉴스를 검색
다음단어 모두 포함	기본검색어와 입력한 단어를 모두 포함하는 뉴스를 검색하는 조건 설정
정확히 일치하는 단어	기본검색어를 포함하면서 입력한 단어/문장이 변형 없이 입력된 순서대로 정확하게 일치하는 단어/문장을 검색
다음단어 제외	입력한 단어가 포함된 뉴스는 검색결과에서 제외하는 조건을 설정

표5. 연구에 적용한 검색 키워드 범위

유형	검색 키워드
운동	운동 AND “신체활동” OR “체육활동” OR “피트니스” OR “트레이닝” NOT 선거, 총선, 민주화, 만세, 독립, 새마을, 시민, 반대, 미투 등
식이	식이 AND “식단” OR “식습관” OR “식생활” OR “건강식” OR “웰빙식” OR “균형식”
영양제	건강기능식품 AND “영양제” OR “건기식” OR “보조제” OR “보충제” OR “기능성식품” OR “건강기능식품”
음주	음주 AND “금주” OR “단주” OR “절주” OR “알코올” OR “술”
흡연	흡연 AND “금연” OR “단연” OR “절연” OR “담배”
휴식	휴식 AND “릴렉스” OR “피로회복” OR “재충전” OR “힐링” OR “안정” OR “휴양”
수면	수면 AND “숙면” OR “잠” OR “램수면” OR “수면질” NOT 해수면, 바다, 해양, 하천, 낚시, 매립, 공장, 공유수면, 건설 등
스트레스	스트레스 AND “화병” OR “화병” OR “화증”

예를 들어 1시기 운동에 대한 데이터를 수집하기 위해 검색창에 ‘운동’ 키워드를 넣은 다음 상세검색 메뉴를 이용하여 기간은 1990.01.01.-1999.12.31.을 설정하고 이 연구에서 적용한 검색 키워드 범위에 맞게 ‘신체활동’, ‘체육활동’ 등 포함시킬 키워드를 추가한다. 이 때 연구 내용과 관계없는 키워드가 도출되지 않도록 ‘선거 운동’, ‘독립 운동’ 과 같이 키워드 정제 과정에서 의미가 없거나 불필요한 단어는 불용어로 설정한다.

4) 시기 구분

이 연구의 데이터 수집을 위한 시기 구분은 삼성경제연구소(2012)에서 제시한 헬스케어 시대 1.0(공중보건의 시대)부터 2.0(질병치료의 시대), 3.0(건강수명의 시대) 변화를 반영하여 1990년대, 2000년대, 2010년대로 설정하였다. 이 연구에서 참고한 헬스케어 시기구분 내용은 아래 <그림5>와 같다.



그림5. 헬스케어 패러다임 변화

경향 파악을 위한 시기구분은 1990년 01월 01일부터 2019년 12월 31일까지 30년 기간을 10년 단위로 나누어 <1시기: 1990.01.01-1999.12.31>, <2시기: 2000.01.01-2009.12.31.>, <3시기: 2010.01.01-2019.12.31> 총 3시기로 구분하였다. 1990년을 데이터 시작점으로 설정한 이유는 빅카인즈에서 제공하는 자료의 최초년도를 반영한 것이며 데이터 종료시점으로 2020년을 포함하지 않은 이유는 코로나19 영향으로 코로나 관련 키워드가 기하급수적으로 증가했기 때문에 데이터 수집에 따른 연구결과에 큰 영향을 미칠 것이라고 판단하여 내린 결정이다.

3. 데이터 클리닝

빅데이터를 활용하여 수집한 텍스트 자료는 데이터 전처리 과정을 거쳐 분석에 적합한 형태로 만드는 것이 선행되어야 한다(Lucas et al., 2015). 이에 따라 빅카인즈 뉴스분석시스템 자체 알고리즘을 적용하여 데이터 정제가 완료된 명사형태의 키워드를 토대로 문서-단어 행렬(Document-term matrix; DTM)을 생성하였다. 빅카인즈에서 추출된 명사형태의 키워드는 python 3 프로그램의 nltk패키지를 이용하여 한자, 숫자, 영어 등 한글 이외에 문자는 제거하고 연구내용에 부적합한 키워드는 추가적인 불용어 처리를 통해 데이터 클리닝을 실시하였다.

4. 데이터 분석

1) 키워드 분석

이 연구는 키워드 분석을 통해 시기별 건강관리 유형에 따른 주요 키워드가 무엇인지 살펴보고 빈도분석을 통해 각 항목별 자주 언급되는 상위 주요 키워드를 파악했다. 추출된 결과는 python 3 프로그램의 pytagcloud 패키지를 이용하여 워드클라우드(word cloud)를 생성하였다. 각 항목별 주요 키워드를 말뭉치로 시각화하기 위한 워드클라우드 생성 코드는 아래 <그림6>과 같다.

```
wc = wordcloud.WordCloud(  
    font_path='malgun.ttf',  
    background_color="white",  
    width=1280,  
    height=800,  
    max_words=20,  
    max_font_size=300  
)  
  
wc.generate(", ".join(keywords))
```

그림6. 워드클라우드 생성 코드

키워드 분석을 위해 python 3 프로그램의 scikit-learn 패키지를 이용하여 문서-단어 행렬(DTM)을 생성한 후 출현빈도가 높은 키워드를 분석하는 단어 빈도(Term Frequency) 분석을 실시하였다. 단어 빈도와 단어빈도-역문서빈도(Term Frequency-Inverse Document Frequency, TF-IDF) 두 가지 기준을 적용하여 빈도가 큰 순으로 나열하였다. TF-IDF는 텍스트 마이닝 분석에서 가장 많이 쓰는 가중치로 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 통계적 수치로 나타낸 것이다. 키워드 빈도는 높지만 대부분의 문서에 분산되어 있어 문서 특성을 잘 설명하지 못하는 단어를 찾는데 활용된다(설동훈 외, 2020).이 때 가중치 값이 클수록, 주제어들이 차지하는 비율이 높을수록 중요한 단어를 의미한다(이나영, 최정화, 2019).

이 연구에서 각 항목별로 추출된 상위20위 주요 키워드는 빈도크기 순으로 나열하고 TF-IDF 가중치 행렬을 적용하여 TF-IDF 중요도가 .01 이상인 결과만 제시하였다. 또한 빈도분석 결과에서 나타난 시기별 공통/차별 키워드 같은 경우 >부등호를 표시하여 상대적 순위를 나타내고 높은 순위 순으로 나열하였다.

키워드 트렌드 분석을 위해 공통 키워드를 하나의 집단으로 그룹화한 뒤 Hot_Cold 분석을 실시하여 각 항목별 공통 키워드에 대한 관심도 증감 추세를 산출하고 연도별 변화를 살펴보기 위해 키워드 트렌드 추이 그래프를 생성하였다. 공통 키워드에 대한 차이검정을 위해 회귀분석을 실시하였으며 모든 자료처리는 python 3 프로그램을 이용하였다.

2) 토픽 분석

이 연구에서 건강관리 유형별 토픽 구조적 특성을 파악하기 위해 python 3 프로그램의 scikit-learn 패키지를 이용하여 LDA 알고리즘 기반 토픽모델링 분석을 실시하였다. 토픽 모델링 기법의 장점은 문서에 사용된 어휘(term)의 관련성을 기반으로 말뭉치(corpus)에 내포되어 잠재된 토픽을 도출해 준다는 것이다(Blei, 2012; Hofmann, 1999). 하나의 문서는 여러 토픽(주제)로 구성되고 문서가 갖는 주제와 해당 단어의 분포에 따라 결정된다고 가정할 때 단위 문서가 생성되며 이러한 개념을 도식화하면 아래 <그림7>과 같다.

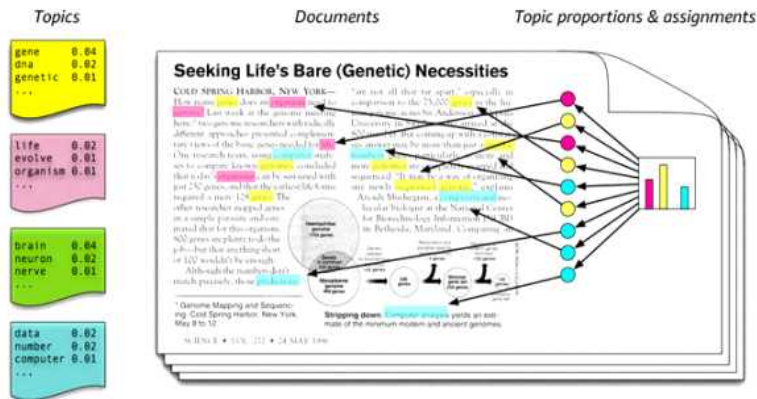


그림7. 잠재 디리클레 할당 분석 모델의 개념

출처: Blei, D.(2012). "Probabilistic topic models". Communications of the ACM, 55(4): 78

토픽 분석을 위해 키워드 간 유사성(Similarity)에 따라 추출되는 결과를 근거로 적정 토픽 수(k)를 7개로 결정하였다. 토픽모델링은 토픽 수에 따라 결과가 달라질 수 있기 때문에 연구자가 해당 분야의 적합성을 고려하여 토픽수를 지정하거나 임의의 수 몇 개를 지정하여 얻은 결과를 비교하여 해석가능성을 기준으로 결정할 수 있다(안주영 외, 2016; 강범일 외, 2013; 배정환 외, 2013). 이 연구에서는 각 토픽을 대표하는 상위20위 키워드 비중 값이 도출된 3개의 토픽을 대상으로 Hot_Cold 분석을 실시하였으며 연도별 토픽 트렌드 추이 그래프를 생성하였다. 또한 토픽에 대한 관심도가 유의미하게 변화되었는지 확인하기 위해 회귀분석을 실시하였다. 이 연구에서 사용된 토픽모델링 생성 코드는 아래 <그림8>과 같다.

```
dictionary = corpora.Dictionary(total_keywords)
corpus = [dictionary.doc2bow(text) for text in total_keywords]
NUM_TOPICS = 7
ldamodel = gensim.models.ldamodel.LdaModel(
    corpus,
    num_topics = NUM_TOPICS,
    id2word=dictionary,
    passes=10)
topics = ldamodel.print_topics(num_words=20)
```

그림8. 토픽모델링 생성 코드

5. 연구의 타당도

이 연구에서는 연구의 설계, 자료수집 및 분석과 해석 등에 대한 타당도를 높이기 위해 개별적인 전문가 회의를 실시하였다. 전문가는 건강교육학을 전공한 박사학위소지자 2명과 S기업, N기업의 데이터분석가 2명으로 구성하여 연구수행 타당성 및 자료수집 범위 설정을 논의하였다. 또한 키워드 및 토픽 분석에서 공통/차별 키워드의 순위 변동을 검증하기 위해 카이스퀘어 검정을 실시하였고 각 토픽 트렌드 추이를 파악하기 위해 Hot_Cold 분석과 회귀분석을 실시하였다.

6. 자료처리

1) 빅카인즈의 분석방법

빅카인즈는 온라인 뉴스기사 비정형 데이터를 분석하는 빅데이터 분석시스템이다. 이 연구에서 이용한 키워드는 빅카인즈 자체 알고리즘에 의해 전처리 과정을 거친 후 명사만 추출된 결과이다.

2) Python 3 프로그램

빅카인즈 뉴스분석시스템을 통해 추출한 키워드는 python 3을 이용하여 연구 디자인에 맞게 빈도분석과 워드클라우드를 생성했고 토픽모델링 기법을 적용하여 토픽을 추출하였다. python 3를 이용한 분석방법별 라이브러리 패키지는 다음 <표6>과 같다.

표6. python 3에서 사용한 라이브러리 패키지

분석방법	패키지
wordcloud	pytagcloud
word tokenizing, stopwords filtering	nltk
LDA visualization	gensim
TF-IDF, LDA	scikit-learn
keyword frequency analysis	pandas, konlpy

III. 연구결과

시기별 건강관리 경향을 파악하기 위해 설정한 분석방법 및 연구내용은 다음과 같다.

첫째, 시기별 건강관리 유형에 따른 주요 키워드 추출이다. 빈도수 분석 기반으로 추출된 상위50위 주요 키워드는 워드클라우드를 이용하여 시각화하였다.

둘째, 시기별 건강관리 유형에 따른 키워드 출현빈도 분석이다. 상위20위 주요 키워드의 빈도(TF), TF-IDF 가중치 값을 산출하여 빈도수가 큰 순서대로 나열하였다.

셋째, 키워드 출현빈도 분석에서 나타난 시기별 공통/차별 키워드 순위 변동을 검정하기 위해 카이스퀘어 검정을 실시하였다.

넷째, 시간흐름에 따라 공통 키워드 트렌드가 유의미하게 증감하였는지 Hot_Cold 분석과 회귀분석을 실시하였다.

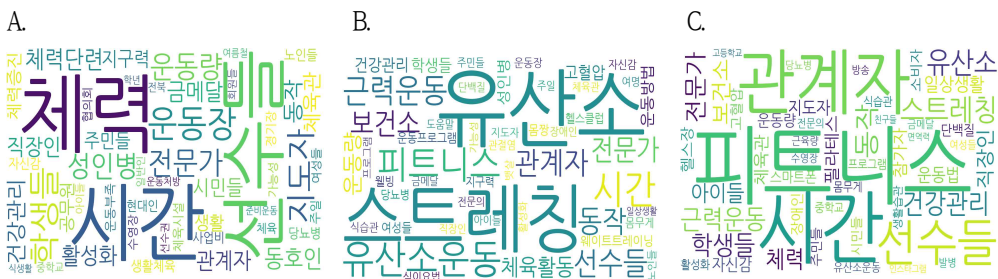
다섯째, 토픽모델링 분석을 통해 토픽 내 키워드의 비중을 확인함으로써 각 토픽마다 어떤 키워드가 주요하게 다뤄지고 있는지 확인하였다.

여섯째, 시간흐름에 따라 각 토픽의 트렌드가 유의미하게 증감하였는지 Hot_Cold 분석과 회귀분석을 실시하였다.

1. 시기별 건강관리 유형에 따른 주요 키워드

1) 운동

시기별 운동 관련 주요 키워드를 워드클라우드 통해 시각화하면 <그림9>와 같다.



A: 1시기_1990년대, B: 2시기_2000년대, C: 3시기_2010년대

그림9. 시기별 운동 워드클라우드

2시기는 ‘유산소’, ‘스트레칭’, ‘피트니스’, ‘시간’ 이 큰 비중을 차지했으며 다이어트 관련 내용이 주를 이루면서 1시기와 다른 양상으로 변화된 결과를 나타냈다. 특히 ‘유산소 운동’, ‘스트레칭’, ‘근력 운동’, ‘웨이트 트레이닝’ 과 같은 운동방법 관련 키워드가 다양하게 도출되었다.

2) 식이

그림10. 시기별 식이 워드클라우드

– 19 –

3시기는 ‘면역력’, ‘항산화’ 키워드가 급부상하였다. 영양소와 관련하여 ‘유산균’, ‘셀레늄’, ‘오메가’ 키워드가 도출되었으며 특히 ‘고함량’, ‘섭취량’, ‘추출물’ 등 영양소 섭취 관련 키워드가 출현하였다. 또한 ‘쇼핑몰’, ‘건강식품’, ‘소비자들’ 과 같이 영양제 제품 관련 키워드가 다양하게 나타났다.

4) 음주

시기별 음주 관련 주요 키워드를 워드클라우드 통해 시각화하면 <그림12>와 같다.



A: 1시기_1990년대, B: 2시기_2000년대, C: 3시기_2010년대

그림12. 시기별 음주 워드클라우드

1시기 2시기 모두 ‘관계자’, ‘가능성’ 키워드 비중이 크게 나타났고 2시기 3시기에서는 ‘절주’가 출현하면서 핵심 키워드로 나타났다. 1시기에서 금주의 의미로 ‘방법’ 키워드가 나타났으며 음주로 인한 문제점 ‘발병’, ‘간염’, ‘중독자’, ‘지방간’, ‘부작용’ 키워드가 도출되었다.

2시기는 ‘술자리’ 비중이 높아졌고 이와 관련된 ‘음주문화’, ‘폭탄주’ 키워드가 도출되었다. 또한 ‘보건소’ 키워드가 출현하면서 ‘예방’, ‘도움말’ 키워드가 함께 도출되었으며 ‘금연’, ‘담배’와 같이 흡연 관련 키워드도 출현하였다.

3시기는 ‘절주’가 압도적으로 큰 비중을 차지하였고 음주예방 관련 내용으로 ‘예방’, ‘생활습관’, ‘식습관’, ‘식생활’, ‘식이요법’, ‘건강관리’, ‘건강검진’ 키워드가 도출되었다. 또한 ‘전문가’, ‘전문’의와 관련하여 ‘보건복지부’, ‘관계자’와 같이 보건 의료 관련 키워드도 함께 나타났다. 음주 관련 키워드에서 ‘간질환’, ‘간경화’, ‘합병증’, ‘고혈압’ 등 질병 관련 키워드는 전체시기에 걸쳐 도출되었다.

5) 흡연

시기별 흡연 관련 주요 키워드를 워드클라우드 통해 시각화하면 <그림13>과 같다.



A: 1시기_1990년대, B: 2시기_2000년대, C: 3시기_2010년대

그림13. 시기별 흡연 워드클라우드

1시기는 ‘담배’가 높은 비중을 차지하였고 동일 키워드로 ‘흡연’이 추출되었다. ‘흡연자’의 경우 전체 시기에서 큰 비중을 차지하였는데 흡연자와 관련하여 ‘애연가’가 도출되었다. 또한 ‘금연운동’과 관련하여 ‘국민건강증진법’, ‘한국금연운동협의회’, ‘보건복지부’, ‘보사부’가 출현하였다. 이 외에 ‘공공장소’, ‘담배연기’, ‘사무실’, ‘구역’과 같이 ‘금연장소’ 관련 키워드가 도출되었다.

2시기는 ‘흡연자’를 비롯하여 ‘보건의’ 키워드의 비중이 크게 나타났다. 특히 ‘교육’과 관련하여 ‘금연교육’, ‘흡연예방’, ‘금연클리닉’, ‘예방’이 도출되었으며 ‘청소년들’, ‘고등학교’, ‘아이들’과 같이 학생을 대상으로 하는 키워드가 다양하게 나타났다. 한편 ‘고지혈증’, ‘호흡기’, ‘당뇨병’, ‘고혈압’, ‘심혈관’, ‘심장병’ 등 음주 워드클라우드에 나타났던 질병 관련 키워드가 다수 도출되었다.

3시기 역시 ‘흡연자’ 즉 흡연을 하는 대상 관련 키워드가 다양하게 출현하였고 ‘보건의’ 비중이 점점 크게 나타났다. 3시기의 특이점은 ‘전자담배’의 출현이다. 또한 ‘간접흡연’, ‘금연구역’에 대한 키워드 비중이 높아지면서 ‘과태료’, ‘부스’, ‘화장실’, ‘구역’ 등 폭넓은 키워드가 출현하였으며 ‘마약류’, ‘유해성’, ‘위험성’, ‘우울증’과 같이 흡연에 대한 부정적 키워드 비중이 커지고 있다. 이에 따라 음주와 흡연은 건강관리 요소 중 부정적 행위에 해당하므로 질병, 정책 관련 내용을 다수 포함하는 것으로 나타났다.

6) 휴식

시기별 휴식 관련 주요 키워드를 워드클라우드 통해 시각화하면 <그림14>와 같다.



A: 1시기_1990년대, B: 2시기_2000년대, C: 3시기_2010년대

그림14. 시기별 휴식 워드클라우드

휴식 관련 키워드는 전체시기에서 ‘재충전’과 ‘시간’ 키워드가 큰 비중을 차지하였다. 1시기는 ‘피로회복’, ‘피서지’가 주된 키워드로 나타났으며 ‘휴가’, ‘여름휴가’, ‘휴가철’과 같이 휴가 관련 키워드가 다수 출현했다. 또한 ‘주일’, ‘여름철’, ‘관광지’, ‘고속도로’, ‘주차장’ 등 휴가를 떠나는 내용과 관련된 키워드가 함께 도출되었다. 특성 있는 키워드로는 ‘플레이오프’, ‘하나님’, ‘자신감’, ‘혈액순환’, ‘생산성’ 등이 나타났다.

2시기는 ‘수영장’을 포함하여 ‘물놀이’, ‘온천수’, ‘온천욕’, ‘해수욕장’, ‘위더파크’, ‘파도풀’과 같이 물놀이 관련 키워드가 다수 출현하였다. 이 외에도 ‘골프장’이나 ‘강원도’, ‘충남’ 등 국내 지역 관련 키워드가 출현하였고 ‘태국’, ‘아시아’, ‘프랑스’와 같이 해외 여행지 관련 키워드도 함께 도출되었다.

3시기에서는 ‘힐링’이 새롭게 출현함과 동시에 가장 큰 비중을 차지하였다. 이는 스트레스 3시기에 나타나는 주요 키워드와 일치된 내용임을 확인하였다. 핵심 키워드인 ‘힐링’과 관련하여 ‘산책로’, ‘쉼터’, ‘휴식공간’이 주요 키워드로 도출되었다. 이 외에도 ‘캠핑장’, ‘자연환경’, ‘휴양림’과 같이 자연친화적 키워드가 출현하였다. 또한 3시기에서 ‘관광객’의 비중이 급부상하였는데 이와 관련하여 ‘관광지’, ‘여행지’, ‘방문객’, ‘이용객’ 키워드가 함께 도출되었다. 특히 ‘도서관’, ‘전망대’, ‘편의시설’과 같이 이전 시기와 차별화된 키워드가 다양하게 출현하였다.

시기별 수면 관련 주요 키워드를 워드클라우드 통해 시각화하면 <그림15>와 같다.



A: 1시기_1990년대, B: 2시기_2000년대, C: 3시기_2010년대
그림15. 시기별 수면 워드클라우드

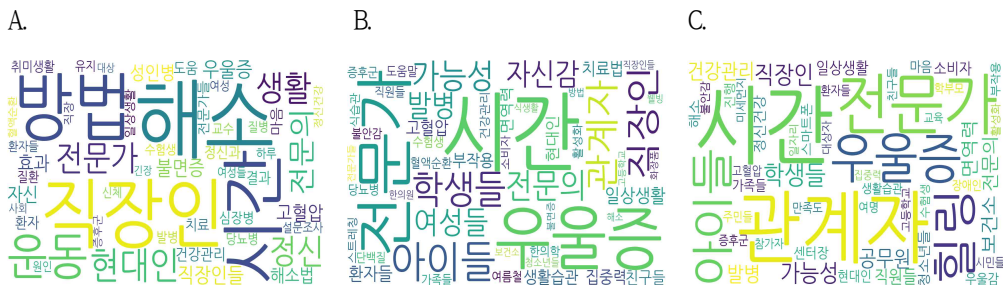
수면 전체시기에서 주로 ‘시간’, ‘불면증’ 이 큰 비중을 차지하였다. 시간은 수면시간과 연관 지어 ‘수면부족’, ‘다음날’, ‘하루’ 키워드가 함께 도출되었다. 1시기는 공통된 키워드 이외에 ‘수면제’, ‘수면부족’, ‘수면장애’가 주요 키워드로 나타났는데 수면장애 또는 불면증 치료와 관련된 ‘치료법’, ‘서울대병원’, ‘정신과’, ‘증후군’이 함께 출현하였다. 또한 도움을 주는 대상으로 ‘전문가’, ‘전문직’, ‘교수’ 키워드가 도출되었다.

2시기는 1시기와 비슷한 내용으로 구성되어 있지만 ‘숙면’ 키워드의 비중이 눈에 띄게 증가하였다. ‘수면중’, ‘코골’, ‘무호흡증’, ‘수면무호흡증’과 같이 수면 중 나타나는 증상 관련 키워드가 다수 도출되었다. 또한 ‘숙면’과 ‘중요’ 키워드를 통해 알 수 있듯이 ‘집중력’, ‘기억력’, ‘생체리듬’ 등 숙면의 중요성과 관련된 키워드가 부각되었다. 특히 ‘수면다원검사’, ‘멜라토닌’, ‘뇌파’와 같이 수면 관련 정보를 포함하고 있다.

3시기에서는 ‘관계자’ 가장 큰 비중을 차지하였으며 ‘가능성’에 이어 ‘전문가’의 비중이 큰 것으로 나타났다. 특히 ‘스마트폰’이 주요 키워드로 급부상하면서 ‘소비자’, ‘사용자’ 키워드가 함께 도출되었다. 또한 일상 속 수면관리 관련 키워드로 ‘건강관리’, ‘생활습관’, ‘일상생활’, ‘활성화’ 키워드가 출현하였다. 이 외에 ‘프로포폴’, ‘유튜브’, ‘박송’ 등 다른 시기와 달리 특정 있는 키워드가 다수 출현하였다.

8) 스트레스

시기별 스트레스 관련 주요 키워드를 워드클라우드 통해 시각화하면 <그림16>과 같다.



A: 1시기_1990년대, B: 2시기_2000년대, C: 3시기_2010년대

그림16. 시기별 스트레스 워드클라우드

1시기는 ‘해소’가 가장 큰 비중을 차지했으며 ‘직장인’, ‘방법’, ‘시간’도 비슷한 수준의 주요 키워드로 나타났다. 스트레스 해소와 관련하여 ‘운동’이 핵심 키워드로 도출되었으며 ‘치료’, ‘해소법’, ‘건강관리’, ‘정신건강’, ‘마음’, ‘정신’, ‘취미생활’ 등이 있다. 스트레스 대상은 직장인뿐 아니라 ‘현대인’, ‘환자들’, ‘수험생’, ‘여성들’로 다양하게 도출되었다.

2시기, 3시기는 ‘시간’이 가장 큰 비중을 차지하였으며 다음으로 ‘전문가’, ‘우울증’이 주요 키워드로 나타났다. 1시기와 마찬가지로 다양한 대상자 관련 키워드가 도출되었는데 ‘현대인’보다 ‘학생들’, ‘여성들’, ‘직장인’의 비중이 커졌음을 확인하였다. 특히 ‘웰빙’, ‘건강관리’, ‘생활습관’과 같이 건강관리 관련 키워드가 도출되었으며 이 외에도 ‘화장품’, ‘한의원’, ‘단백질’, ‘스트레스’ 등과 같이 특성 있는 키워드가 다양하게 출현하였다.

3시기는 ‘시간’이 외에도 ‘관계자’, ‘힐링’ 키워드가 큰 비중을 차지하였다. 또한 이전 시기보다 ‘아이들’, ‘우울증’, ‘보건소’ 비중이 증가하였고 ‘미세먼지’, ‘스마트폰’, ‘센터장’, ‘만족도’와 같이 다른 시기와 구분된 키워드가 다수 출현하였다. 스트레스는 전체시기에 걸쳐 ‘공무원’, ‘직원들’ 등 대상자 관련 키워드가 꾸준히 나타나고 있다.

2. 시기별 건강관리 유형에 따른 키워드 출현빈도

시기별 건강관리 유형에 따른 주요 키워드를 단어빈도(TF), 단어빈도-역문서빈도(TF-IDF) 기준으로 산출한 결과는 다음 <표7>과 같다. 이 연구의 결과에서 역문서빈도(IDF) 값은 0이 나오지 않게 최소 0.1로 조절하여 그 이상의 값을 제시하였다.

표7. 시기별 운동 상위20 키워드 출현빈도

순위	1시기			2시기			3시기		
	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF
1	체력	478	.24	유산소	1335	.32	피트니스	2664	.34
2	시간	405	.26	유산소운동	947	.23	시간	2046	.26
3	선수들	316	.18	스트레칭	930	.22	관계자	1613	.20
4	운동장	280	.14	피트니스	791	.34	선수들	1435	.18
5	학생들	266	.13	시간	790	.26	유산소	1229	.32
6	지도자	263	.13	근력운동	740	.18	스트레칭	1212	.22
7	전문가	261	.13	동작	624	.15	체력	1129	.24
8	운동량	222	.11	선수들	590	.18	동작	1098	.15
9	동호인	218	.11	관계자	573	.20	건강관리	1070	.13
10	성인병	216	.11	보건소	572	.14	근력운동	1052	.18
11	체력단련	212	.11	전문가	558	.13	학생들	1006	.13
12	건강관리	210	.13	고혈압	518	.12	전문가	949	.13
13	체육	209	.10	체육활동	451	.11	보건소	928	.14
14	동작	207	.15	운동량	427	.11	프로그램	742	-
15	직장인	206	.10	운동방법	422	.10	자신감	687	-
16	금메달	199	-	건강관리	421	.13	직장인	679	.10
17	관계자	198	.20	성인병	413	.11	아이들	660	-
18	활성화	195	-	학생들	408	.13	참가자	637	-
19	지구력	191	-	당뇨병	394	-	단백질	606	-
20	생활	191	-	웨이트 트레이닝	385	-	스마트폰	599	-

*공통 키워드는 강조 표시함

운동 상위 빈출 20개 키워드 빈도분석 결과 TF 기준과 TF-IDF 기준에 따른 순위 차이는 있었지만 <표 7>과 같이 운동 1시기에 ‘시간’, ‘체력’, ‘관계자’, 2시기 ‘피트니스’, ‘유산소’, ‘시간’, 3시기 ‘피트니스’, ‘유산소’, ‘시간’ 등의 키워드가 중요하게 다뤄졌음을 확인하였다.

표8. 시기별 식이 상위20 키워드 출현빈도

순위	1시기			2시기			3시기		
	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF
1	식이요법	631	.62	식이요법	1865	.62	식이요법	3600	.62
2	당뇨병	270	.26	당뇨병	653	.26	단백질	1854	.22
3	요법	167	.16	고혈압	635	.20	식이섬유	1751	.21
4	성인병	163	.16	단백질	482	.22	영양소	1426	.17
5	단백질	157	.22	요법	472	.16	고혈압	1353	.20
6	치료법	148	.14	식이섬유	460	.21	당뇨병	1350	.26
7	합병증	147	.14	운동	407	.13	전문가	1179	.14
8	식생활	136	.13	합병증	361	.14	식이조절	1126	.13
9	고혈압	132	.20	식생활	348	.13	식습관	1101	.13
10	운동요법	131	.13	부작용	348	.11	다이어트	1042	.12
11	영양소	101	.17	전문가	338	.14	운동	995	.13
12	발병	100	-	성인병	318	.16	관계자	957	.11
13	환자들	92	-	영양소	316	.17	시간	923	.11
14	부작용	91	.11	시간	312	.11	합병증	864	.14
15	전문의	89	-	고지혈증	304	-	비타민	811	-
16	음식물	81	-	운동요법	304	.13	요법	808	.16
17	전문가	81	.14	보건소	304	-	발병	790	-
18	소비자	81	-	전문의	294	-	섬유소	779	-
19	약물요법	78	-	발병	292	-	소비자	777	-
20	심장병	70	-	치료법	283	.14	부작용	771	.11

*공통 키워드는 강조 표시함

식이 상위 빈출 20개 키워드 빈도분석 결과 TF 기준과 TF-IDF 기준에 따른 순위 차이는 있었지만 <표 8>과 같이 식이 전체시기에서 ‘식이요법’, ‘당뇨병’, ‘단백질’ 키워드를 중점으로 1시기에 ‘고혈압’, 2시기와 3시기는 ‘식이섭유’ 등의 키워드가 중요하게 다뤄졌음을 확인하였다.

표9. 시기별 영양제 상위20 키워드 출현빈도

순위	1시기			2시기			3시기		
	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF
1	영양	47	.28	비타민	586	.35	비타민	2084	.35
2	비타민	35	.35	의약품	483	.26	영양소	1490	.25
3	의약품	31	.26	관계자	378	.24	관계자	1407	.24
4	관계자	24	.24	소비자	358	.20	소비자	1172	.20
5	부작용	23	.14	영양	335	.28	의약품	1090	.26
6	단백질	23	.14	영양소	308	.25	면역력	976	.16
7	영양소	22	.25	건강식품	263	.14	오메가	938	.20
8	전문가	20	.12	부작용	251	.14	단백질	726	.14
9	소비자	18	.20	단백질	246	.14	임산부	725	.12
10	항생제	17	.10	화장품	234	.12	영양	718	.28
11	건강보조식품	15	-	치료제	205	.11	항산화	677	.11
12	주사	15	-	아이들	204	.11	건강기능식품	672	.15
13	가능성	13	-	항산화	202	.11	건강관리	659	.11
14	성인병	13	-	제품	200	.11	전문가	634	.12
15	비타민제	13	-	소비자들	180	-	아이들	615	.11
16	식생활	11	-	웰빙	180	.12	부작용	603	.14
17	임산부	11	.12	비타민제	171	-	시간	543	-
18	영양상태	11	-	시간	164	-	건강식품	501	.14
19	음식물	11	-	건강관리	164	.11	현대인	493	-
20	건강식품	11	.14	보건소	159	-	쇼핑몰	482	.10

*공통 키워드는 강조 표시함

영양제 상위 빈출 20개 키워드 빈도분석 결과 TF 기준과 TF-IDF 기준에 따른 순위 차이는 있었지만 <표 9>와 같이 영양제 전체시기에서 ‘비타민’, ‘영양’, ‘의약품’ 키워드를 중점으로 1시기에 ‘영양소’, 2시기 ‘관계자’, 3시기 ‘영양소’ 등의 키워드가 중요하게 다뤄졌음을 확인하였다.

표10. 시기별 음주 상위20 키워드 출현빈도

순위	1시기			2시기			3시기		
	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF
1	관계자	99	.24	관계자	340	.24	절주	578	.19
2	가능성	82	.20	가능성	296	.20	관계자	541	.24
3	방법	70	.17	절주	261	.19	음주	531	.18
4	전문가	52	.14	음주	242	.18	술자리	502	.17
5	음주	44	.18	술자리	236	.17	시간	500	.17
6	시간	38	.17	시간	233	.17	고혈압	478	.16
7	간염	36	.16	간염	225	.16	간염	469	.16
8	금주중	33	-	전문가	213	.14	지방간	459	.15
9	지방간	30	.15	고혈압	200	.16	가능성	447	.20
10	중독자	30	-	당뇨병	198	.13	발병	447	.15
11	공무원	26	-	지방간	195	.15	생활습관	408	.14
12	문제점	26	-	금연	186	.12	전문가	405	.14
13	학생들	25	-	방법	179	.17	식습관	336	.11
14	치료	25	-	직장인	167	.11	당뇨병	335	.13
15	주일	24	-	발병	159	.15	전문의	320	.11
16	운동	24	-	간질환	145	-	합병증	305	.10
17	치료법	23	-	보건소	144	-	직장인	290	.11
18	발병	23	.15	전문의	141	.11	금연	271	.12
19	기업들	23	-	청소년들	139	-	간암	265	-
20	환자들	22	-	음주문화	119	-	고지혈증	262	-

*공통 키워드는 강조 표시함

음주 상위 빈출 20개 키워드 빈도분석 결과 TF 기준과 TF-IDF 기준에 따른 순위 차이는 있었지만 <표 10>과 같이 음주 전체시기에서 ‘관계자’, ‘가능성’ 키워드를 중점으로 1시기에 ‘음주’, 2시기와 3시기는 ‘절주’ 등의 키워드가 중요하게 다뤄졌음을 확인하였다.

표11. 시기별 흡연 상위20 키워드 출현빈도

순위	1시기			2시기			3시기		
	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF
1	담배	1171	.42	흡연	4614	.36	흡연자	4758	.36
2	흡연	984	.36	담배	4525	.42	보건소	4033	.31
3	흡연자	852	.36	흡연자	4338	.36	금연구역	3354	.26
4	금연운동	447	.16	보건소	3358	.31	관계자	3087	.24
5	증진법	340	.12	관계자	2866	.24	흡연	3032	.36
6	애연가	337	.12	금연구역	1673	.26	간접흡연	2879	.22
7	금연구역	325	.26	흡연율	1665	.13	과태료	2412	.18
8	사무실	297	.11	청소년	1369	.11	담배	2273	.42
9	범칙금	291	.11	학생들	1324	.10	보건복지부	1486	.11
10	흡연율	264	.13	대상	1322	.10	건강관리	1365	.10
11	비흡연자	263	-	교육	1284	-	증진법	1350	.12
12	관계자	262	.24	청소년들	1216	-	전자담배	1346	.10
13	운동	249	-	흡연자들	1208	-	흡연율	1331	.13
14	백만	246	-	금연운동	1207	.16	절주	1278	-
15	담배연기	239	-	담뱃값	1155	-	흡연자들	1256	-
16	청소년들	233	-	보건복지부	1140	.11	학생들	1201	.10
17	공공장소	231	-	간접흡연	1131	.22	담뱃값	1183	-
18	장소	229	-	담배연기	1098	-	흡연예방	1165	-
19	금연장소	215	-	금연교육	1069	-	고혈압	1076	-
20	애연가들	213	-	운동	1028	-	구역	1057	-

*공통 키워드는 강조 표시함

흡연 상위 빈출 20개 키워드 빈도분석 결과 TF 기준과 TF-IDF 기준에 따른 순위 차이는 있었지만 <표 11>과 같이 흡연 전체시기에서 ‘담배’, ‘흡연’, ‘흡연자’ 키워드를 중점으로 1시기에 ‘금연구역’, 2시기와 3시기는 ‘보건소’ 등의 키워드가 중요하게 다뤄졌음을 확인하였다.

표12. 시기별 휴식 상위20 키워드 출현빈도

순위	1시기			2시기			3시기		
	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF
1	재충전	343	.69	재충전	1402	.69	힐링	8214	.72
2	시간	140	.34	시간	751	.34	시간	2782	.34
3	피로회복	62	.12	수영장	259	.12	재충전	2212	.69
4	피서지	43	-	피로회복	237	.12	관계자	1924	.17
5	여름휴가	43	-	직장인	230	.10	관광객	1594	.14
6	휴가	42	-	웰빙	220	.13	산책로	967	-
7	직장인	37	.10	관계자	216	.17	시민들	950	-
8	전문가	36	-	관광객	211	.14	관광지	938	-
9	여름철	35	-	아이들	201	-	수영장	918	.12
10	휴가철	33	-	선수들	186	-	아이들	914	-
11	근로자	33	-	휴양지	178	-	직장인	770	.10
12	관계자	32	.17	물놀이	175	-	쉽터	748	-
13	주일	30	-	직원들	157	-	전문가	728	-
14	관광지	26	-	여름휴가	153	-	사업비	701	-
15	피로	26	-	온천수	145	-	여행지	673	-
16	고속도로	26	-	가족들	145	-	주민들	639	-
17	건강관리	24	-	스트레칭	143	-	방문객	617	-
18	해수욕장	24	-	전문가	141	-	직원들	601	-
19	휴식시간	24	-	삼성	136	-	휴식공간	597	-
20	자녀들	24	-	휴가철	135	-	가족들	562	-

*공통 키워드는 강조 표시함

휴식 상위 빈출 20개 키워드 빈도분석 결과 TF 기준과 TF-IDF 기준에 따른 순위 차이는 있었지만 <표 12>와 같이 휴식 1시기와 2시기의 ‘재충전’, ‘시간’ 을 중점으로 1시기에 ‘관계자’, 2시기 ‘관광객’ 키워드가 중요하게 나타난 반면 3시기는 ‘힐링’ 키워드가 압도적인 비중(72%)을 차지하며 핵심 키워드로 나타났다.

표13. 시기별 수면 상위20 키워드 출현빈도

순위	1시기			2시기			3시기		
	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF
1	시간	326	.50	시간	415	.50	관계자	2205	.33
2	불면증	192	.32	불면증	266	.32	시간	1699	.50
3	수면시간	109	.15	숙면	195	-	가능성	1246	.19
4	수면제	107	.15	수면장애	138	.17	소비자	913	.14
5	수면장애	104	.17	수면시간	121	.15	전문가	874	.13
6	수면부족	94	.13	집중력	109	.13	스마트폰	656	.17
7	전문의	91	.12	호흡증	108	.13	불면증	563	.32
8	전문가	84	.13	코골	93	.11	우울증	513	-
9	운동	78	.11	수면제	90	.15	사용자	489	-
10	수면증	73	-	수면무	87	.11	한국인	425	-
11	집중력	69	.13	수면무호흡증	86	.10	연구원	403	-
12	결과	68	-	기억력	80	-	아이들	401	-
13	뇌파	68	-	다음날	79	-	수면장애	382	.17
14	교수	66	-	전문가	79	.13	프로포폴	376	-
15	상태	64	-	우울증	77	-	의약품	368	-
16	방법	64	-	전문의	75	.12	삼성전자	362	-
17	숙면	63	.24	고혈압	68	-	전문의	352	.12
18	우울증	61	-	열대야	66	-	부작용	349	-
19	호흡증	61	.13	연구팀	65	-	시청자	342	-
20	정신과	60	-	수면부족	64	.13	발병	322	-

*공통 키워드는 강조 표시함

수면 상위 빈출 20개 키워드 빈도분석 결과 TF 기준과 TF-IDF 기준에 따른 순위 차이는 있었지만 <표 13>과 같이 수면 전체시기에서 ‘시간’, ‘불면증’ 키워드를 중점으로 1시기 ‘숙면’, 2시기 ‘수면장애’, 3시기 ‘관계자’ 등의 키워드가 중요하게 다뤄졌음을 확인하였다.

표14. 시기별 스트레스 상위 20 키워드 출현빈도

순위	1시기			2시기			3시기		
	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF	키워드	빈도	TF-IDF
1	해소	198	.28	시간	2248	.40	시간	3191	.40
2	직장인	162	.23	전문가	1393	.18	관계자	1796	.22
3	방법	145	.21	우울증	1338	.18	전문가	1344	.18
4	시간	139	.40	아이들	1293	.17	우울증	1243	.18
5	현대인	125	.18	직장인	1198	.23	힐링	1129	.18
6	운동	115	.17	관계자	1080	.22	아이들	1008	.17
7	전문가	102	.18	학생들	1074	.14	일들	969	.12
8	생활	96	.14	전문의	1017	.13	직장인	895	.23
9	정신	94	.14	가능성	988	.13	학생들	870	.14
10	직장인들	86	.12	자신감	842	.11	가능성	844	.13
11	전문의	79	.13	여성들	830	.11	프로그램	806	.10
12	우울증	74	.18	발병	824	.11	친구들	788	-
13	불면증	70	.10	고혈압	783	.10	자신감	727	.11
14	효과	67	-	일상생활	736	-	공무원	718	-
15	고혈압	66	.10	현대인	735	.18	면역력	646	-
16	자신	65	-	부작용	728	-	건강관리	616	-
17	성인병	64	-	환자들	703	-	보건소	612	-
18	해소법	61	-	집중력	694	-	발병	601	.11
19	건강관리	58	-	치료법	679	-	소비자	588	-
20	전문가들	57	-	면역력	658	-	가족들	563	-

*공통 키워드는 강조 표시함

스트레스 상위 빈출 20개 키워드 빈도분석 결과 TF 기준과 TF-IDF 기준에 따른 순위 차이는 있었지만 <표 14>와 같이 스트레스 전체시기에서 ‘시간’ 키워드를 중점으로 1시기 ‘해소’, ‘직장인’, 2시기와 3시기는 ‘직장인’, ‘관계자’ 등의 키워드가 중요하게 다뤄졌음을 확인하였다.

이 연구는 뉴스 빅데이터에 나타난 건강관리 경향을 파악하기 위해 키워드 빈도 분석을 실시하여 시기별 상위20 주요 키워드를 추출하고 TF, TF-IDF 두 가지 방식을 적용하여 키워드 간 순위를 비교하였다. 이러한 시도는 키워드의 순위 파악에 앞서 키워드 도출방식의 적합성을 확인하고자 했으며 뉴스기사와 같은 비정형 문서에서 특정 단어가 빈번하게 나타나는 경우를 고려한 것이다. 이와 같이 TF, TF-IDF 두 가지 방식을 비교해보면 도출된 주요 키워드 구성은 유사하지만 부분적으로 순위가 다르게 나타났다.

운동 1시기에서 TF기준 ‘체력’, ‘시간’, ‘선수들’ 순으로 순위가 높게 나타난 반면 TF-IDF기준 ‘시간’, ‘체력’, ‘관계자’ 순으로 큰 비중을 차지하였다. 영양제 3시기는 TF 기준 ‘비타민’, ‘영양소’, ‘관계자’ 순으로 순위가 높게 나타난 반면 TF-IDF기준 ‘비타민’, ‘영양’, ‘의약품’ 순으로 큰 비중을 차지하였다. 음주 2시기는 TF, TF-IDF 두 기준에서 ‘관계자’, ‘가능성’, ‘절주’ 순으로 일치하는 결과가 나타났다. 한편 음주 2시기에 ‘방법’ 이 TF기준 13위에서 TF-IDF기준 공동5위로 나타났으며 3시기 ‘가능성’ 또한 9위에서 2위로 순위가 높아지는 것을 확인하였다. 휴식 1시기에서 TF-IDF기준 ‘관계자’ 는 12위에서 3위로 변화하였으며 특히 수면 1시기 ‘숙면’ 은 17위에서 3위로 순위 변화가 큰 것으로 나타났다. 스트레스 2시기에서도 TF-IDF기준 ‘현대인’ 이 15위에서 공동2위로 부상하면서 TF, TF-IDF 두 가지 방식의 결과 차이를 보여주고 있다.

3. 시기별 건강관리 유형에 따른 공통·차별 키워드

상위20위 주요 키워드의 빈도분석 결과를 바탕으로 시기별 공통/차별 키워드를 구분하여 회귀분석을 실시한 내용은 <표15>와 같다.

표15. 시기별 운동 공통/차별 키워드

구분	1시기	2시기	3시기
공통 키워드	시간>선수들>학생들>전문가>건강관리>동작>관계자	시간>동작>선수들>관계자>전문가>건강관리>학생들	시간>관계자>선수들>동작>건강관리>학생들>전문가
차별 키워드	체력>운동장>지도자>운동량>동호인>성인병>체력단련>체육>직장인>금메달>활성화>지구력>생활	유산소>유산소운동>스트레칭>피트니스>근력운동>보건소>고혈압>체육활동>운동량>운동방법>성인병>당뇨병>웨이트트레이닝	피트니스>유산소>스트레칭>체력>근력운동>보건소>프로그램>자신감>직장인>아이들>참가자>단백질>스마트폰

$$\chi^2 = 3442, df = 44, p < .001$$

*부등호는 키워드의 상대적 순위를 의미함

시기별 운동 관련 공통 키워드는 총7개이고 나머지 13개는 차별 키워드로 나타났다. <표15>와 같이 운동에서 ‘시간’ 키워드가 공통 키워드로 출현하면서 전체시기에 걸쳐 가장 높은 비중을 차지하였다. 또한 ‘선수’, ‘건강관리’ 관련 내용이 꾸준히 등장하면서 시기별로 이와 관련된 운동유형, 목적 등 다양한 차별 키워드가 도출되었다. 시기별 공통/차별 키워드 순위에 유의미한 변화가 있는지 알아보기 위해 카이검정을 실시한 결과 $p < .001$ 유의확률 수준에서 차이가 있는 것으로 나타났다.

표16. 시기별 식이 공통/차별 키워드

구분	1시기	2시기	3시기
공통 키워드	식이요법>당뇨병>요법>단백질>합병증>고혈압>영양소>발병>부작용>전문가	식이요법>당뇨병>고혈압>단백질>요법>합병증>부작용>전문가>영양소>발병	식이요법>단백질>영양소>고혈압>당뇨병>전문가>합병증>요법>발병>부작용

차별 키워드	성인병>치료법>식생활 >운동요법>환자들>전 문의>음식물>소비자> 약물요법>심장병	식이섭유>운동>식생활> 성인병>시간>고지혈증> 운동요법>보건소>전문 의>치료법	식이섭유>식이조절>식습 관>다이어트>운동>관계 자>시간>비타민>섭유소 >소비자
-------------------	---	--	--

$$\chi^2 = 1221, df = 33, p < .001$$

*부등호는 키워드의 상대적 순위를 의미함

시기별 식이 관련 공통 키워드는 총10개이고 나머지 10개는 차별 키워드로 나타났다. <표 16>과 같이 키워드 빈도분석에서 알 수 있듯 식이에서는 ‘식이요법’ 키워드가 전체시기에 걸쳐 압도적인 비중을 차지하였다. 1시기와 2시기의 경우 ‘식이요법’ > ‘당뇨병’ 순위가 동일하게 나타났으며 식이요법을 중심으로 ‘당뇨병’, ‘고혈압’, ‘합병증’ 키워드가 공통 키워드로 분류되었다. 또한 시기 변동에 따라 ‘단백질’, ‘전문가’ 키워드의 상대적 순위가 증가하였다. 차별 키워드에서는 ‘식이섭유’, ‘운동’ 키워드가 상위 순위에 머물렀으며 2시기와 3시기에서 ‘식이섭유가’ 공통된 우선순위로 나타났다. 시기별 공통/차별 키워드 순위에 유의미한 변화가 있는지 알아보기 위해 카이검정을 실시한 결과 $p < .001$ 유의 확률 수준에서 차이가 있는 것으로 나타났다.

표17. 시기별 영양제 공통/차별 키워드

구분	1시기	2시기	3시기
공통 키워드	영양>비타민>의약품>관계자>부작용>단백질>영양소>소비자>건강식품	비타민>의약품>관계자>소비자>영양>영양소>건강식품>부작용>단백질	비타민>영양소>관계자>소비자>의약품>단백질>영양>부작용>건강식품
차별 키워드	전문가>항생제>건강보조식품>주사>가능성>성인병>비타민제>식생활>임산부>영양상태>음식물	화장품>치료제>아이들>항산화>제품>소비자들>웰빙>비타민제>시간>건강관리>보건소	면역력>오메가>임산부>항산화>건강기능식품>건강관리>전문가>아이들>시간>현대인>쇼핑몰

$$\chi^2 = 1033, df = 18, p < .001$$

*부등호는 키워드의 상대적 순위를 의미함

시기별 영양제 관련 공통 키워드는 총9개이고 나머지 11개는 차별 키워드로 나타났다. <표 17>과 같이 영양제 공통 키워드에서는 ‘비타민’ 이 상위 순위에 유지되었고 1시기에서 영양소가 아닌 ‘영양’ 이 큰 비중을 차지하였다. 2시기는 ‘영양’ > ‘영양소가’ 나열된 순위로 나타났으며 2시기와 3시기에서는 ‘비타민’ 이 공통된 우선순위로 나타났다. 특히 차별 키워드 2시기에서 ‘화장품’, ‘웰빙’, 3시기의 ‘임산부’, ‘쇼핑몰’ 과 같이 시기 간 상이한 키워드가 특성 있게 도출되었다. 시기별 공통/차별 키워드 순위에 유의미한 변화가 있는지 알아보기 위해 카이검정을 실시한 결과 $p < .001$ 유의확률 수준에서 차이가 있는 것으로 나타났다.

표18. 시기별 음주 공통/차별 키워드

구분	1시기	2시기	3시기
공통 키워드	관계자>가능성>전문가> 음주>시간>간염>지방간 >발병	관계자>가능성>음주>시 간>간염>전문가>지방간 >발병	관계자>음주>시간>간염 >지방간>가능성>발병> 전문가
차별 키워드	방법>금주중>중독자>공 무원>문제점>학생들>치 료>주일>운동>치료법> 기업들>환자들	절주>술자리>고혈압>당 뇨병>금연>방법>직장인 >간질환>보건소>전문 >청소년들>음주문화	절주>술자리>고혈압>생 활습관>식습관>당뇨병> 전문의>합병증>직장인> 금연>간암>고지혈증

$$\chi^2 = 685, df = 6, p < .001$$

*부등호는 키워드의 상대적 순위를 의미함

시기별 음주 관련 공통 키워드는 총8개이고 나머지 12개는 차별 키워드로 나타났다. <표 18>과 같이 음주의 공통 키워드는 전체시기에서 ‘관계자’ 가 우선순위로 나타났고 1시기와 2시기에서 ‘관계자’ > ‘가능성’ 순위가 동일하게 나타났다. 차별 키워드는 2시기와 3시기에서 ‘절주’ > ‘술자리’ > ‘고혈압’ 순위까지 동일하게 나타났다. 시기별 공통/차별 키워드 순위에 유의미한 변화가 있는지 알아보기 위해 카이검정을 실시한 결과 $p < .001$ 유의확률 수준에서 차이가 있는 것으로 나타났다.

표19. 시기별 흡연 공통/차별 키워드

구분	1시기	2시기	3시기
공통 키워드	담배>흡연>흡연자>금연 구역>흡연율>관계자	흡연>담배>흡연자>관 계자>금연구역>흡연율	흡연자>금연구역>관계 자>흡연>담배>흡연율
차별 키워드	금연운동>증진법>애연가 >사무실>법칙금>비흡연 자>운동>백만>담배연기 >청소년들>공공장소>장 소>금연장소>애연가들	보건소>청소년>학생들 >대상>교육>청소년들> 흡연자들>금연운동>담 뱃값>보건복지부>간접 흡연>담배연기>금연교 육>운동	보건소>간접흡연>과태 료>보건복지부>건강관 리>증진법>전자담배>절 주>흡연자들>학생들>담 뱃값>흡연예방>고혈압> 구역

$$\chi^2 = 4119, df = 15, p < .001$$

*부등호는 키워드의 상대적 순위를 의미함

시기별 흡연 관련 공통 키워드는 총6개이고 나머지 14개는 차별 키워드로 나타났다. <표 19>와 같이 흡연 공통 키워드의 내용은 ‘담배’, ‘흡연’ 관련 내용으로 대체로 비슷하게 나타났는데 시기별 일치된 순위를 보이지는 않았다. 차별 키워드 2시기와 3시기에서 ‘보건소’가 우선순위로 고정되어 공통된 순위를 나타냈으며 1시기에서는 ‘금연장소’, 2시기는 ‘교육’, 3시기는 ‘건강관리’ 관련 내용의 키워드가 포함되었다. 특히 차별 키워드 3시기에서 ‘전자담배’가 흡연을 대체하는 방법으로 새롭게 등장하였다. 시기별 공통/차별 키워드 순위에 유의미한 변화가 있는지 알아보기 위해 카이검정을 실시한 결과 $p < .001$ 유의확률 수준에서 차이가 있는 것으로 나타났다.

표20. 시기별 휴식 공통/차별 키워드

구분	1시기	2시기	3시기
공통 키워드	재충전>시간>직장인> 전문가>관계자	재충전>시간>직장인> 관계자>전문가	시간>재충전>관계자> 직장인>전문가

차별 키워드	피로회복>피서지>여름 휴가>휴가>여름철>휴가 철>근로자>주일>관광지 >피로>고속도로>건강관 리>해수욕장>휴식시간> 자녀들	수영장>피로회복>웰빙> 관광객>아이들>선수들> 휴양지>물놀이>직원들> 여름휴가>온천수>가족들 >스트레칭>삼성>휴가철	힐링>관광객>산책로> 시민들>관광지>수영장 >아이들>컴터>사업비> 여행지>주민들>방문객 >직원들>휴식공간> 가족들
-------------------	--	--	--

$$\chi^2 = 601, df = 19, p < .001$$

*부등호는 키워드의 상대적 순위를 의미함

시기별 휴식 관련 공통 키워드는 총5개로 다른 유형에 비해 공통 키워드의 수가 적은편이다. <표 20>과 같이 나머지 15개는 차별 키워드로 나타났는데 5개의 공통 키워드 중에서도 1시기, 2시기에서 ‘재충전’ > ‘시간’ > ‘직장인’ 순서까지 동일한 순위를 나타냈다. 한편 차별 키워드의 수는 다른 건강관리 유형에 비해 많은 것으로 나타났는데 차별 키워드 1시기에서 ‘휴가’ 관련 키워드가 다수 나타났고 2시기에서는 ‘웰빙’, 3시기는 ‘힐링’ 과 같이 특성 있는 키워드가 도출되었다. 시기별 공통/차별 키워드 순위에 유의미한 변화가 있는지 알아보기 위해 카이검정을 실시한 결과 $p < .001$ 유의확률 수준에서 차이가 있는 것으로 나타났다.

표21. 시기별 수면 공통/차별 키워드

구분	1시기	2시기	3시기
공통 키워드	시간>불면증>수면장애> 전문의>전문가>우울증	시간>불면증>수면장애> 전문가>우울증>전문의	시간>전문가>불면증>우 울증>수면장애>전문의
차별 키워드	수면시간>수면제>수면 부족>운동>수면중>집중 력>결과>뇌파>교수>상 태>방법>숙면>호흡중> 정신과	숙면>수면시간>집중력> 호흡중>코골>수면제>수 면무>수면무호흡증>기 억력>다음날>고혈압>열 대야>연구팀>수면부족	관계자>가능성>소비자> 스마트폰>사용자>한국 인>연구원>아이들>프로 포폴>의약품>삼성전자> 부작용>시청자>발병

$$\chi^2 = 3995, df = 16, p < .001$$

*부등호는 키워드의 상대적 순위를 의미함

시기별 수면 관련 공통 키워드는 총6개이고 나머지 14개는 차별 키워드로 나타났다. <표 21>과 같이 수면 공통 키워드 전체시기에서 ‘시간’ 이 우선순위로 나타났으며 1시기와 2시기에 ‘시간’ > ‘불면증’ > ‘수면장애’ 순으로 동일한 순위가 나타났다. 수면은 흡연과 마찬가지로 공통 키워드 내용 범주가 넓지 않아 비슷한 내용의 키워드로 구성되었다. 한편 차별 키워드는 다양한 키워드가 나타났는데 1시기 ‘운동’, ‘뇌파’, 2시기 ‘집중력’, ‘기억력’, 3시기 ‘스마트폰’, ‘프로포폴’, ‘삼성전자’ 등 시기별 특성 있는 키워드가 나타났다. 시기별 공통/차별 키워드 순위에 유의미한 변화가 있는지 알아보기 위해 카이검정을 실시한 결과 $p < .001$ 유의확률 수준에서 차이가 있는 것으로 나타났다.

표22. 시기별 스트레스 공통/차별 키워드

구분	1시기	2시기	3시기
공통 키워드	직장인>시간>전문가>우울증	시간>전문가>우울증>직장인	시간>전문가>우울증>직장인
차별 키워드	해소>방법>현대인>운동>생활>정신>직장인들>전문가의>불면증>효과>고혈압>자신>성인병>해소법>건강관리>전문가들	아이들>관계자>학생들>전문의>가능성>자신감>여성들>발병>고혈압>일상생활>현대인>부작용>환자들>집중력>치료법>면역력	관계자>힐링>아이들>일들>학생들>가능성>프로그램>친구들>자신감>공무원>면역력>건강관리>보건소>발병>소비자>가족들

$$\chi^2 = 8232, df = 22, p < .001$$

*부등호는 키워드의 상대적 순위를 의미함

시기별 스트레스 관련 공통 키워드는 총4개로 전체 건강관리 유형 중에서 공통 키워드의 수가 가장 적은편이고 <표 22>와 같이 나머지 16개는 차별 키워드로 나타났다. 특히 2시기와 3시기의 공통 키워드에서 ‘시간’ > ‘전문가’ > ‘우울증’ > ‘직장인’ 4개의 공통 키워드 모두 일치된 순위를 나타냈다. 차별 키워드에서는 1시기, 2시기에서 ‘전문의’ 그리고 2시기, 3시기에서 ‘아이들’, ‘관계자’, ‘가능성’, ‘면역력’ 키워드가 중복되어 나타났다. 특히 3시기에서 휴식 3시기의 ‘힐링’ 과 동일한 키워드가 나타났다. 시기별 공통/

차별 키워드 순위에 유의미한 변화가 있는지 알아보기 위해 카이검정을 실시한 결과 $p < .001$ 유의확률 수준에서 차이가 있는 것으로 나타났다.

4. 시기별 건강관리 유형에 따른 키워드 트렌드

시간흐름에 따라 키워드 트렌드 관심도가 증가 또는 감소하였는지 확인하기 위해 빈도 분석결과에 나타난 공통 키워드를 이용하여 Hot_Cold 분석과 회귀상관 분석을 실시하였다. 유의수준 5%에서 유의한 확률을 갖는 주제에 대해 회귀계수 값이 양수(+)이면 Hot, 음수(-)이면 Cold로 구분하였다. 시기별 건강관리 유형에 따른 공통 키워드 Hot_Cold 분석결과는 다음 <표23>과 같다.

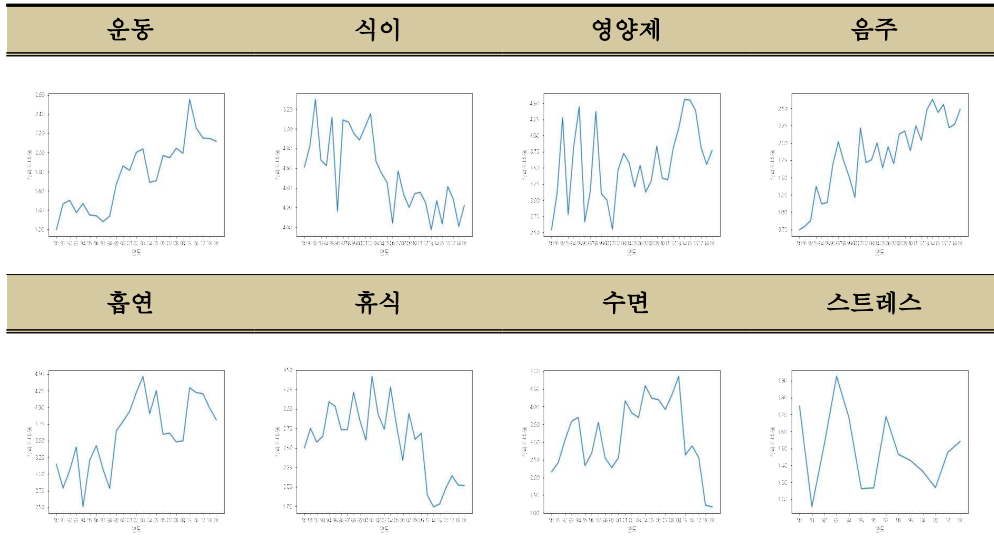
표23. 건강관리 유형에 따른 공통 키워드 Hot_Cold

구분	coefficient	p_value	Hot_Cold
운동	.043	.001	Hot
식이	-.032	.001	Cold
영양제	.028	.02	Hot
음주	.053	.001	Hot
흡연	.05	.001	Hot
휴식	-.036	.001	Cold
수면	.001	.994	-
스트레스	-.013	.398	-

*Hot은 증가, Cold는 감소 추세를 나타냄

키워드 빈도 분석 결과에서 나타난 공통 키워드를 이용하여 연도별 트렌드 추이를 파악하기 위해 Hot_Cold 분석을 실시하였다. 운동, 영양제, 음주, 흡연에서 Hot(정적상관)이, 식이, 휴식에서 Cold(부적상관)이 유의미하게 나타났으며($p < .05$) 수면, 스트레스에서는 유의미한 차이가 나타나지 않았다. Hot_Cold 분석에서 얻은 결과를 바탕으로 연도와 키워드 비중 관계를 그래프로 나타내면 다음 <표24>와 같다.

표24. 건강관리 유형에 따른 공통 키워드 트렌드 추이



5. 시기별 건강관리 유형에 따른 주요 토픽

이 연구는 LDA 알고리즘 기반 토픽모델링 분석을 실시하여 토픽 내 키워드 비중을 확인함으로써 토픽을 구성하는 주요 키워드가 무엇인지 확인하고 토픽의 구조적 특성을 파악하였다.

수차례 반복한 샘플링 분석을 통해 적정 토픽 수($k=7$)를 결정하고 이 중 토픽 비중 값이 나타난 토픽($k=3$)을 채택하였다. 각 토픽을 구성하는 키워드는 토픽 내 상위 빈출 20개 주요 키워드이며 비중은 주요 키워드들이 해당 토픽에 차지하는 비중을 나타낸 것이다. 이와 같이 키워드 간 유사성(Similarity)에 따라 추출되는 토픽모델링 결과를 근거로 Intertopic Distance Map(IDM)에 이용하여 토픽의 분포를 시각화 하였다. 이 때 전체에서 토픽 비중이 큰 순으로 토픽1, 토픽2, 토픽3의 순으로 토픽 번호를 부여하였다. Intertopic Distance Map(IDM) 그림에 나타난 각각의 원형 도형들은 하나의 주제를 나타내며 원의 거리가 분산될수록 또는 거리가 멀수록 판별 타당도가 높고 주제가 뚜렷하게 구분되는 것으로 해석할 수 있다. 시기별 건강관리 유형에 따른 토픽모델링 분석 결과는 다음과 같다.

토픽	비중	토픽명	키워드
1	.049	피트니스	피트니스, 시간, 관계자, 선수들, 유산소, 스트레칭, 체력, 동작, 건강관리, 근력운동, 학생들, 전문가, 보건소프로그램, 자신감, 직장인, 아이들, 참가자, 단백질
2	.057	유산소 운동	유산소, 유산소운동, 스트레칭, 피트니스, 시간, 근력운동, 동작, 선수들, 관계자, 보건소, 전문가, 고혈압, 체육활동, 운동량, 운동방법, 건강관리, 성인병, 학생들, 당뇨병
3	.029	체력	체력, 시간, 선수들, 운동장, 학생들, 지도자, 전문가, 운동량, 동호인, 성인병, 체력단련, 건강관리, 체육, 동작, 직장인, 금메달, 관계자, 활성화, 지구력



표26. 식이 토픽모델링 분석결과

토픽	비중	토픽명	키워드
1	.069	식이요법	식이요법, 단백질, 식이섬유, 영양소, 고혈압, 당뇨병, 전문가, 식이조절, 식습관, 다이어트, 운동, 관계자, 시간, 합병증, 비타민, 요법, 발병, 섬유소, 소비자
2	.062	운동과 식이요법	식이요법, 당뇨병, 고혈압, 단백질, 요법, 식이섬유, 운동, 합병증, 부작용, 식생활, 전문가, 성인병, 영양소, 시간, 운동요법, 보건소, 고지혈증, 전문의, 발병
3	.06	환자 치료식	식이요법, 당뇨병, 요법, 성인병, 단백질, 치료법, 합병증, 식생활, 고혈압, 운동요법, 영양소, 발병, 환자들, 부작용, 전문의, 소비자, 음식물, 전문가, 약물요법

〈표26〉과 같이 식이 토픽모델링 분석 결과 토픽1은 식이요법 관련 키워드가 큰 비중을 차지하였고 토픽2는 운동과 식이요법, 토픽3은 환자 치료식으로 분류하였다. 식이 토픽모델링 결과를 IDM을 이용하여 시각화하면 〈그림18〉과 같다.

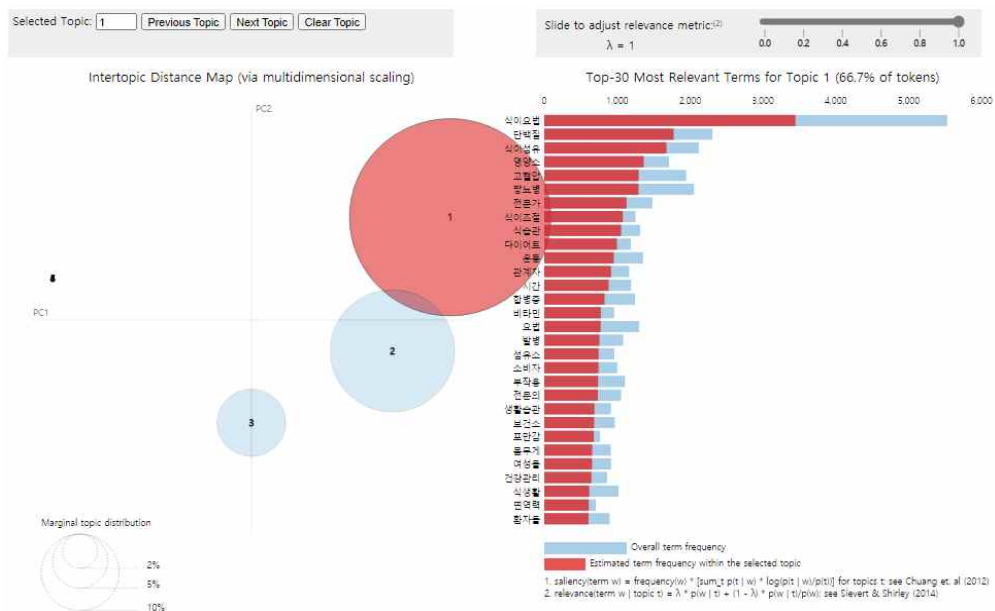


그림18. 식이 Intertopic Distance Map(IDM)

표27. 영양제 토픽모델링 분석결과

토픽	비중	토픽명	키워드
1	.069	건강 기능식품	비타민, 영양소, 관계자, 소비자, 의약품, 면역력, 오메가, 단백질, 임신부, 영양, 항산화, 건강기능식품, 건강관리, 전문가, 아이들, 부작용, 시간, 건강식품, 현대인
2	.053	제품	비타민, 의약품, 관계자, 소비자, 영양, 영양소, 건강식품, 부작용, 단백질, 화장품, 치료제, 아이들, 항산화, 제품, 소비자들, 웰빙, 비타민제, 시간, 건강관리
3	.024	의약품	영양, 비타민, 의약품, 관계자, 부작용, 단백질, 영양소, 전문가, 소비자, 항생제, 주사, 건강보조식품, 성인병, 가능성, 비타민제, 영양상태, 음식물, 식생활, 임신부

〈표27〉과 같이 영양제 토픽모델링 분석 결과 토픽1은 건강기능식품 관련 키워드가 큰 비중을 차지하였고 토픽2는 제품, 토픽3은 의약품으로 분류하였다. 영양제 토픽모델링 결과를 IDM을 이용하여 시각화하면 〈그림19〉와 같다.

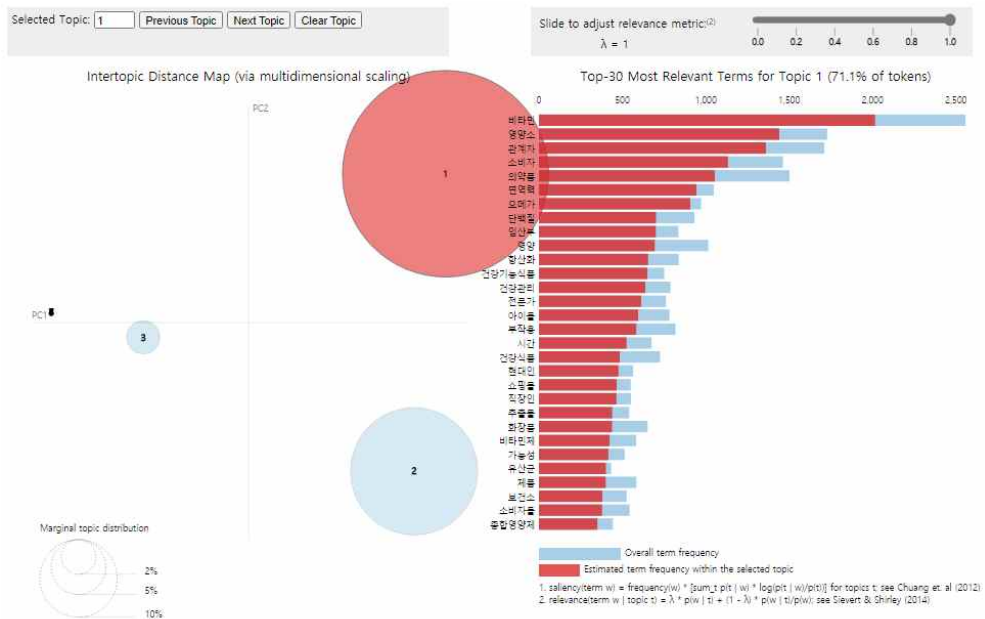


그림19. 영양제 Intertopic Distance Map(IDM)

표28. 음주 토픽모델링 분석결과

토픽	비중	토픽명	키워드
1	.05	절주	절주, 관계자, 음주, 술자리, 시간, 고혈압, 간염, 지방간, 가능성, 발병, 생활습관, 전문가, 식습관, 당뇨병, 전문의, 합병증, 직장인, 금연, 간암
2	.035	부작용 및 문제점	관계자, 가능성, 절주, 음주, 술자리, 시간, 간염, 전문가, 고혈압, 당뇨병, 지방간, 금연, 방법, 직장인, 발병, 간질환, 보건소, 전문의, 청소년
3	.023	치료법	관계자, 가능성, 방법, 전문가, 음주, 시간, 간염, 중독자, 지방간, 문제점, 공무원, 학생들, 치료, 주일, 운동, 치료법, 기업들, 발병, 환자들

<표28>과 같이 음주 토픽모델링 분석 결과 토픽1은 절주 관련 키워드가 큰 비중을 차지하였고 토픽2는 부작용 및 문제점, 토픽3은 치료법으로 분류하였다. 음주 토픽모델링 결과를 IDM을 이용하여 시각화하면 <그림20>과 같다.

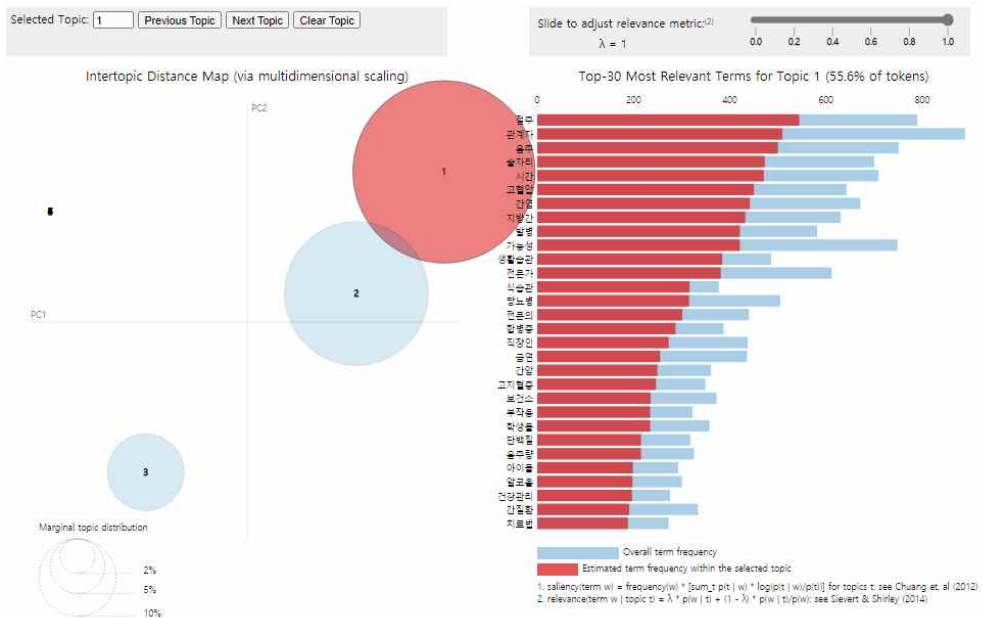


그림20. 음주 Intertopic Distance Map(IDM)

표29. 흡연 토픽모델링 분석결과

토픽	비중	토픽명	키워드
1	.09	금연정책	흡연자, 보건소, 금연구역, 관계자, 흡연, 간접흡연, 과태료, 담배, 보건복지부, 건강관리, 증진법, 전자담배, 흡연율, 절주, 흡연자들, 학생들, 담뱃값, 흡연예방, 고혈압
2	.054	금연구역	담배, 흡연, 흡연자, 금연운동, 증진법, 애연가, 금연구역, 사무실, 범칙금, 흡연율, 비흡연자, 관계자, 운동, 백만, 담배연기, 청소년들, 공공장소, 장소, 금연장소
3	.07	금연교육	흡연, 담배, 흡연자, 보건소, 관계자, 금연구역, 흡연율, 청소년, 학생들, 대상, 교육, 청소년들, 흡연자들, 금연운동, 담뱃값, 보건복지부, 간접흡연, 담배연기, 금연교육

〈표29〉와 같이 흡연 토픽모델링 분석 결과 토픽1은 금연정책 관련 키워드가 큰 비중을 차지하였고 토픽2는 금연구역, 토픽3은 금연교육으로 분류하였다. 흡연 토픽모델링 결과를 IDM을 이용하여 시각화하면 〈그림21〉과 같다.

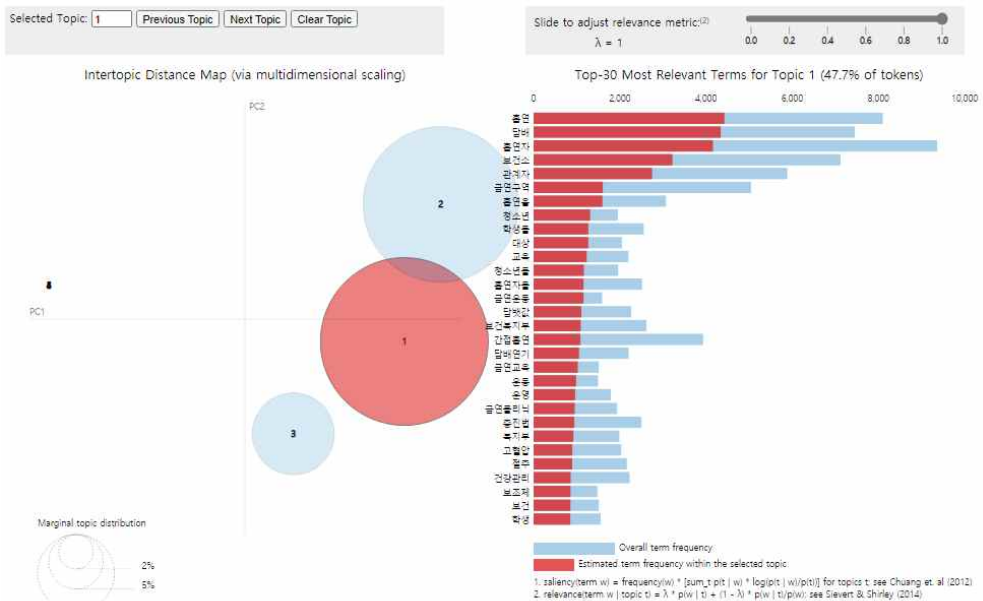


그림21. 흡연 Intertopic Distance Map(IDM)

표30. 휴식 토픽모델링 분석결과

토픽	비중	토픽명	키워드
1	.063	힐링	힐링, 시간, 재충전, 관계자, 관광객, 산책로, 시민들, 관광지, 수영장, 아이들, 직장인, 쉼터, 전문가, 사업비, 여행지, 주민들, 방문객, 직원들, 휴식공간
2	.047	휴가여행	재충전, 시간, 수영장, 피로회복, 직장인, 웰빙, 관계자, 관광객, 아이들, 선수들, 휴양지, 물놀이, 직원들, 여름휴가, 가족들, 온천수, 스트레칭, 전문가, 삼성
3	.033	재충전	재충전, 시간, 피로회복, 여름휴가, 피서지, 휴가, 직장인, 전문가, 여름철, 근로자, 휴가철, 관계자, 주일, 피로, 고속도로, 관광지, 자녀들, 휴식시간, 건강관리

〈표30〉과 같이 휴식 토픽모델링 분석 결과 토픽1은 힐링 관련 키워드가 큰 비중을 차지하였고 토픽2는 휴가여행, 토픽3은 재충전으로 분류하였다. 휴식 토픽모델링 결과를 IDM을 이용하여 시각화하면 〈그림22〉와 같다.

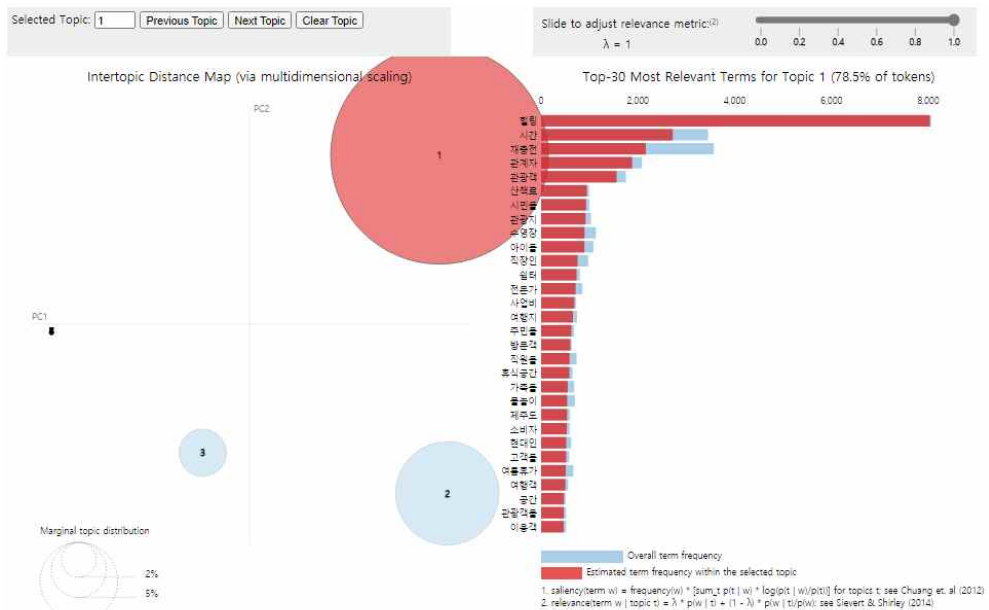


그림22. 휴식 Intertopic Distance Map(IDM)

표31. 수면 토픽모델링 분석결과

토픽	비중	토픽명	키워드
1	.046	스마트폰 사용	시간, 관계자, 가능성, 불면증, 전문가, 소비자, 스마트폰, 우울증, 집중력, 수면장애, 면역력, 전문의, 아이들, 일상생활, 수면시간, 직장인, 멜라토닌, 사용자, 발병
2	.056	수면부족	시간, 불면증, 집중력, 우울증, 전문가, 전문의, 수험생, 아이들, 가능성, 도움말, 수면장애, 증후군, 스트레칭, 고혈압, 수면부족, 수면시간, 기억력, 직장인, 부작용
3	.042	수면시간	시간, 불면증, 전문가, 전문의, 수면부족, 수면시간, 수면장애, 수면제, 치료, 수험생, 운동, 우울증, 스트레스, 집중력, 건강관리, 수면중, 전문가들, 습관, 여름철

〈표31〉과 같이 수면 토픽모델링 분석 결과 토픽1은 스마트폰 사용 관련 키워드가 큰 비중을 차지하였고 토픽2는 수면부족, 토픽3은 수면시간으로 분류하였다. 수면 토픽모델링 결과를 IDM을 이용하여 시각화하면 〈그림23〉과 같다.

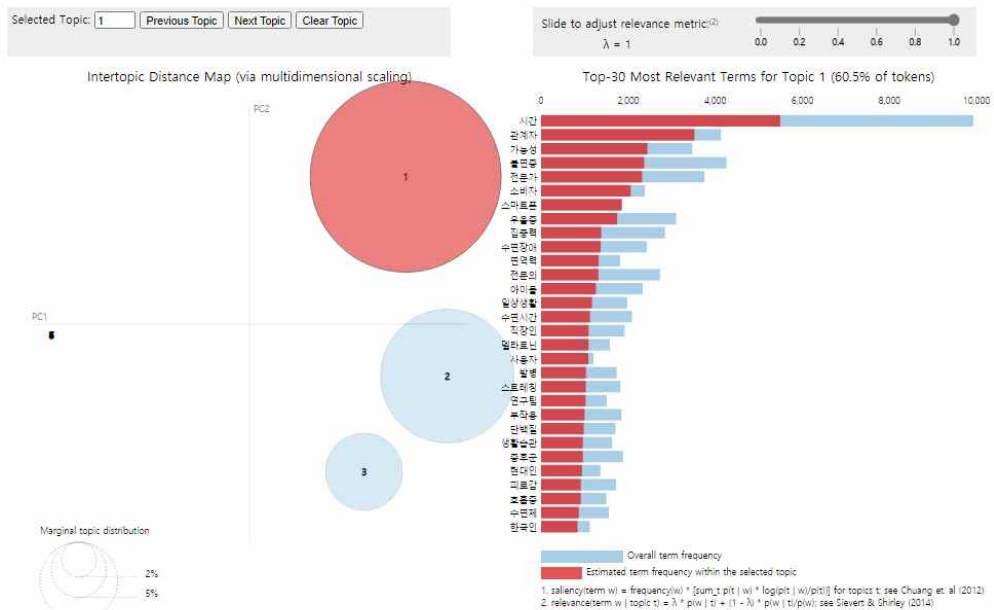


그림23. 수면 Intertopic Distance Map(IDM)

표32. 스트레스 토픽모델링 분석결과

토픽	비중	토픽명	키워드
1	.041	발병 및 치료	시간, 전문가, 우울증, 아이들, 직장인, 관계자, 학생들, 전문의, 가능성, 자신감, 여성들, 발병, 고혈압, 일상생활, 현대인, 부작용, 환자들, 집중력, 치료법
2	.044	힐링	시간, 관계자, 전문가, 우울증, 힐링, 아이들, 일들, 직장인, 학생들, 가능성, 프로그램, 친구들, 자신감, 공무원, 면역력, 건강관리, 보건소, 발병, 소비자
3	.035	스트레스 해소	해소, 직장인, 방법, 시간, 현대인, 운동, 전문가, 생활, 정신, 직장인들, 전문의, 우울증, 불면증, 효과, 고혈압, 자신, 성인병, 해소법, 건강관리

〈표32〉와 같이 스트레스 토픽모델링 분석 결과 토픽1은 발병 및 치료 관련 키워드가 큰 비중을 차지하였고 토픽2는 힐링, 토픽3은 스트레스 해소로 분류하였다. 스트레스 토픽모델링 결과를 IDM을 이용하여 시각화하면 〈그림24〉와 같다.

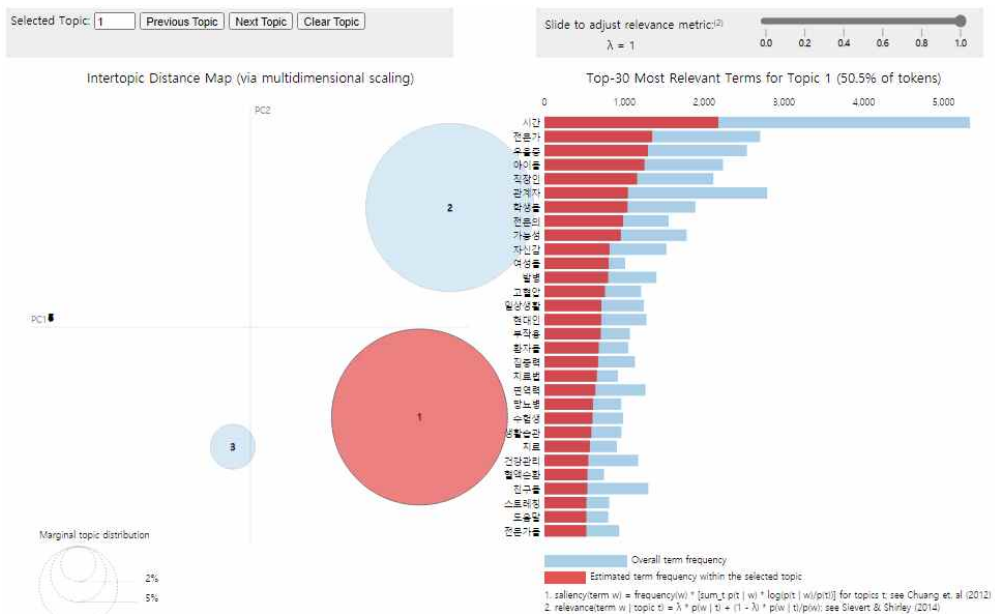


그림24. 스트레스 Intertopic Distance Map(IDM)

6. 시기별 건강관리 유형에 따른 토픽 트렌드

각 토픽의 시기별 증감 추세를 Hot_Cold 로 구분하여 토픽 관심도에 대한 상승과 하락을 판단하는 기준으로 선형 회귀분석의 회귀계수 값을 산출하였다. 유의수준 5%에서 유의한 확률을 갖는 토픽에 대해 회귀계수 값이 양수(+)이면 Hot, 음수(-)이면 Cold로 구분하였다. 시기별 건강관리 유형에 따른 토픽의 Hot_Cold 분석결과는 다음 <표33>과 같다.

표33. 건강관리 유형에 따른 토픽별 Hot_Cold

구분	Topic 1			Topic 2			Topic 3		
	coef	p_value	Hot_Cold	coef	p_value	Hot_Cold	coef	p_value	Hot_Cold
운동	.163	.001	Hot	.161	.001	Hot	-.018	.022	Cold
식이	.065	.001	Hot	-.02	.057	-	-.083	.001	Cold
영양제	.112	.001	Hot	.072	.001	Hot	-.022	.245	-
음주	.149	.001	Hot	.122	.001	Hot	.045	.001	Hot
흡연	.253	.001	Hot	.16	.001	Hot	-.054	.008	Cold
휴식	.122	.001	Hot	-.01	.612	-	-.134	.001	Cold
수면	.029	0.556	-	.081	.031	Hot	-.07	.129	-
스트레스	.162	.001	Hot	.05	.118	-	-.33	.001	Cold

*Hot은 증가, Cold는 감소 추세를 나타냄

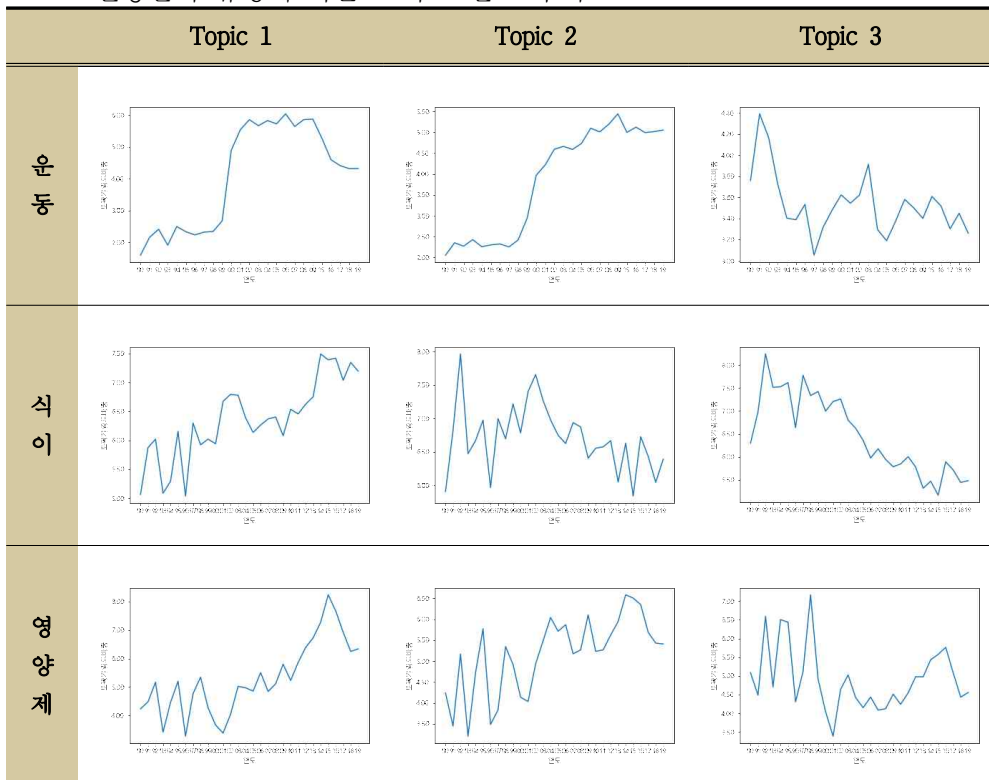
토픽모델링 분석 결과를 바탕으로 연도별 토픽 트렌드를 파악하기 위해 Hot_Cold 분석을 실시한 결과 <표33>과 같이 각 항목별 부분적으로 Hot(정적상관), Cold(부적상관)에 유의한 차이가 나타났다(α .05).

운동 토픽1과 토픽2는 Hot으로 나타난 반면 토픽3은 Cold로 나타났다(α .05). 식이 토픽1은 Hot, 토픽3은 Cold로 유의한 차이가 나타난 반면 토픽2는 감소 추세를 보였지만 통계적으로 유의미한 차이가 없는 것으로 나타났다(p .057). 영양제 토픽1과 토픽2는 Hot으로 나타난 반면 토픽3은 감소 추세를 보였지만 통계적으로 유의미한 차이가 없는 것으로 나타났

다($p=.245$). 음주는 전체 토픽에서 Hot으로 나타나 음주에 대한 관심도가 꾸준히 높아졌음을 확인하였다($p<.001$). 흡연 토픽1과 토픽2는 Hot으로 나타난 반면 토픽3에서 Cold로 나타났다($p<.01$). 휴식 토픽1은 Hot, 토픽3은 Cold로 유의한 차이가 나타난 반면 토픽2는 감소 추세를 보였지만 통계적으로 유의미한 차이가 없는 것으로 나타났다($p=.612$). 수면은 토픽2에서 Hot으로 나타났지만 토픽1은 증가 추세를 보였지만 통계적으로 유의미한 차이가 없는 것으로 나타났다($p=.556$). 반면 토픽3은 감소 추세를 보였지만 통계적으로 유의미한 차이가 없는 것으로 나타났다($p=.129$). 스트레스 토픽1은 Hot, 토픽3은 Cold로 유의한 차이가 나타난 반면 토픽2는 증가 추세를 보였지만 통계적으로 유의미한 차이가 없는 것으로 나타났다($p=.118$).

이와 같이 Hot_Cold 분석에서 얻은 결과를 토대로 연도와 토픽의 키워드 비중 관계를 그래프로 나타내면 다음 <표34>와 같다.

표34. 건강관리 유형에 따른 토픽 트렌드 추이





토픽 트렌드 내용을 정리하면 다음과 같다. (운동_피트니스와 유산소운동↑ 체력↓), (식이_식이요법↑ 환자 치료식↓), (영양제_기능성식품과 영양제 제품↑), (음주_절주, 부작용 및 문제점, 치료법 모두↑), (흡연_금연정책과 금연구역↑ 금연교육↓), (휴식_힐링↑ 재충전↓), (수면_수면부족↑ 수면시간↓), (스트레스_발병 및 치료↑ 스트레스 해소↓)

IV. 논 의

이 연구는 빅데이터를 활용하여 시기변동에 따른 건강관리 경향을 파악하는 것이 목적이다. 연구의 목적을 달성하기 위하여 지난30년 동안 수집된 건강관리 관련 뉴스기사를 이용하여 키워드 및 토픽 분석을 실시하였다. 이 연구의 분석결과를 토대로 다음과 같이 논의하고자 한다.

첫째, 1990년대는 운동과 관련해서 크게 엘리트체육과 생활체육 내용으로 나눌 수 있다. 이 연구의 키워드 분석 결과에서 운동 1시기에 ‘체력단련’, ‘선수들’, ‘금메달’ 과 같이 선수 관련 키워드와 ‘동호인’, ‘건강관리’, ‘활성화’, ‘생활’ 등 국민생활체육 관련 키워드가 다수 출현하였다. 이러한 결과는 1995년 국민건강 증진법을 제정하면서 국민건강증진사업이 활발히 시행됨에 따라 국가나 지역사회의 사회적, 환경적 지원개입으로 인해 실 내외 운동시설이 확충되고 건강증진 캠페인, 학교체육 강화 등 엘리트체육뿐 아니라 생활 체육 활성화가 반영된 결과라고 사료된다. 운동 2시기에서는 ‘유산소’, ‘유산소운동’ 이 가장 높은 비중을 차지하였으며 1시기에 ‘성인병’ 에 그치던 질병 관련 키워드가 ‘고혈압’, ‘성인병’, ‘당뇨병’ 으로 증가하였다. 또한 ‘보건소’, ‘건강관리’ 키워드와 연관 지어 등장하였는데 이러한 결과는 현대인의 좌식생활, 운동부족, 서구화된 식습관 등으로 인해 비만인구가 증가하면서 운동을 통한 예방적 건강관리가 강조되었음을 알 수 있었다. 1시기의 운동 주요키워드가 ‘체력’ 에 치중되었다면 2시기에서는 ‘유산소운동’, ‘피트니스’, ‘근력운동’, ‘웨이트 트레이닝’ 등 다양한 운동형태가 나타났다. 김경식 외(2019)의 연구에 따르면 2000년대 초반에 시작된 ‘몸짱’ 신드롬의 영향으로 젊은 층을 중심으로 몸 만들기 운동이 성행하였는데 이 연구의 결과에서도 ‘몸짱’ 키워드가 도출되어 부합된 내용인 것을 확인하였다. 흥미로운 사실은 새롭게 출현한 ‘스트레칭’ 키워드가 상위 3번째 키워드로 급부상하였는데 현대사회에서 좌식생활 시간이 늘고 각종 스트레스에 쉽게 노출되는 상황에서 고정된 자세와 긴장된 근육을 풀어주는 스트레칭 운동의 선호도가 높아진 것

으로 판단된다. 이러한 결과는 1990년대 이후 체력향상을 위한 운동에서 건강관리를 위한 운동 패러다임으로 전환되는 것을 의미한다는 점에서 Padden & Padden(2002)의 연구내용과 부합한다. 운동 3시기에서는 1시기, 2시기의 내용을 포함하면서 ‘피트니스’ 키워드가 높은 비중을 차지하였다. 2010년대 들어 근력운동에 대한 관심이 높아지면서 각종 피트니스 대회 및 교육기관, 자격증 관련 내용이 핫이슈였는데 이 연구결과에서도 ‘근력운동’, ‘자신감’, ‘참가자’, ‘단백질’과 같이 피트니스 선수 관련키워드가 다양하게 나타나 시기변동에 따라 운동 목적이 변화했음을 확인하였다. 한편 ‘스마트폰’ 키워드는 정보기술(IT)과 스마트 헬스케어 변화를 잘 반영한 결과로써 최근 스마트폰과 웨어러블 기기를 통해 사용자의 심박수, 걸음 수, 열량소모량 등 다양한 개인 활동정보를 제공받을 수 있다. 이처럼 현재의 운동관리는 단순히 신체를 단련시키는 것뿐만 아니라 개인 활동정보를 이용하여 피드백을 제공하는 등 빠르게 변화하는 정보사회 발달에 발맞춰 진화하고 있다.

둘째, 식이는 식생활과 연관 지을 수 있는데 식이의 핵심 키워드는 전체시기에서 ‘식이요법’으로 나타났다. 시기별로 질병 관련 키워드가 나타났는데 1시기에 ‘당뇨병’, ‘성인병’, ‘고혈압’, ‘합병증’, ‘심장병’, ‘치료법’, ‘환자들’, ‘발병’ 등 가장 많은 질병 키워드가 도출되었고 2시기에서는 ‘고지혈증’ 키워드가 새롭게 출현하였다. 또한 2시기, 3시기에서 ‘식이섭유’가 상위 키워드에 자리매김하는가 하면 ‘뱃살’, ‘몸무게’, ‘운동’, ‘체중감량’ 등 다이어트 관련 키워드가 다양하게 나타났다. 특히 2시기를 잘 반영해주는 ‘웰빙’ 키워드가 출현하면서 과거 환자를 위한 식단관리에서 2000년대 이후 비만환자가 증가하고 비만에 대한 관심이 높아지면서 ‘다이어트’, ‘건강관리’, ‘생활습관’, ‘식생활’ 키워드와 같이 질병예방을 위한 개인의 자발적, 적극적 식이관리로 변화된 것을 알 수 있었다. 흥미로운 점은 ‘면역력’, ‘항산화’, ‘비타민’과 같이 영양제 관련 키워드가 식이 키워드 워드클라우드에 나타난 것이다. 이러한 결과는 영양제 섭취가 일상생활 속 식문화로 자리매김하면서 음식을 통한 영양섭취 외에 건강보조식품을 통해 식이관리를 하는 것으로 사료된다. 또한 전체시기에서 ‘단백질’ 키워드가 중요하게 다뤄지는 동시에 ‘운동’, ‘운동요법’ 키워드가 함께 도출됨에 따라 운동과 식이는 건강관리를 위해 병행되어

아 할 중요한 요소인 것을 확인할 수 있었다.

셋째, 영양제 섭취는 식이에 포함되기도 하며 주로 영양결핍 또는 영양불균형에 대한 예방적 건강관리로 손꼽힌다. 영양제 관련 키워드에서는 ‘영양’, ‘비타민’, ‘의약품’을 중심으로 영양제 종류 및 효능, 부작용에 대한 키워드가 도출되었다. 1시기에서는 영양제 공급과 관련된 키워드로 ‘주사’, ‘주사제’가 있다. 또한 영양상태에 따른 영양보충이나 질병치료 목적으로 처방된 ‘치료제’, ‘항생제’, ‘비타민제’, ‘소화제’ 키워드가 함께 나타났다. 영양제 명칭과 관련해서 1시기에서만 ‘건강보조식’ 키워드가 출현했으며 2시기에서 ‘건강식품’, 3시기에서는 ‘건강기능식품’으로 나타났다. 1990년 후반부터 ‘건강보조식품’이라는 명칭으로 기능성을 표시했지만 효능과 부작용에 대한 논란이 끊이지 않았다. 이에 따라 2004년부터 식품의약품안전청에서 과학적 근거를 바탕으로 기능성 원료를 인정하여 ‘건강기능식품’ 명칭이 사용되었다. 2시기에서는 식이와 마찬가지로 2000년대를 잘 반영해주는 ‘웰빙’ 키워드가 출현하였으며 ‘항산화’ 키워드가 새롭게 출현했다. 흥미로운 점은 ‘화장품’ 키워드가 도출됨에 따라 이너뷰티(inner beauty)에 대한 관심이 높아졌음을 확인하였다. 또한 ‘아이들’, ‘제품’ 키워드를 통해 영양제 섭취 대상이 다양해지고 수요가 증가하면서 건강기능식품 관련 제품 출시가 본격화 된 것으로 판단된다. 식품의약품 안전처(2015)의 연도별 건강기능식품 생산실적 보고에 따르면 2011년에서 전년대비 28.2% 성장하여 가장 큰 폭으로 증가한 것으로 나타났는데 이 연구결과에서도 ‘제품’, ‘소비자들’, ‘쇼핑몰’ 키워드가 출현한 것으로 볼 때 영양제 제품 시장이 확대되었다는 통계자료 내용을 지지하고 있다. 2시기 이후부터는 영양제 종류와 관련된 ‘오메가’, ‘셀레늄’, ‘무기질’ 뿐 아니라 성분 및 함량에 대한 ‘고함량’, ‘추출물’ 키워드가 다양하게 도출되었다. 특히 3시기에서 ‘면역력’ 키워드가 새롭게 출현하며 상위 순위로 급부상하였는데 이는 최근 인체가 가진 자연치유력과 면역증진에 대한 관심이 반영된 결과라고 판단된다.

넷째, 건강관리를 위해서는 긍정적 행위를 증진 또는 유지시키는 것과 부정적 행위를 통제 또는 조절하는 것으로 나누어 볼 때 건강에 부정적인 영향을 미치는 요인은 대표적으로 음주와 흡연이 있다. 이 연구결과에서도 음주와 흡연은 서로 교차된 결과가 나타났는데 가

령 음주 2시기, 3시기에 ‘금연’ 이 나타났고 흡연 2시기, 3시기에 ‘절주’ 가 나타났다. Kwak et al., 2000; Fisher, Goden, 1985) 연구에서도 음주와 흡연은 상호연관성이 높아 음주량과 흡연량에서 양적 상관관계가 있는 것으로 나타났다. 음주 키워드 분석결과 대부분 잘못된 음주습관으로 인한 질병 관련 키워드 비중이 높게 나타났는데 1시기에서 ‘간염’, ‘지방간’, ‘간경화’, ‘간질환’ 등 음주 섭취가 간 건강에 여러 문제점을 야기 시키는 것으로 나타났다. 2시기는 ‘술자리’ 키워드 비중이 높게 나타났으며 이와 관련된 ‘음주 문화’, ‘폭탄주’ 키워드가 함께 도출되었다. 이는 현대사회 인간관계속에서 알코올 섭취가 빠질 수 없는 만큼 성인의 알코올 섭취 비율은 꾸준히 증가하고 있으며 젊은 층을 중심으로 다양한 음주 문화가 형성되고 있는 것을 확인하였다. 음주 1시기는 ‘금주’ 가 나타난 반면 2시기에 ‘절주’ 키워드가 나타났는데 3시기에서도 ‘절주’ 키워드가 유지됨과 동시에 압도적인 비중을 차지하였다. 이러한 결과는 ‘생활습관’, ‘건강관리’, ‘건강검진’, ‘예방’ 키워드가 같이 출현하면서 절주를 위한 예방적 건강관리로 확대된 것으로 판단된다.

다섯째, 흡연은 우리나라 사망원인 1위인 암과 직접적인 연관이 있는 부정적 행위로 잘 알려져 있다. 흡연의 경우 1시기는 ‘담배’, ‘흡연자’ 와 함께 ‘증진법’, ‘금연구역’ 키워드 비중이 높게 나타났는데 1995년 국민건강 증진법이 제정된 이후 흡연구역 지정, 담배광고 금지, 미성년자 담배판매 금지 등 담배에 대한 규제조치가 본격화된 시대배경을 반영한 것으로 판단된다(김동진, 2017). 또한 ‘금연운동’ 과 관련하여 ‘국민건강 증진법’, ‘한국금연운동협의회’, ‘보건복지부’, ‘보사부’ 키워드를 통해 금연지원정책이 강화된 것을 알 수 있었다. 2시기는 2002년 담배에 대한 건강증진부담금이 부과되기 시작하면서 ‘담뱃값’ 키워드가 새롭게 나타났다. 특히 ‘보건소’ 가 큰 비중을 차지하면서 ‘예방’, ‘교육’ 과 관련된 ‘금연교육’, ‘흡연예방’, ‘금연클리닉’ 키워드가 새롭게 출현하였다. 이러한 결과는 2002년부터 운동, 영양, 절주, 금연을 중심으로 건강증진사업이 실시되면서 금연사업 기반 마련 및 금연사업이 전개된 결과로 사료된다(양숙자, 2008). 3시기는 ‘보건소’ 비중이 지속적으로 증가하였으며 ‘구역’, ‘과태료’, ‘부스’, ‘화장실’ 과 같이 ‘금연구역’ 관련 키워드가 다수 도출되었다. 국민건강 증진법을 통해 2015년부터 금연구역을 지정하여 간접

흡연을 예방하고 담뱃갑에 혐오스러운 그림을 삽입하거나 담뱃값을 인상하는 등 ‘금연’ 이 적극 강조되었음을 알 수 있었다. 이와 같이 음주와 흡연에서는 관리방법 보다 국가 정책 사업이나 규제에 관한 키워드가 다수 출현하였다. 이는 뉴스기사가 갖는 특성과 관련지을 수 있는데 뉴스기사는 개인 블로그나 SNS와 달리 사실위주의 공식화 된 보도형식을 갖추기 때문에 음주와 흡연에 대한 개별적이고 상세한 예방관리 방법을 파악하는데 한계가 있는 것으로 판단된다. 따라서 음주와 흡연에 나타난 키워드를 개인의 예방관리 방법내용으로 정리하면 다음과 같다. 음주 2시기에서는 ‘절주’, ‘금연’, ‘운동’, ‘생활습관’ 으로 나타났고 3시기는 ‘식습관’, ‘식생활’, ‘식이요법’, ‘단백질’, ‘생활습관’, ‘금연’ 이 나타났다. 흡연의 경우 2시기에 ‘금연침’, ‘보조제’, ‘절주’, ‘운동’ 으로 나타났고 3시기는 ‘보조제’, ‘전자담배’, ‘생활습관’, ‘절주’ 로 나타났다. 반면 음주, 흡연 모두 1시기에서는 국가정책을 제외하면 ‘치료’ 이외에 다양한 예방관리 키워드가 나타나지 않았다.

여섯째, 휴식은 ‘재충전’ 과 ‘시간’ 키워드의 비중이 큰 것으로 나타나 몸과 마음의 재충전을 위한 시간으로 해석된다. 건강 및 외모관리 관련 트렌드 평가 학회자료(2012)에 따르면 건강 회복을 위해 집에서 휴식한다는 응답이 68.9%를 차지하며 압도적으로 높게 나타났다. 휴식의 의미를 여러 가지로 나누어 볼 때 체력충전을 위한 몸의 휴식, 여가시간을 위한 휴식활동, 운동휴식 등이 있다. 이 연구에서 나타난 휴식은 여가시간을 위한 휴식활동 관련 내용을 중심으로 1시기에서 ‘피로회복’, ‘피서지’ 키워드가 큰 비중을 차지하면서 피서지, 휴가 관련 내용이 주를 이루었다. 2시기는 ‘물놀이’ 관련 키워드가 다양하게 출현했으며 ‘휴가지’ 에 대한 내용으로 ‘강원도’, ‘충남’ 등 국내 지역뿐 아니라 ‘태국’, ‘아시아’, ‘프랑스’ 와 같이 해외 여행지 관련 키워드도 함께 출현하였다. 이러한 결과는 2000년 말 IMF 외환위기가 완화된 이후 국민생활수준이 향상되고 2005년 주5일제 근무가 시행된 이후 여가시간이 확보됨에 따라 외식 증가, 레저 활동의 증가, 해외여행 증가 등 과거 업무중심에서 개인을 위한 생활중심의 라이프스타일로 변화하기 시작하였다(김경식 외, 2019). 3시기는 ‘힐링’ 키워드가 핵심 키워드로 나타나 스트레스 3시기와 일치된 결과가 나타났는데 이는 스트레스관리에서 몸과 마음의 회복을 위한 휴식이 포함된 것으로

사료된다. ‘힐링’과 관련하여 ‘산책로’, ‘쉼터’, ‘휴식공간’이 주요 키워드로 도출되었고 이 외에도 ‘캠핑장’, ‘자연환경’, ‘휴양림’과 같이 자연친화적 키워드가 출현하였다. 이러한 결과는 김경식 외(2019)연구에서 여가와 관련하여 자연환경이 어우러진 장소나 친환경 여가활동으로 제시한 내용과 부합하여 휴식을 위한 여가공간으로서의 의미를 갖는 것을 확인하였다. 이 외에도 ‘관광객’의 비중이 급부상하면서 ‘관광지’, ‘여행지’, ‘방문객’, ‘이용객’ 키워드가 함께 도출되었는데 실제 2008년부터 해외여행객 수가 늘어나 2015년에 2469만 명으로 2배 가까이 증가한 것으로 조사되었다(국가지표체계, 2016). 3시기에서 특정장소와 관련하여 ‘도서관’, ‘전망대’, ‘편의시설’과 같이 이전 시기와 차별화된 키워드가 다양하게 출현하면서 휴식활동, 휴식장소에 대한 다양성을 확인할 수 있었다.

일곱째, 수면은 단순한 휴식을 넘어 몸의 휴식과 건강을 위해 꼭 필요한 요인이다. 수면 하면 불면증을 빼놓을 수 없는데 이 연구결과에서도 1시기와 2시기에서 ‘불면증’과 ‘시간’ 키워드가 큰 비중을 차지하였다. 또한 전체시기에서 불면증, ‘수면부족’, ‘수면장애’와 같이 수면 관련 문제점이 꾸준히 나타나면서 수면 중 증상, 질병, 치료 관련 키워드가 함께 증가하여 나타났다. 국민건강보험공단에서 불면증 치료환자를 조사한 결과 2012년에 비해 2016년에 34.3% 증가한 것으로 나타나 이 연구결과와 같이 수면관리에서 불면증의 영향력은 상당한 것으로 판단된다. 2시기는 ‘숙면’ 키워드의 비중이 눈에 띄게 증가하였는데 ‘숙면’과 ‘중요’ 키워드를 통해 알 수 있듯이 숙면의 중요성이 부각되었음을 알 수 있었다. 이는 오수연(2019)이 제시한 현대인의 숙면 트렌드는 잠에도 돈을 투자하는 슬리포노믹스 시대를 반영한 결과로 사료된다. 3시기는 ‘스마트폰’ 키워드가 출현하면서 취침 전 스마트폰 사용이 숙면을 방해하는 요소로 급부상하였다. 한편 수면 관련 앱, 웨어러블 기기 등 스마트폰을 통해 수면패턴을 분석하거나 수면을 유도하려는 긍정적 변화도 함께 나타나고 있다.

여덟째, 스트레스는 휴식, 수면 내용과 중복된 부분이 나타났으며 스트레스관리를 위해 잠을 자거나 휴식을 취하기도하고 휴식과 관련하여 다양한 여가활동을 즐기는 것을 알 수 있었다. 스트레스 키워드 분석결과에서 ‘현대인’, ‘학생들’, ‘직장인들’, ‘환자들’, ‘아

이들', '여성들', '수험생' 등 스트레스 대상자 관련 키워드가 끊임없이 나타났다. 그만큼 스트레스는 남녀노소, 직업이나 신분과 상관없이 우리 삶에 혼재되어 있음을 알 수 있었다. 1시기에서 스트레스 '해소'가 큰 비중을 차지하였고 스트레스 해소법으로 '운동'이 핵심키워드로 나타나났는데 이는 운동을 통해 마음과 정신건강을 도모하고 우울증을 예방하려는 것으로 판단된다. 시기별 스트레스관리는 2시기의 '웰빙', 3시기의 '힐링'과 같이 당시 사회문화 변화를 받는 것으로 나타났으며 '화장품', '한의원', '미세먼지', '스마트폰' 등 특성 있는 키워드가 다양하게 출현하였다. 특히 수면 3시기와 마찬가지로 '스마트폰' 키워드가 출현하면서 스마트폰 활용의 이점과 스마트폰 사용으로 인한 문제점이 함께 나타나는 것을 알 수 있었다. 무엇보다 스트레스관리가 곧 건강관리인 만큼 이 연구결과의 '마음', '만족도', '정신건강' 키워드를 통해 알 수 있듯이 외부환경으로부터 받는 스트레스를 자신의 심신단련을 통해 스스로 해소하는 것이 중요한 것으로 사료된다.

종합해보면 건강관리에 대한 관심과 노력은 날로 증가하고 있으며 시기마다 사회문화적 영향을 받아 다양하게 변화하고 있다. 운동, 식이는 요법과 관련하여 공통적인 결과가 나타나 질병치료를 위해서는 운동과 식이요법이 병행되어야하는 것을 알 수 있었고 영양제가 새로운 식문화로 자리매김하면서 섭취행위와 관련하여 식이, 영양제 결과가 일부 중복되는 것으로 나타나 영양제 섭취가 곧 식이관리와 연결되어 있음을 확인하였다. 또한 휴식, 수면, 스트레스의 경우 힐링과 같이 몸과 마음의 치유와 회복, 정신건강과 관련하여 상호적인 관계인 것을 파악할 수 있었다. 반면 음주, 흡연은 건강 관련 부정적 요인으로서 공포소구(fear appeal)와 관련하여 국가정부차원의 지원정책 및 지원사업, 규제방안에 대한 내용을 다수 포함하고 있으며 이 연구의 결과를 통해 음주관리를 위해 금연을, 금연관리를 위해 절주를 동시에 관리하는 것이 보다 효과적인 건강관리 방법인 것으로 사료된다.

V. 결론

이 연구는 뉴스 빅데이터를 활용하여 지난 30년간 건강관리 경향이 어떻게 변화하였는지 파악하는 것이 목적이다. 연구의 목적을 달성하기 위해 키워드 및 토픽 분석을 실시하여 얻은 결과를 바탕으로 다음과 같은 결론을 내렸다.

첫째, 운동관리 변화에서 운동방법 및 운동형태, 운동 목적이 시기에 따라 다양해지고 복잡해졌음을 알 수 있었다. 과거 선수와 체력단련 위주의 운동이 2000년대에 스트레칭과 유산소운동으로 전환되면서 다이어트를 위한 운동이 주를 이루었고 최근까지도 그 인기가 이어지고 있다. 2010년대 운동은 피트니스 개념이 추가되어 PT, 피트니스 선수, 트레이너 등 새로운 직업, 운동문화를 형성하고 있다.

둘째, 식이관리 변화는 과거 질병 치료를 위한 식단관리, 식이요법에서 2000년대 웰빙 시대에는 다이어트와 관련하여 단백질, 식이섬유에 관심이 높게 나타났다. 또한 식이요법과 운동요법이 함께 등장하면서 운동과 식이는 건강관리를 위해 병행되어야 할 중요한 요소인 것을 확인할 수 있었다. 2010년부터 최근까지 면역력, 항산화, 비타민 등 영양제 관련 내용이 식이와 맞물리면서 영양제가 하나의 식문화로 자리매김한 것을 확인하였다.

셋째, 영양제는 건강기능식품(건기식)이라고도 칭하며 2000년대부터 건강기능식품의 종류 및 효능과 부작용에 대한 관심이 높아진 것으로 나타났다. 1990년대 영양제는 영양부족과 같이 영양상태에 따라 처방 받거나 질병 치료를 위한 의약품의 역할이 컸다면 2000년대 이후 면역력, 항산화 기능이 주목받으면서 건강관리를 위해 누구나 스스로 챙겨먹는 식품이 되었다. 특히 2010년대는 피부, 화장품과 관련하여 이너뷰티(inner beauty)시대가 도래된 것을 알 수 있었다.

넷째, 음주 같은 경우 전반적으로 잘못된 음주습관으로 인한 질병 또는 질병 발생 가능성과 관련된 결과가 나타났다. 2000년대는 음주문화 관련 내용이 부각됨에 따라 알코올 섭취 비율이 꾸준히 증가하고 있음을 알 수 있었다. 2010년대 들어 절주와 같은 예방관리 행동이 확대되었고 식이요법이나 건강검진을 포함한 복합적인 음주관리가 이루어진 것으로 나타났다.

다섯째, 흡연은 개인의 예방관리 보다 국가 정책 관련 내용으로 금연운동, 금연구역, 금연 교육으로 분류되었다. 1990년대 금연운동, 2000년대 금연클리닉, 2010년대 보건소를 중심으로 흡연 예방관리가 실시되었으며 개인 차원의 노력은 전자담배가 대표적인 것으로 나타났다. 이를 통해 간접흡연에 대한 민감도가 증가함에 따라 흡연의 직간접적 피해 예방을 위한 제도 및 정책이 강화된 것을 알 수 있었다.

여섯째, 과거 휴식은 재충전, 피로회복과 같이 완전휴식의 성격이 강한 반면 2000년대 웰빙 열풍의 영향으로 행복과 만족감이 더해진 휴식, 여가활동으로 변화하였다. 특히 온천, 수영장, 워터파크 등 물놀이가 주된 내용이었으며 이 밖에도 해외여행이 급부상하였다. 2010년대는 힐링을 중심으로 자연친화적 콘셉트부터 개인취향에 맞는 특정장소까지 이전과 다르게 특색 있고 다양한 휴식방법이 나타났다.

일곱째, 수면은 불면증과 시간이 큰 비중을 차지하며 현대인의 수면부족, 수면장애 같은 문제와 연관되어 있는 것으로 나타났다. 1990년대는 불면증과 관련하여 수면제, 병원, 치료로 제한되었으나 2000년대 수면의 중요성이 대두되면서 수면다원검사, 뇌파검사와 같이 수면질환에 대한 예방적 검진이 활발히 이루어졌음을 확인하였다. 2010년대 수면방해 요인으로 스마트폰 사용이 주요원인으로 나타나는 동시에 수면관리를 위해서 스마트폰이나 웨어러블 기기를 이용하는 것으로 나타나 스마트폰의 양면성을 확인할 수 있었다.

여덟째, 스트레스는 부정적 요인으로 인식되어 스트레스관리 방법으로 해소, 해결 등이 수식되어왔다. 또한 시기를 막론하고 스트레스를 받는 대상이 다양하게 나타나 스트레스의 심각성을 짐작할 수 있었다. 1990년대의 스트레스 관리는 마음, 정신을 가다듬는 정적인 형태로 나타났으며 2000년대 웰빙과 관련하여 평소 생활습관이나 건강관리를 통해 스트레스를 줄이거나 예방하는 노력이 이루어졌다. 2010년대는 힐링 열풍의 영향으로 개인의 만족감을 중요시하는 것으로 나타났다. 한편 미세먼지가 스트레스의 새로운 원인으로 나타났고 우울증에 대한 비중도 꾸준히 증가하면서 스트레스관리의 한계점도 나타났다.

결론적으로 시기별 건강관리 유형에 따른 주요 키워드와 토픽은 서로 다른 경향을 보이며 변화하고 있다. 현대사회에서 나타나는 건강위험요인이 많아질수록 이에 대응하는 예방관리 방법도 다양하게 나타나고 있으며 사회문화적 트렌드를 반영한 건강관리 서비스 및 정책, 교육, 사업 등이 활발하게 진행되고 있다. 무엇보다 건강관리는 스스로 긍정적인 행위를 도모하고 부정적 행위에 대해서 통제 및 조절관리 할 수 있는 것으로 자신을 주체로 보다 능동적이고 적극적인 관리가 중요하다.

끝으로 건강은 환경에 적응하는 조건과 능력을 발휘하여 능동적인 건강관리를 통해 개인이 갖는 건강 잠재력을 극대화시키는데 궁극적인 목적이 있다. 본인 스스로 하는 능동적 건강관리 곧 건강행동은 삶의 질을 높여줄 뿐 아니라 질병이나 사망을 지연시키고 건강상태를 오래 유지할 수 있다. 건강에 위협을 주는 위험요소가 많은 현대사회에서 질병으로 인한 사망원인으로 생활습관이 손꼽히는 만큼 환자뿐 아니라 정상인까지도 질병예방·관리 차원의 건강관리는 매우 중요하며 개개인의 건강은 지역사회의 건강, 더 나아가 국가의 건강과 긴밀하게 연결되어있음을 잊지 말아야 한다.

참고 문헌

- 강범일, 송민, 조화순(2013). 토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구. **한국문헌정보학회지** 47(4), 315-334.
- 곽재현, 홍지숙(2018). 빅데이터를 활용한 옴로(YOLO) 현상 분석. **관광연구저널** 32.2: 21-34.
- 권영옥(2013). 빅데이터를 활용한 맞춤형 교육 서비스 활성화 방안연구. **한국지능정보시스템학회, 학술대회논문집**, 2013(6), 90-96.
- 권태연, 임자영, 박유성(2017). 국민건강보험 표본코호트DB를 이용한 한국인의 건강기대수명 연구. **응용통계연구**, 30(3), 475-486.
- 김경식, 이연주, 한승진, 한승백(2019). 빅데이터를 활용한 여가트렌드 분석. **한국여가레크리에이션학회지**, 43(1), 25-41.
- 김동진(2017). 제4차 국민건강증진종합계획 추진을 위한 정책 방향과 과제 : 건강생활실천 및 만성질환 관리를 중심으로. **보건복지포럼**, 246, 6-21.
- 김소연, 김선이(2012). 의료소비자의 건강 라이프스타일 유형에 따른 온라인 정보탐색에 관한 연구. **한국소비문화연구**, 15(2), 75-97.
- 김현정, 조남옥, 신경식(2014). 항공산업 미래유망분야 선정을 위한 텍스트 마이닝 기반의 트렌드 분석. **한국지능정보시스템학회, 학술대회논문집**, 11, 194-202.
- 김홍남, 최정웅(2008). 체육시설 현황 분석 결과에 의한 생활체육 활성화 방안. **한국사회체육학회지**, 33(1), 317-326.
- 류지혜(2018). 토픽 모델링을 이용한 보건학 분야 국내학술지 연구동향 분석. 고신대학교 박사학위논문.
- 문옥륜, 이상이, 정백근, 이신재, 김남순, 장원기, 윤태호(2000). 우리나라의 사회계층별 건강행태의 차이. *Journal of Preventive Medicine and Public Health* 33.4: 469-476.
- 박명은, 유소이(2019). 유기가공식품 선택에 대한 소비자 특성 분석: 계획적 행동이론의 확장과 식생활 라이프스타일을 중심으로. **소비자학연구**, 30(6), 143-170.

- 박상준, 원도연(2020). 2000년대 이후 미디어를 통한 소비와 행복의 트렌드 분석 - 빅카인즈와 구글 트렌드 분석을 중심으로. **문화경제연구** 23(2): 59-84.
- 박승정, 전진오, 김선우, 김성태(2017). 국내 주요일간지의 대통령 이슈소유권에 대한 빅데이터 분석 : LDA(latent Dirichlet allocation) 토픽 모델링(Topic Modeling)을 중심으로. **정치정보연구** 20(3), 25-56.
- 박주섭, 김나랑, 한은정(2018). 키워드 네트워크 분석을 활용한 과학기술동향 분석. **한국산업정보학회논문지**, 23(2), 63-73.
- 배정환, 손지은, 송민(2013). 텍스트 마이닝을 이용한 2012년 한국대선 관련 트위터 분석. **지능정보연구** 19(3), 141-156.
- 사공진(2003). 건강보험 재정안정화 방안에 대한 연구. **보건경제와 정책연구**, 9(2), 53-78.
- 설동훈, 고재훈, 유승환, 이기재(2020). 한국조사연구학회 발표 논문 내용분석, 1999~2019년: 학문분야방법론연구대상 데이터분석기법의 지속과 변동. **조사연구** 21(1), 1-32.
- 손준석, 송광현, 이원주, 함영식, 오재민, 전인호, 김영익(2012). 무선랜 AP 기반 지하철 객차별 승객수 추정 기법 연구. **한국통신학회 학술대회논문집** 2012(11), 197-198.
- 송태민(2013). 인터넷 건강정보 게이트웨이 시스템 구축 및 운영. 서울: **한국보건사회연구원**, 37, 299-305.
- _____(2015). 소셜 빅데이터 분석과 활용 방안. **보건복지포럼**, 277(9), 29-49.
- 식품의약품안전처(2015). 2014년 건강기능식품 국내 시장 규모 동향분석. **식품의약품안전처**, 2015
- 안영창(2014). 국민건강보험 인식수준과 만족도의 관계. **한국콘텐츠학회논문지**, 14(8), 277-286.
- 안주영, 안규빈, 송민(2016). 텍스트 마이닝을 이용한 매체별 에볼라 주제 분석. **한국문헌정보학회지** 50(2), 289-307.
- 안창원, 황승구(2012). 빅 데이터 기술과 주요 이슈. **정보과학회지** 30(6), 10-17.
- 양숙자(2008). 흡연 · 금연 관련 연구동향 분석. **한국보건간호학회지**, 22(2), 255-265.

- 엄경희, 정보리(2014). 힐링 트렌드가 반영된 기능성 섬유 제품 사례 분석. **한국디자인문화 학회지**, 20(3), 343-352.
- 오수연(2019). 꿀잠 삽니다. **마케팅**, 53(12), 51-57.
- 유단비(2010). 신자유주의적 통치성 관점에서 본 한국 사회의 ‘건강한 라이프스타일’ 담론 연구. 연세대학교 커뮤니케이션대학원 국내석사학위논문.
- 유예림(2017). 빅데이터 분석 기법을 활용한 2015 개정 교육과정 정책에 대한 언론보도 분석. 서울대학교 박사학위논문.
- 이나영, 최정화(2019). 빅데이터 분석을 통한 밀키트에 대한 소비자 인식 조사. **한국외식산업학회지**, 15(4), 211-222.
- 이미경(2018). 소셜 빅데이터를 이용한 한국인의 여행트렌드 분석 ‘가족여행’ 과 ‘나홀로 여행’ 을 중심으로. **관광학연구** 42(10), 111-134.
- 이상훈, 김영갑, 정현중(2015). WANO Analysis Report(2015-3), 전자기기 노후화 관련 국내 원전 현황 연구. **한국에너지학회 학술발표회**, 205-205.
- 이수정(2020). 키워드 네트워크 분석을 통한 건강보험연구 동향 분석. 이화여자대학교, 박사학위논문.
- 이은별, 전진오, 백지선(2017). 서울의 다문화 공간 연구. **미디어경제와 문화** 15(2), 7-43.
- 이종구(2004). 건강증진 정책 방향과 연구 : 우리나라 건강증진정책의 현황과 향후 정책 방향. **서울대학교 간호과학연구소 학술발표 논문집**, 2004, 3-26.
- 조경두(2014). 지속가능한 힐링환경을 위한 새로운 도전. **환경논총**, 53, 25-34.
- 학회자료(2012). 건강 및 외모관리 관련 트렌드 평가. **리서치보고서**, 2012(9), 83-118.
- 한승희(2018). 뉴스 빅데이터를 이용한 우리나라 언론의 기록관리 분야 보도 특성 분석: 1999~2018 뉴스를 중심으로. **정보관리학회지** 35(3), 41-75.
- 국가지표체계[웹사이트].(2020). 해외여행자수.
https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1655
- 국민건강보험 통계[웹사이트].(2016). 불면증환자 4년 새 34% 급증.

https://m.health.chosun.com/svc/news_view.html?contid=2018033002265

빅카인즈[웹사이트]. <http://www.bigkinds.or.kr>

질병관리본부[웹사이트].(2016). **건강행태 개선사업 통계**. <http://www.kdca.go.kr/yhs>.

통계청 공식 블로그[웹사이트].(2019). **운동하면 어떻게 좋을까**.

https://blog.naver.com/hi_nso/221714228279

통계청 공식 블로그[웹사이트].(2020). **끝나지 않은 셀프케어 트렌드**.

https://blog.naver.com/hi_nso/222121068537

Bertino, E. (2013). Big data--opportunities and challenges panel position paper. In 2013 IEEE 37th Annual Computer Software and Applications Conference, 479-480.

Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of science. *The Annals of Applied Statistics*, 17-3.

Fisher M, Gorden T. (1985). The relation of drinking and smoking habits to diet: the lipid research clinic prevalence study. *Am J Clin Nutr*, 41, 623-630.

Gould D. W., Graham-Brown M. P., Watson E. L., Viana J. L., Smith A. C. (2014). Physiological benefits of exercise in pre-dialysis chronic kidney disease. *Nephrology* 19, 519-527.

Hofmann (1999), "Probabilistic Latent Smentic Indexing."

Kokkinos, P. F., & Papademetriou, V. (2000). Exercise and hypertension. *Coronaryartery disease*, 11(2), 99-102.

Kwak CS, Lee JW, Hyun WJ. (2000). The effects of smoking and alcohol drinking on nutritional status and eating habits in adult males. *Korean J Community Nutr*, 5, 161-171.

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., & Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis*, 23(2),

254-277.

- Manovich, L. (2011). Trending: the promises and the challenges of big social data. *Debates in the digital humanities*, 2(1), 460-475.
- Padden D. L., Padden M. N. (2002). The role of the advanced practice nurse in the promotion of exercise and physical activity. *Top Adv Pract Nurs J*, 2, 93-115.
- Solomon, T. P., Sistrun, S. N., Krishnan, R. K., DelAguila, L. F., Marchetti, C. M., O'Carroll, S.M., O'Leary, V. B., & Kirwan, J. P. (2008). Exercise and diet enhance fat oxidation and reduce insulin resistance in older obese adults. *Journal of Applied Physiology*, 104(5): 1313-1319.
- World Health Organization. (1998). The Role of the pharmacist in self-care and self-medication : report of the 4th WHO Consultative Group on the Role of the Pharmacist, The Hague, The Netherlands, 26-28 August 1998. World Health Organization. <https://apps.who.int/iris/handle/10665/65860>
- Zikopoulos, P., Eaton, C., deRoos, D., Deutsch, T. and Lapis, G. (2012). *Understanding Big Data - Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw Hill.

ABSTRACT

Trend Analysis of Health Care Using Big Data

Ji-Eun, Kim

(Graduate School of Korea National Sport University)

Hyun-Tae, Kim(Advisor)

This study aims to identify the trends of health care by period using big data. To achieve this purpose, keyword and topic analyses were conducted of the news articles collected from bigkinds (www.bigkinds.or.kr) for the past 30 years with python 3 program. In the first place, the period was subcategorized into period 1 (1990s), period 2 (2000s), and period 3 (2010s) before extracting main keywords by period depending on type of health care and visualizing with word cloud. In the second keyword frequency analysis, the frequency of top 20 keywords and TF-IDF (term frequency-inverse document frequency) values were drawn and arranged in order of high frequency. Third, to identify the changes in ranking of common/differential keywords by period in the keyword appearance frequency analysis, chi-squared test was conducted. Fourth, to identify if common keyword trends increased or decreased significantly with the passage of time, Hot_Cold analysis and regression analysis were conducted. Fifth, the percentage of keywords within a topic was identified through topic modeling analysis before checking what keywords were mainly handled in each topic. Sixth, to identify if trends of each topic increased or decreased significantly with the passage of time, Hot_Cold analysis and regression analysis were conducted. As a result, it appeared that in <exercise>, ‘athlete’ and ‘physical strength’ keywords occupied

high percentage in period 1, 'aerobic' and 'stretching' in period 2, and 'fitness' and 'time' in period 3. In <diet>, 'dietary therapy' occupied an overwhelmingly great percentage in all periods and 'dietary fiber' keyword occupied a high percentage in period 2 and 'dietary control' in period 3. In <nutritional supplement>, 'vitamin' appeared as keyword in all periods, and 'nutrition' was drawn as main keyword in period, 'medicine' in period 2, and 'nutrient' in period 3. Besides, 'lactic acid bacillus', 'selenium', and 'omega' keywords also appeared. In <drinking>, 'the persons interested' appeared as keyword in all periods and 'possibility' appeared as main keyword in period 1 and period 2 and 'moderate drinking' in period 3. In <smoking>, general contents overall appeared similarly, and centering on 'cigarette' and 'smoking' related keywords, 'anti-smoking campaign' was drawn as main keyword in period 1 and 'public health center' in period 2 and period 3. In <break>, 'recharge' and 'time' appeared as main keywords in period 1 and period 2 and 'healing' occupied the highest percentage in period 3. In <sleep>, 'time' appeared as main keyword, and 'sleeplessness' was drawn as main keyword in period 1 and period 2 and 'the persons interested' in period 3. In <stress>, 'stress reduction' appeared as main keyword in period 1 and 'time' in period 2 and period 3 and the keywords related to the subjects who got stressed appeared multiply. Like this, it was identified that major keywords appeared differently depending on type of health care and common/differential keyword showed a significant difference in ranking ($p<.001$). As a result of looking at the trends of common keyword by year using regression analysis and Hot_Cold analysis, it was found that Hot(positive correlation) appeared significant in exercise, nutritional supplement, drinking, and smoking and Cold(negative correlation) appeared significant in diet and break ($p<.05$) and there was no significant difference in sleep and stress. In LDA topic modeling analysis, number of topic ($k=3$) was designated by type of health care and topic was

numbered in order of higher percentage. As a result, it was classified as follows: in <exercise>, topic 1: fitness, topic 2: aerobic exercise, and topic 3: physical strength; in <diet>, topic 1: dietary therapy, topic 2: exercise and dietary therapy, and topic 3: therapeutic diet; in <nutritional supplement>, topic 1: health functional food, topic 2: product, and topic 3: medicine. In <drinking>, it was classified as topic 1: moderate drinking, topic 2: side effects and problems, and topic 3: therapeutics, and in <smoking>, topic 1: antismoking policy, topic 2: non-smoking area, and topic 3: anti-smoking education. It was classified as in <break>, topic 1: healing, topic 2: holiday trip, and topic 3: recharge; in <sleep>, topic 1: smartphone use, topic 2: sleep insufficiency, and topic 3: sleeping time; and in <stress>, topic 1: onset and treatment, topic 2: healing, and topic 3: stress reduction. As a result of looking at the trends of topic using regression analysis and Hot_Cold analysis, there was a significant difference in Hot(positive correlation) and Cold(negative correlation) partially by item ($p < .05$). In conclusion, this study identified the trends of health care by period through keyword and topic analyses using big data and provided health care information reflecting the characteristics of the times and is expected to be used as basic data for future health policy and tasks.