


빅데이터 텍스트마이닝을 이용한 초기상담 기능강화 방안

최항석·김성훈



머리말

빅데이터(Big data) 분석은 기존의 데이터 분석의 한계를 극복하기 위해 대용량의 정형·비정형 데이터를 수집, 분석하여 숨겨진 정보와 규칙을 탐색해, 정책적으로 가치 있는 결과를 도출하고자 최근 들어 널리 사용되는 기법이다. 최근에는 행정데이터와 빅데이터를 사회보장 분야에도 적용하여 의사결정 절차를 효율화하고, 사회의 요구와 위험 등을 선제적으로 파악하고자 적극, 대응하고 있다.

일반적으로 복지급여의 제공여부를 판단하기 위해서 상담이 수행되고, 이를 통해 정확한 욕구의 파악과 적절한 서비스의 제공이 이루어진다.

공공분야의 복지전달체계의 경우, 읍면동 사무소에서 사회복지통합 업무의 ‘신청’ 단계 이전에 ‘초기상담’을 통하여 대상자의 기본적인 욕구를 파악하고, 시군구에서 복합욕구를 종합적으로 상담과 관리하는 ‘통합사례관리’를 수행하는 방식으로 이루어지고 있다. 여기에 더해 최근 ‘복지허브화’ 정책이 도입되면서 읍면동 사무소에서 사례관리의 역할을 담당하게 되어 복지전달체계에서 상담의 역할을 읍면동으로 집중시켜 업무체계를 강화하면서 일원화하였다. 여기서 말하는 상담은 상담자와 내방객간의 관계를 통하여 내방객이 가진 문제를 해결하기 위한 방안을 모색하는 것이라 할 수 있다.

본 연구에서는 상담의 전문성을 확보하기 위해서 지속적인 교육,

상담 매뉴얼 개발이 필요하다는 문제의식 하에서, 대상자의 욕구를 파악할 수 있는 텍스트를 수집하여 그것들을 텍스트마이닝 기법으로 적절한 욕구들을 추출하여 상담기능을 강화할 수 있는 방안을 제시하고자 하였다.

본 연구는 최항석 박사와 김성훈 연구원에 의해 수행되었는데, 연구계획과 결과의 활용 등에 관해서 다양한 의견을 주신 전문가와 관련 담당자께 감사의 말씀을 드린다.

마지막으로 본 연구의 주요 연구결과가 정책전문가와 관련 연구자들에게 유용하게 활용되길 바라며, 본 연구의 모든 내용은 우리원의 공식적인 견해가 아니며 참여한 연구진의 의견임을 밝힌다.

2016년 12월
사회보장정보원장

임 병 인

목 차

요 약	i
제1장 서론	1
제1절 연구의 배경	1
제2절 연구의 목적	5
제2장 통합사례관리의 현황과 상담의 역할	11
제1절 공공분야 사회복지 전달체계 개편과 상담기능의 강화	11
제2절 희망복지지원단과 통합사례관리	18
제3절 초기상담의 역할과 기능	21
제3장 빅데이터 분석과 텍스트마이닝	27
제1절 빅데이터 분석	27
제2절 데이터마이닝과 텍스트마이닝	31
제3절 텍스트마이닝 활용사례	36
제4장 복지영역별 토픽 분석	43
제1절 토픽 분석의 방법과 절차	43
제2절 복지영역별 텍스트마이닝 분석결과	50
제3절 초기상담 기능강화 방안	110

제5장 결론	117
--------------	-----

참고문헌	121
------------	-----

표 목 차

<표 1-1> 맞춤형 복지팀 담당별 통합사례관리 역할 구분	5
<표 2-1> 읍면동 복지기능 강화 시범사업 주요내용	13
<표 2-2> 희망복지지원단의 주요업무	19
<표 3-1> 빅데이터를 분석하기 위한 시각의 변화	27
<표 4-1> TF(단어 빈도)의 수학적 측정 방법	47
<표 4-2> IDF(역문서 빈도)의 수학적 측정 방법	47
<표 4-3> 복지영역별 포털사이트 검색 현황	50
<표 4-4> ‘임신출산’ 관련 TF-IDF를 고려한 다빈도 형태소	53
<표 4-5> ‘임신출산’ 관련 다빈도 형태소별 유사도	54
<표 4-6> ‘영유아’ 관련 TF-IDF를 고려한 다빈도 형태소	56
<표 4-7> ‘영유아’ 관련 다빈도 형태소별 유사도	57
<표 4-8> ‘아기’, ‘아이’ 관련 키워드 유사도	60
<표 4-9> ‘아동청소년’ 관련 TF-IDF를 고려한 다빈도 형태소	61
<표 4-10> ‘아동청소년’ 관련 다빈도 형태소별 유사도	62
<표 4-11> ‘어린이’ 관련 다빈도 형태소별 유사도	64
<표 4-12> ‘청년’ 관련 TF-IDF를 고려한 다빈도 형태소	65
<표 4-13> ‘청년’ 관련 다빈도 형태소별 유사도	66
<표 4-14> ‘중장년’ 관련 TF-IDF를 고려한 다빈도 형태소	68
<표 4-15> ‘노년’ 관련 TF-IDF를 고려한 다빈도 형태소	69
<표 4-16> ‘노년’, ‘노인’ 관련 다빈도 형태소별 유사도	70
<표 4-17> ‘장애인’ 관련 TF-IDF를 고려한 다빈도 형태소	72
<표 4-18> ‘장애인’ 관련 다빈도 형태소별 유사도	73
<표 4-19> ‘한부모’ 관련 TF-IDF를 고려한 다빈도 형태소	75
<표 4-20> ‘한부모’ 관련 다빈도 형태소별 유사도	76
<표 4-21> ‘다문화(새터민)’ 관련 TF-IDF를 고려한 다빈도 형태소	79

<표 4-22> '다문화', '새터민' 관련 다빈도 형태소별 유사도	80
<표 4-23> '저소득' 관련 TF-IDF를 고려한 다빈도 형태소	83
<표 4-24> '저소득' 관련 다빈도 형태소별 유사도	84
<표 4-25> '교육' 관련 TF-IDF를 고려한 다빈도 형태소	86
<표 4-26> '교육' 관련 다빈도 형태소별 유사도	87
<표 4-27> '고용' 관련 TF-IDF를 고려한 다빈도 형태소	90
<표 4-28> '고용' 관련 다빈도 형태소별 유사도	91
<표 4-29> '주거' 관련 TF-IDF를 고려한 다빈도 형태소	94
<표 4-30> '주거' 관련 다빈도 형태소별 유사도	95
<표 4-31> '건강' 관련 TF-IDF를 고려한 다빈도 형태소	98
<표 4-32> '건강' 관련 다빈도 형태소별 유사도	99
<표 4-33> '서민금융' 관련 TF-IDF를 고려한 다빈도 형태소	101
<표 4-34> '생계자금', '주거자금', '고용지원', '학자금' 관련 유사도	102
<표 4-35> '문화' 관련 TF-IDF를 고려한 다빈도 형태소	105
<표 4-36> '문화' 관련 다빈도 형태소별 유사도	105
<표 4-37> 통합사례관리 10대 복지욕구 영역	112

그림 목 차

[그림 1-1] 2016년 보건복지부의 목표와 3대 과제	1
[그림 1-2] 읍면동 복지허브화 4대 추진전략	2
[그림 1-3] 사회복지 통합업무 절차	5
[그림 2-1] 복지제도 확산을 위한 중점과제	11
[그림 2-2] 129콜센터(왼쪽), 복지로(오른쪽) 홈페이지	15
[그림 2-3] 통합사례관리 절차	20
[그림 2-4] 사회보장정보시스템(행복e음) 욕구조사 관리화면	21
[그림 2-5] 통합사례관리 기본형 흐름도	22
[그림 2-6] 사회보장정보시스템(행복e음) 초기상담 기본화면	23
[그림 3-1] 데이터마이닝 기본절차	32
[그림 4-1] 복지로 '한눈에 보는 복지정보' 복지영역별 분류	43
[그림 4-2] 복지영역별 토픽 분석 수행 절차	45
[그림 4-3] Word2Vec 알고리즘의 아키텍처	48
[그림 4-4] 텍스트마이닝 결과를 이용한 상담 절차	111

요 약

제1장 서론

- 정부는 2016년 보건복지 분야 ‘맞춤형복지’ 계획을 수립하고 ‘맞춤형복지 체감도 제고’를 목표로 함
 - 3대 과제로 ‘읍면동 중심의 복지전달체계 구축’, ‘맞춤형 복지 제도 내실화’, ‘복지사각지대 적극 해소’를 마련함
- ‘읍면동 중심의 복지전달체계 구축’은 ‘복지허브화’를 통하여
 - 전국의 읍면동이 복지기능을 강화해 인적안전망을 구축하고
 - 민간의 복지자원을 활용하여 통합서비스를 제공함
- 읍면동 ‘복지허브화’로 읍면동 사무소의 기능을 개선·강화하여 통합서비스를 지원함
 - 주요 추진전략으로 ‘맞춤형 복지 전담팀 구성 및 운영’, ‘복지 인력 전문성 강화 및 처우 개선’ 그리고 ‘민관 협력으로 공공 복지 보완 및 강화’를 수립함
 - 전문성을 갖춘 전담팀을 운영해 고품질의 복지서비스 제공함

- 더불어 공공분야 복지전달체계에서 ‘상담’의 역할이 시군구에서 읍면동으로 전환되어 전문성을 갖춘 인력이 필요함
 - 전문적인 ‘상담’의 능력을 단시간에 갖추기는 어려움
 - 결국, 표준화된 업무절차와 지침은 상담업무 수행에 필수적임
- ‘맞춤형 복지팀원’의 역할은 ‘사례관련 대상자 발굴’, ‘초기상담 및 욕구조사’, ‘서비스 계획 수립 및 실행’, ‘사후 모니터링’임
 - 이는 대상자 발굴과 서비스 제공·연계의 효과적인 업무수행을 위한 기능 재편임
 - 특히 ‘초기상담 및 욕구조사’는 대상자의 상태를 파악할 수 있는 중요한 기능임
- 한편 통합사례관리는 다양한 욕구를 가진 대상자의 복지급여를 종합적으로 제공 및 연계하는 등의 역할을 수행하는 것임
 - ‘복지허브’로 읍면동에서 초기상담과 통합사례관리를 동일 선상에서 수행하여 효율적인 업무처리 체계를 갖춘
 - 즉, 초기상담에서 정확한 욕구를 파악해 서비스를 연계하고 복합욕구 대상자는 통합사례관리로 연계해 기능을 수행함
 - 이를 통해 복지체감도 향상, 맞춤형 서비스 제공, 복지서비스 내실화, 행정업무의 간소화 가능함
- 본 연구의 주요 내용은 복지영역별(임신출산, 영유아, 건강 등) 텍스트를 수집하고 토픽을 분석해 상담 기능강화 방안 제언
 - 상담 영역별 다빈도 단어, 연관성 높은 토픽 등을 분석
 - 이를 통한 상담 매뉴얼과 효율적인 상담의 절차 마련, 10대 욕구조사 분류 세분화 등의 활용 방안을 제언함

제2장 통합사례관리의 현황과 상담의 역할

- 정부는 생애주기별 맞춤형 복지제도 체계의 완성과 국민들의 만족도를 높이기 위해 3대 중점과제를 마련함
 - ‘읍면동 중심 복지전달체계 구축’, ‘맞춤형 복지제도 내실화’, ‘복지 사각지대 적극 해소’가 3대 과제임
 - 읍면동 기능강화로 읍면동 중심의 원스톱 서비스를 제공함
 - 즉, 시군구와 읍면동으로 이원화 된 전달체계를 읍면동 중심으로 통합해 효율적으로 업무를 추진함
- 읍면동 중심 복지전달체계 구축을 위해 ‘읍면동 복지기능 강화 시범사업’을 추진(2014.7.~2015.12.)
 - 주요사업 내용은 상담의 기능강화, 서비스 연계와 효과적인 자원의 관리 운영이 핵심임
 - 결국, 상담기능 강화를 위한 인력의 배치와 적합한 서비스의 제공으로 복지 사각지대 발생을 줄이고자 함
 - 2016년에는 700여개 읍면동에서 본격적 추진하고, 2018년에 전국에서 수행하여 복지체감도 향상과 이용 효율성을 높임
- 사회복지통합업무절차는 급여를 제공받기 위해서 필수적으로 수행되는 과정(신청 → 조사 → 보장결정 → 급여 → 변동·사후관리)
 - 신청단계에서 상담이 진행되며 이를 초기상담이라 함
 - 초기상담은 지원대상자의 주요 문제와 욕구를 파악하고, 사회보장급여 제공의 필요성을 판단하고 적합한 서비스를 안내함
 - 공공분야 서비스가 어려운 경우 민간 서비스 연계도 지원함

- 희망복지지원단은 복합적 욕구의 대상자에게 통합사례관리를 수행하고 지역단위 통합서비스를 제공하는 조직임
- 업무체계는 ‘대상자 발굴’, ‘통합사례관리 실시’, ‘자원관리’임
 - 읍면동에서 초기상담을 수행 후 사례관리 대상자로 등록
 - 복합 욕구를 충족시키는 서비스 제공, 점검, 사후관리
 - 지역 사회의 자원의 총괄 관리와 조사 발굴을 수행함
- 특히, 통합사례관리는 희망복지지원단의 핵심 업무로 다양한 욕구의 대상에게 통합적인 서비스를 제공하여 안정적인 삶을 지원하여 복지제도의 효과성을 제고함
- ‘복지허브화’로 기존 시군구의 통합사례관리가 읍면동 중심으로 역할을 수행하게 됨
- 초기상담의 주요영역은 안전, 건강, 가족관계 등 10개 영역의 복지욕구로 이는 통합사례관리의 조사내용과 동일
- 결국 초기상담의 내용을 기준으로 통합사례관리가 수행될 수 있어 초기상담의 역할의 중요성이 증대됨

제3장 빅데이터 분석과 텍스트마이닝

- 빅데이터(Big Data)는 자료의 복잡성이 증가해 기존의 데이터 처리와 관리 도구로 다루기 어려운 자료의 집합임
- 빅데이터의 출현 배경으로
 - 게놈 프로젝트와 같은 빅데이터를 이용한 연구의 증가
 - 디지털화 등 정보기술의 발전으로 자료의 양 급증
 - 다양한 고객 정보가 축적되어 가치를 창출 기회가 확대됨

- 빅데이터를 분석하는 관점의 3가지 변화
 - 대상 측면에서 통계적 표본에서 전체 대상으로 확대
 - 자료 측면에서 데이터의 질보다는 양의 증가
 - 모델 측면에서 인과관계 보다는 상관관계 연구로 변화됨
- 빅데이터 분석은 비즈니스 모델 개발에 활용되며 연관규칙학습, 기계학습, 감성분석, 소셜네트워크분석 등의 방법이 있음
- 최근 비정형 데이터 분석에 관심이 증가하여 텍스트, 영상 등의 연구가 증대되고 있음
 - 특히, 텍스트 분석은 비정형인 문서에서 자연어 처리 기술에 기반을 두어 정보를 추출 및 가공하는 방법으로
 - 소셜 미디어 등의 텍스트의 긍정, 부정 선호도 파악, 상품의 시장규모, 소비자 반응 및 입소문 분석 등에 활용됨
- 데이터마이닝은 대용량의 데이터에서 관계, 패턴, 규칙을 탐색하고 모델링하는 일련의 과정임
 - 통계학의 데이터 탐색, 모델 구축 등의 개념은 동일하나
 - 데이터가 방대하고 업무가 복잡해 전문가의 통찰력에 한계가 발생하여 데이터에 기반을 두어 지식 패턴을 추출하게 됨
 - 다양한 데이터에서 정보의 일반화 또는 예측에 초점을 둠
- 통계학의 기본적인 분석 데이터는 정형화된 수치 및 범주이며, 데이터마이닝도 정형화된 데이터를 주로 다루었음
- 그러나 비정형 데이터의 분석도 활발한데, 이는 텍스트, 영상, 음성 등 비정형 데이터의 처리와 분석방법의 발전에 기인하며 이를 비정형데이터마이닝이라고 함

- 비정형데이터마이닝 중 텍스트마이닝은 문서의 텍스트를 인식하여 자동으로 정보를 추출 및 패턴을 분석하는 것으로
 - 텍스트마이닝 전처리 과정에는 텍스트를 정형화된 데이터로 변환하는 과정이 포함되며 분석에서 큰 비중을 차지함
 - 분석 기법으로 ‘동시단어’, ‘문헌키워드’, ‘계량서지학’ 등의 연구동향 및 현황 분석과 ‘네트워크 분석’, ‘토픽 모델링’ 등 예측과 일반화 방법이 알려짐
- 텍스트마이닝의 활용사례는 다양한 분야에서 소개되고 있는데, 일반적으로 마케팅 전략 계획, 고객의 신용평가, 환자의 질병 예측 등에 이용
 - 국민권익위원회의 ‘민원정보분석시스템을 통한 국민과 정보의 소통 활성화’로 민원 자료를 수집해 지원정보의 획득과 민원 분석시간 단축 등의 업무 효율화
 - 한국도로공사의 ‘고객목소리분석 시스템을 통한 서비스 혁신’으로 키워드와 토픽을 추출해 연관성과 추이 분석하여, 주요 이슈 사항 파악 및 대응 등의 의사결정 지원
 - 국민건강보험공단의 ‘국민건강주의 예보 시범서비스 구축’은 전염병이 유행할 때 사전에 알림으로 경제적 비용의 발생을 최소화를 목적으로 예측과 예방 가능한 서비스 제공
 - 국외에서 ‘SNS를 활용한 할리우드 흥행 수익 예측’은 SNS의 관심 내용의 텍스트를 분석해 시장을 예측하는 것임
 - 코카콜라의 ‘SNS 데이터 활용을 통한 가치 향상’은 트위터와 페이스북에서 자사와 관련된 데이터를 수집·분석하여 제품의 판매와 관련된 의사결정을 지원함

제4장 복지영역별 토픽 분석

□ 복지영역별 토픽 분석으로

- 복지로의 ‘한눈에 보는 복지정보’의 임신출산, 아동청소년 등 16개 영역별로 다빈도 단어와 연관성 높은 토픽을 분석함
- 이 영역은 생애주기별, 인구집단별, 생활영역별로 범주화됨

□ 분석을 위한 데이터는 포털사이트 상세검색(2006.1.~2016.6)의 결과를 웹 크롤링을 통해 수집함

- 분석 절차는 ‘자료수집 → 전처리 → 형태소분석 → 다빈도분석 → TF-IDF → Word Embedding’으로 구분됨
- 자료수집에서 16개 복지영역별로 ‘복지’가 반드시 포함되는 텍스트를 약 10년간 768,124건 수집함
- 전처리, 형태소분석에서는 수집된 데이터에서 오타자 검사와 형태소로 분해해 유의어, 동의어 등의 변환을 수행함
- 다음으로 관심 및 핵심어 선정을 위한 다빈도와 TF-IDF의 분석 및 결과 확인 후 Word Embedding으로 기준 단어와 연관성 높은 토픽을 이끌어 냄

□ 분석결과는 16개 영역별로 TF-IDF를 고려한 다빈도 형태소와 다빈도 형태소 중심의 유사도를 각각 제시함

- ‘임신출산’ 영역의 결과로
 - ‘시댁’, ‘와이프’ 등의 가족 관련된 형태소가 높은 순으로 나타났고, ‘양육’, ‘휴직’ 등 복지정책 이슈, 이혼(위자료), 미혼모 등 사회적 이슈가 나타남

- 유사도는 ‘출산’, ‘육아’ 관련하여 퇴직이나 권고사직 등의 이슈와 나타나고, ‘미혼모’의 경우 보호시설과 한부모가정, 기초생활수급 등 복지급여 관련 형태소가 확인됨

○ ‘중장년’ 영역은 수집된 텍스트가 2,163 건이며

- 이를 분석한 결과로 ‘모친’, ‘형님’, ‘자식’ 등 가족관계의 형태소와 ‘사주’, ‘운’, ‘삶’ 등의 불확실한 인생의 고민을 내포하는 형태소가 다수 나타남
- 한편 유사도의 결과는 분석에 이용된 텍스트의 수가 적어 모델이 수렴하지 않아 결과가 도출되지 않음

○ ‘한부모’ 영역의 결과는 복지급여와 관련된 것이 분석되었고, 특히 급여 신청에 필요한 서류 등에 대한 형태소도 있음

○ 이외의 나머지 영역에서도 분야별 사회, 복지정책, 전문분야, 주거환경 등의 이슈가 나타남

□ 데이터 수집과 분석의 결과를 통해 도출된 한계로

- 불특정 다수가 대상인 인터넷 문서를 수집해 분석한 결과로 오타자, 은어, 단축단어 등으로 모델의 적합에 약점이 있고
- 포털사이트의 특성상 검색어와 관계가 없는 문서들도 상당수 포함되어 일반화에 어려움이 있으며
- 특히 인터넷을 사용하는 연령대, 소득수준이 급여의 대상자로 규정하기에는 부족한 점도 있음

□ 분석결과로 대상자의 욕구를 정확하게 파악할 수 있는 정보의 구축을 통해서 이를 활용하는 방안으로

- 복지영역별 요구사항 등을 파악하여 복지급여 발전방안, 신규 서비스 개발 등에 적용할 수 있으며

- 복지로 ‘전문가상담’, ‘관련정보’를 이용자의 요구사항에 맞게 정보를 제공해 이용률 제고와 검색 기능을 강화하며
- 초기상담 등 전문성이 필요한 상담에서 유의어를 시스템화해 대상자의 욕구를 질문 및 유의어 검색으로 유도할 수 있으며
- 토픽을 이용해 통합사례관리 10대 복지욕구 영역을 세분화해 구체적인 상담을 수행함

제5장 결론

- 정부는 ‘읍면동 중심의 복지전달체계 구축’을 위해
 - ‘복지허브화’를 추진해 원스톱으로 통합서비스를 제공받도록 읍면동 중심의 전달체계를 지속적으로 강화함
- 읍면동을 공적 부조의 집행에서 통합서비스 기관으로 전환은
 - 현장 맞춤형 통합서비스 기능을 강화하는 방안으로 재편하고
 - 읍면동에 맞춤형 복지전담팀을 설치해 방문상담과 사례관리 수행으로 상담의 중요성을 증대함
- 초기상담은 급여를 제공받기 위해 신청단계에서 대상자를 상담하는 것으로
 - 대상자 및 가구의 특성과 복지욕구 및 서비스 파악과
 - 지원 가능한 서비스 안내와 신청할 수 있는 서비스 지원함
- 정확한 욕구와 적절한 복지급여 안내가 중요하며
 - 이를 통해 복지체감도 향상, 맞춤형 서비스 제공, 복지서비스 내실화, 행정업무의 간소화가 가능함

- 본 연구에서는 복지영역별(임신출산, 영유아 등 16개 영역)로 텍스트마이닝을 이용한 정보의 정형화 수행함
- 분석의 방법으로 복지영역별 중요도를 고려한 다빈도 형태소 도출과 이 형태소와 연관성 높은 토픽을 정보화함
- 분석 결과를 활용하는 방안으로
 - 초기상담에서 대상자의 언어 및 전달 내용을 적극적으로 이해하여 상담업무의 전문성을 높이고
 - 복지로 이용률 향상을 위한 사용자 요구에 부응하는 정보 추가적인 구축과
 - 현재 복지사업의 개선점 등의 복지영역별 관심 및 요구를 반영한 정책 의사결정 지원이 있음
- 본 연구는 텍스트마이닝을 이용한 복지영역별 다빈도 단어와 토픽 등의 연계어 도출로 정보화를 하여 읍면동 초기상담 기능 강화 등의 활용방안을 제언함
- 그러나 데이터의 한계로 직접적인 활용보다는 거시적인 관점으로 적용방안을 기술하였으며
- 분석결과와 효과적인 활용을 위해서 희망복지지원단의 통합 사례관리 욕구상담 자료를 이용한 분석이 요구됨

서 론

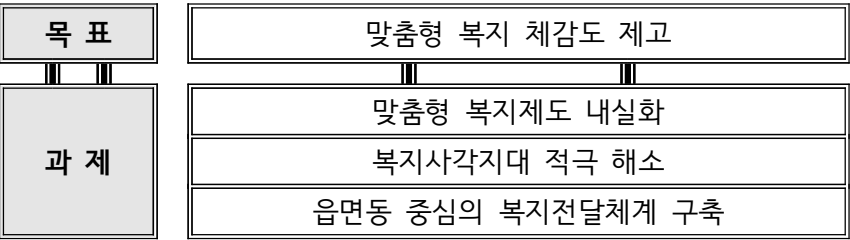
제1절 연구의 배경

제2절 연구의 목적

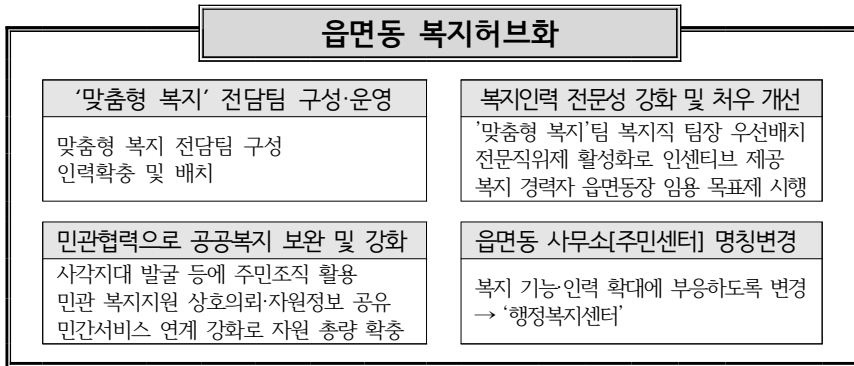
제1장 서론

제1절 연구의 배경

2016년 정부는 국민행복 분야 합동 업무보고 자리에서 보건복지부, 교육부, 고용노동부, 여성가족부가 함께 ‘청년일자리 창출 및 맞춤형 복지’ 계획을 밝혔다. 각 부처별로 관련 정책을 발굴하고 업무 계획을 수립하기 위해 각종 정책간담회, 현장방문 등으로 전문가와 국민의 의견을 수렴하였다. 그리고 보건복지부는 [그림 1-1]과 같이 ‘맞춤형 복지 체감도 제고’라는 목표를 세우고, 이를 달성하기 위해서 ‘맞춤형 복지제도 내실화’, ‘복지사각지대 적극 해소’ 그리고 ‘읍면동 중심의 복지전달 체계 구축’의 3대 과제를 마련하였다.



[그림 1-1] 2016년 보건복지부의 목표와 3대 과제



[그림 1-2] 읍면동 복지허브화 4대 추진전략

전술한 과제의 내용을 살펴보면, 우선 ‘맞춤형 복지제도 내실화’는 생애주기별 맞춤형 복지의 실질적인 체감도를 높이기 위해서 맞춤형 급여체계로 개편한 기초생활보장제도를 안정적으로 정착시키고, 4대 중증질환 등과 같은 의료보장을 지속적으로 추진한다. 그리고 맞춤형 보육으로의 개편과 기초연금, 장기요양 지원을 지속적으로 확대한다. 둘째, ‘복지사각지대 적극 해소’는 복지급여 및 서비스와 같은 정책적 변화 등을 온·오프라인을 활용하여 보다 적극적으로 홍보하고, 지역의 다양한 자원과 인적 네트워크의 확충 및 정부 3.0의 정보 공유·개방을 통해 복지 사각지대 대상자 및 위기가구를 우선적으로 발굴하고 독거 노인, 결손가정 등의 취약계층을 선제적으로 지원한다. 마지막으로 ‘읍면동 중심의 복지전달체계 구축’은 ‘읍면동 복지허브화’를 통하여 전국의 모든 읍면동이 강화된 복지기능으로 인적안전망을 구축하고, 민간의 복지자원을 활용하여 통합서비스를 제공한다.

‘읍면동 복지허브화’를 위한 4대 추진전략은 [그림 1-2]와 같이 ① 맞춤형 복지 전담팀 구성 및 운영, ② 복지인력 전문성 강화 및 처우 개선, ③ 민관협력으로 공공복지 보완과 강화, ④ 읍면동 사무소[주민

센터] 명칭변경이다. 이를 통하여 기존 읍면동 사무소의 기능인 ‘복지 서비스 신청 및 접수’를 개선 및 강화하여 ‘통합서비스 제공’이 가능하도록 하였다. 이는 국민이 보다 손쉽게 접근할 수 있고, 체감할 수 있는 현실적인 맞춤형 복지를 제공하고자 하는 것이다.

이중에서 ‘복지 전담팀 구성 및 운영’으로 국민이 체감할 수 있는 양질의 통합서비스를 제공하기 위해 ‘찾아가는 서비스 활성화’, ‘통합 서비스 지원¹⁾’, ‘민간조직·자원 적극 활용’을 수행한다. 그리고 ‘복지 인력 전문성 강화 및 처우 개선’은 통합조사와 사례관리 등의 업무로 전문성을 갖춘 인력을 중심으로 전문상담을 지원한다.

즉, ‘읍면동 복지허브화’는 전국의 읍면동 주민복지센터를 맞춤형 복지서비스를 제공하기 위한 통합 창구의 기관으로 개편하는 것이다. 2014~15년 전국의 15개소 읍면동에서 실시한 시범사업의 성과로는 사각지대 발굴 건수 6.2배, 방문상담 건수 3.3배, 서비스 연계 건수가 3.4배 증가하였고, 이용자의 만족도도 82.0점에서 92.8점으로 증가하였다. 정부는 시범사업을 통하여 찾아가는 서비스, 사례관리 등 복지 기능이 강화되는 읍면동 사무소를 전국으로 확산하는 계획을 세우고 있다(2016년 700개 읍면동 → 2018년 전국 시행).

기존의 공공분야의 복지전달체계에서 읍면동은 사회복지통합업무의 ‘신청’ 단계 이전에 ‘초기상담’을 통하여 대상자의 기본적인 욕구를 파악하였고, 시군구의 희망복지지원단에서는 대상자의 다양한 욕구를 종합적으로 상담 및 관리하는 ‘통합사례관리’를 수행하였다. 그러나 ‘읍면동 복지허브화’를 통해 읍면동에서 사례관리의 역할을 수행하게 되었다²⁾. 이는 공공분야 복지전달체계에서 상담의 역할이 시군구에서

1) 대상자별로 욕구에 맞게 필요로 하는 서비스를 맞춤형으로 제공하고, 가구별 서비스 제공 계획을 수립 및 연계 등으로 통합사례관리를 수행

2) 위기사례의 경우에는 읍면동이 아닌 시군구 희망복지지원단에서 수행

읍면동으로 전환을 의미하며, 상담에서 담당자가 갖추어야 할 자질의 전문성도 요구된다.

한편 사회복지 실천현장에서의 ‘상담’은 상당한 전문성이 요구되는 것으로 단시간 내에 일반 읍면동 담당공무원의 상담 능력을 향상시킨다는 것은 쉽지 않다. 따라서 상담에 필요한 표준화된 절차와 안내 내용은 업무의 수행을 위해서 담당자에게는 필수적이다. 그래서 기존의 상담전문 담당공무원이 수행한 희망복지지원단의 통합사례관리 상담 노하우(Know-How)를 읍면동의 상담업무에 적용할 수 있다면 앞서 언급한 읍면동 전달체계 개편과정에서의 상담과 사례관리의 안정적인 정착이 가능할 것이다. 결국 비정형인 상담사례(텍스트자료)의 욕구 관련 데이터를 정형화하는 정보화의 노력이 필요하다. 이를 통하여 상담 담당자에게 도움이 되는 정보를 제공할 수 있고, 정보화된 욕구 기반의 데이터를 활용하여 기존 복지서비스의 개선 방안 마련과 신규 서비스의 발굴 및 개발도 가능할 것이다.

제2절 연구의 목적

2016년 읍면동 ‘맞춤형 복지팀원’의 통합사례관리 업무에서 역할은 <표 1-1>에서처럼 ‘사례관련 대상자 발굴’, ‘초기상담 및 욕구조사’, ‘서비스 계획 수립 및 실행’, ‘사후 모니터링’이다. 기본적으로 상담과 서비스계획을 중심으로 역할이 구성되었고, 이 중에서도 ‘초기상담 및 욕구조사’는 ‘맞춤형 복지팀원’의 고유한 역할이라고 할 수 있다. 즉, 대상자 발굴과 서비스 제공 및 연계의 접점에서 가장 활발한 기능의 수행이 요구된다.

<표 1-1> 맞춤형 복지팀 담당별 통합사례관리 역할 구분

읍면동장	맞춤형 복지팀장	맞춤형 복지팀원	지역사회 보장협의체
·사례관리 대상자 발굴	·사례관리대상자 발굴 ·사례회의 총괄 ·사례판정과 배분 ·사례관리대상자 발굴 ·서비스 연계 지원	·사례관리대상자 발굴 ·초기상담 및 욕구조사 ·서비스계획수립 및 실행 ·사후 모니터링	·사례회의 참여 ·자체복지자원연계

공공분야 복지전달체계 내에서 사회복지 통합업무 절차는 [그림 1-3]과 같고, ‘사회복지 통합업무 안내지침’에 제시되어 있듯이 크게 ‘신청-조사-보장결정-급여·서비스실시-변동·사후관리’의 순으로 진행된다.



[그림 1-3] 사회복지 통합업무 절차

이중 ‘신청’단계인 상담의 목적은 대상자가 처해있는 상황과 욕구를 정확히 파악하여 적합한 서비스를 연계하는 것이라 할 수 있다. 이런 과정에서 상호 신뢰를 쌓을 수 있다면 복지 체감도 제고에 기여할 수 있으며, 맞춤형 서비스 제공과 함께 복지 급여에 대한 내실화 그리고 전문적인 상담을 통한 행정업무가 진행되어 효율적인 업무의 수행도 가능할 것이다.

특히 초기상담이란 지원대상자의 현재 상태와 주요 문제 및 욕구를 파악하여 복지서비스의 필요성 여부를 판단하고, 대상자에게 적합한 복지서비스를 안내하는 것이 주요 역할이라고 할 수 있다. 기본적인 상담내용은 지원대상자 및 가구의 특성 파악으로 인적사항, 가구원, 부양의무자 유무 등과 지원대상자가 가지고 있는 주요 문제 및 복지 욕구 등이다.

한편 통합사례관리는 다양한 욕구를 가진 지원대상자에게 복지급여 및 서비스를 종합적으로 제공 및 연계하고, 이를 관리하는 역할들을 수행한다. 통합사례관리는 기존에 읍면동에서 수행하기 어려웠으나, ‘읍면동 복지허브화’라는 전달체계의 개편으로 읍면동에서 초기상담과 함께 통합사례관리를 동일 선상에서 수행하게 되어 효율적인 업무의 체계를 마련하였다. 그러나 앞에서 언급한 것처럼 상담은 전문성을 필요로 하는 분야이기 때문에 단순히 조직을 개편하고, 인력증원 및 재배치로만으로 상담의 전문성을 확보하는데 한계가 있다.

상담 담당자의 노하우를 마련하고 수준을 높이기 위해서 상담자를 대상으로 지속적인 교육과 상담 매뉴얼, 가이드라인을 개발하여 제공하는 것이 필요하다. 즉, 체계적인 상담절차와 사례들이 개발되어야 하는데 신규업무는 유사한 업무를 기반으로 개발이 가능하나 업무의 안정화를 위해서는 많은 노력이 필요하다. 한편 기존 업무의 개선을 위해서는 그동안 축적된 업무(상담사례)의 자료 등을 활용하는 것이

효과적이다. 그리고 초기상담과 통합사례관리의 상담 내용은 사회보장 정보시스템에 기록된다. 기록된 자료는 텍스트 기반의 정보로 이를 빅데이터 방법론 중 텍스트마이닝을 이용하여 정보의 일반화(키워드, 토픽, 연관 단어) 등이 가능하다. 적용된 결과를 이용해 개선된 상담 절차와 서비스를 개발할 수 있다.

본 연구는 상담을 위한 매뉴얼과 효율적 상담의 절차 마련을 위한 것으로 복지영역별(임신출산, 건강, 영유아, 문화 등 복지로에 분류된 16개 영역)로 관련된 텍스트를 수집(웹 크롤링)하여 각 영역별 관심 사항과 관련된 토픽들을 텍스트마이닝 방법론을 적용하여 분석하고자 한다. 이를 통해서 현재의 초기상담 및 사례관리의 10대 욕구조사의 분류를 세분화하는 방안을 제시하여 상담에 활용할 수 있도록 하는데 연구에 초점을 둔다.

본 연구에서는 우선 제2장에서 읍면동 복지허브화에서 초기상담과 통합사례관리의 기능 및 역할 등의 현황에 대해서 살펴보고자 한다. 다음으로 제3장에서 빅데이터 분석에서 텍스트마이닝과 분석에 적용하기 위한 방법론과 활용사례를 제시하고, 제4장에서는 텍스트마이닝 분석결과 제시와 이를 활용한 초기상담 기능강화 방안을 기술한다. 마지막으로 제5장에서는 연구의 내용 요약과 텍스트마이닝을 적용해 상담사례를 활용할 수 있는 향후 과제 등을 정리한다.

통합사례관리의 현황과 상담의 역할

제1절 공공분야 사회복지 전달체계 개편과
상담기능의 강화

제2절 희망복지지원단과 통합사례관리

제3장 초기상담의 역할과 기능

제2장 통합사례관리의 현황과 상담의 역할

제1절 공공분야 사회복지 전달체계 개편과 상담기능의 강화

정부는 복지 체감도의 제고를 위해서 ‘맞춤형 복지제도 내실화’, ‘복지 사각지대 적극 해소’, ‘읍면동 중심 복지전달체계 구축’을 통해 국민들 누구나 현실적으로 체감하는 맞춤형 복지를 확산하고자 한다. 이는 기존의 영유아, 아동·청소년, 청장년, 노인의 생애주기별 맞춤형 복지제도 체계의 완성과 함께 국민의 만족도를 높이기 위한 정부의 중점과제이다.

맞춤형 복지제도 내실화	복지 사각지대 적극 해소
① 맞춤형 기초생활보장제도 정착	① 구석구석 복지안내 강화
② 4대 중증질환 등 의료보장 지속	② 정부3.0 위기가구 선제 발굴
③ 맞춤형 보육 개편	③ 취약계층 필수서비스 지속 확충
④ 기초연금 및 장기요양 지원 확대	④ 노후준비 등 불안요인 해소 지원
읍면동 중심 복지전달체계 구축	
① 읍면동 복지허브화	② 읍면동 중심 통합서비스 제공

[그림 2-1] 복지제도 확산을 위한 중점과제

[그림 2-1]에서 알 수 있듯이 ‘맞춤형 복지제도 내실화’, ‘복지 사각지대 적극 해소’는 ‘읍면동 중심 복지 전달체계 구축’을 기반으로 해 발전 및 완성될 수 있다. 즉, 읍면동의 복지기능을 강화하여 읍면동 중심의 원스톱 서비스와 찾아가는 서비스 제공 등을 통해 생애주기별 맞춤형 복지제도의 대국민 만족도를 높이고, 복지 사각지대를 적극적으로 해소하고자 한다. 이는 기존의 시군구와 읍면동으로 이원화 된 공공분야 복지전달체계를 주민과 인접한 읍면동을 중심으로 통합된 서비스를 제공하기 위한 전환이다. 더불어 생애주기별 맞춤형 복지 제도를 안착시키고, 위기가구의 선제적 발굴과 취약계층에게 필요한 서비스 확충을 통하여 복지 사각지대를 해소하고자 한다.

1. 읍면동 중심 복지전달체계 구축

2014년7월 ~ 2015년12월까지 ‘읍면동 복지기능 강화 시범사업’(15개소)이 추진되었다. 이는 전국의 읍면동 주민센터를 맞춤형 복지허브 기관으로 개편하기 위해 개발된 모델의 시범 및 확대 적용 가능성을 검토하고자 하는 것이 목적이다.³⁾ <표 2-1>과 같이 주요사업 내용은 ① 사회복지서비스 수급 대상자에 대한 상담의 기능 강화, ② 서비스 수요자의 사례특성에 부합하는 맞춤형 서비스의 제공, ③ 효과적인 사회복지서비스 공급을 위한 민관협력 활성화 추진, ④ 복지와 보건 서비스 연계의 활성화, ⑤ 합리적으로 기능을 배분해 서비스 공급의 전문성과 효율성을 제고하는 것이다. 이는 상담 기능강화와 서비스 연계 및 자원의 효율적 관리 운영을 기본적으로 내포하고 있다.

3) 보건복지부, 읍면동 맞춤형 복지팀 업무 매뉴얼(2016)

〈표 2-1〉 읍면동 복지기능 강화 시범사업 주요내용

구 분	주요 내용
상담 및 서비스 연계	· 사회복지서비스 수급 대상자에 대한 상담기능 강화
	· 서비스 수요자의 사례특성에 부합하는 맞춤형 서비스의 제공
효과적인 자원 활용	· 효과적인 사회복지서비스 공급을 위한 민관협력 활성화 추진
	· 복지·보건서비스 연계의 활성화
	· 합리적인 기능배분으로 서비스 공급의 전문성과 효율성 제고

특히, ‘사회복지서비스 수급 대상자에 대한 상담기능 강화’를 위해 상담역량과 경험이 풍부한 경력직 공무원을 복지코디네이터로서 배치하여 대상자의 다양한 복지욕구에 대응하고, 사례별 욕구에 부합하는 서비스 정보를 제공함으로써 복지 사각지대의 발생을 줄이고자 한다. 그리고 주요 사회복지서비스 기능을 분야 별로 분담함으로써 읍면동 복지담당 인력운영의 효율화 방안을 모색하고자 한다.

2016년에는 700여개 읍면동에서 통합사례관리를 본격적으로 추진하게 되었으며, 이를 위해 복지인력의 확충과 읍면동 배치, 복지담당 공무원 전문 교육이 추진된다. 이와 같은 정책적인 변화로 국민들은 인접한 읍면동에서 개인별 복합 욕구에 부합하는 통합서비스를 제공받게 되고, 복지 사각지대 발굴, 방문상담·모니터링을 통한 밀착지원 확대로 복지체감도 향상과 이용 효율성을 높이게 된다.

2. 맞춤형 복지제도 내실화와 복지 사각지대 적극 해소

정부는 읍면동 중심의 원스톱 서비스, 찾아가는 서비스의 복지업무 강화를 통해서 생애주기별 맞춤형 복지 제도에 대한 국민의 체감도를 높이고 복지 사각지대를 해소하고자 한다. 그리고 ‘맞춤형 복지제도 내실화’를 위해서 2016년도 기준 중위소득(2015년 대비 4% 인상)과 생계급여 수급자 선정 기준(중위소득의 29%로 확대)의 완화 등으로 기초생활보장 지원 대상자도 확대하였다. 따라서 맞춤형 급여 개편에 따른 수급자 규모, 생활실태 변화, 최저생계비 파악 등을 위해서 실태 조사를 하였으며, 이를 토대로 급여 기준의 적정성 등에 대한 평가를 하여 기초생활보장제도의 내실화와 정착에 노력한다. 또한 4대 중증 질환의 보장성을 강화 계획을 2016년 내 완결해 환자의 부담을 경감하는 등 생애주기별 보장성 강화를 이행한다. 그리고 보육서비스를 맞춤형으로 개편하여 보육수요에 맞게 질 높고 다양한 보육서비스를 제공할 수 있도록 하고, 안정적인 노후생활의 기반을 보장하는 기초연금의 지급과 요양 서비스의 다양화 및 치매노인 서비스의 확대를 통한 장기요양보험의 지원 등으로 노인 복지를 강화한다.

‘복지 사각지대 적극 해소’ 측면에서는 [그림 2-2]와 같이 온라인 서비스를 이용해 긴급복지 안내, 자살예방, 수화서비스 등 129콜센터 (<http://www.129.go.kr>)를 활용한 복지 상담기능을 강화하고, 복지로 (<http://www.bokjiro.go.kr>)를 통해 누구나 복지서비스 정보에 쉽게 접근할 수 있도록 하였다. 그리고 읍면동사무소 중심의 인적안전망의 전국적인 확산과 민간의 사회복지사 등 시설종사자를 적극 활용하여 사각지대의 발굴과 신고를 활성화 하였다. 그리고 공공 빅데이터를 활용하여 복지 사각지대 소외계층을 선제적으로 발굴하는 등 다양한 노력을 하고 있다.



[그림 2-2] 129콜센터(왼쪽), 복지로(오른쪽) 홈페이지

3. 사회복지통합업무절차 내에서의 상담기능

사회복지 통합업무절차는 사회보장정보시스템(행복e음) 구축과 함께 신청, 조사, 사후관리 등이 개인과 가구단위로 통합적으로 수행될 수 있도록 제도적 근거의 마련과 함께 효율적인 업무처리의 절차를 제시하였다. 그리고 각종 사회복지사업의 수행에서 사용할 공통 기준과 절차를 안내하여 사회복지사업별로 다양한 기준을 일선 담당공무원이 쉽게 이용할 수 있게 체계화하는 것에 목적이 있다.⁴⁾ 이를 달성하기 위한 기본적인 업무절차는 앞에서 설명하였듯이 [그림 1-3]과 같이 ‘신청 → 조사 → 보장결정 → 급여·서비스 실시 → 변동·사후관리’로 구성된다.

각 단계별로 상세하게 살펴보면, ‘신청’ 단계에서는 사회복지급여 또는 서비스를 받으려는 사람은 관계 법령에서 정하는 바에 따라서 국가나 지방자치단체에 신청하는 것을 말한다. 이때, 급여·서비스가 필요한 대상자의 욕구를 정확히 파악하고 필요한 서비스를 누락 없이

4) 보건복지부, 사회복지통합안내지침(2016)

제공하기 위해서 지자체에서는 욕구 등에 대한 통합조사를 실시하여 일괄신청 하도록 하고 있다. 따라서 읍면동의 사회복지업무 담당자는 제공 가능한 급여를 확인하고 제출 서류와 신청방법 등을 안내하고, 필요한 급여를 일괄적으로 신청 받고, 접수 하고 있다.

‘조사’ 단계에서는 신청서 및 소득재산신고서 등 신고내용 확인 및 보장가구 구성을 확정하고, 공적 자료 조회를 요청한 후 조회결과를 확인하여 반영한다. 사업별로 다른 소득·재산의 항목들과 조사방법을 표준화하여 한 번에 조사가 가능하도록 항목들의 값을 단일화했다. 통합조사 담당의 소득·재산 조사 수행에서 정보시스템을 이용해 조회 되는 소득·재산 공적 자료를 우선적으로 적용하고, 보유 여부만 통보 하여 가액산정이 필요한 경우 등은 조사·확인 후에 반영한다.

‘보장결정’ 단계에서는 조사 결과를 토대로 해당 사업팀이 소득과 재산조사 결과 및 기타 자격 요건을 조사·확인하고, 선정기준의 부합 여부를 종합적으로 파악해 보장여부·급여내용(급여액, 서비스 내용)을 결정하여 신청인에게 통지한다.

‘급여·서비스실시’ 단계에서는 통합조사팀이 소득·재산, 가구원 등 변동사항을 반영한 결과를 토대로 급여지급자료를 매월 정보시스템을 통해 자동 생성한다. 각 사업팀 급여 담당자는 자동 생성된 급여의 지급 내역을 확인하여 급여액을 확정된 후, 결재되는 급여 지급자료(PDF파일)를 지방재정시스템인 ‘e-호조’를 통하여 회계부서로 지급을 의뢰한다. ‘e-호조’를 통해서 의뢰된 급여내역은 ‘e-지로시스템’(금융결제원) 또는 각 시·군·구 금고시스템을 통해 암호화 파일을 전송해 급여를 입금한다. 수급자별로 급여는 1인 1계좌로 단일화하고 매월 계좌의 유효성을 확인(실명 인증) 후 지급한다. 그리고 복지서비스는 서비스 계획을 수립하고 서비스 제공기관을 통해 해당 서비스를 제공한다.

‘변동·사후관리’ 단계에서는 수급자 가구에 대하여 정기적인 확인조사를 실시하거나, 수급자의 신고 등에 의하여 수급자격, 급여액변동 여부 등을 지속적으로 확인조사 하는 것이다. 수급자의 거주지, 가구 구성, 근로능력, 취업상태, 소득·재산과 부양의무자의 소득·재산, 부양 능력 유무, 부양 여부 등 수급자의 수급자격과 급여종류, 급여액 보장 기관 등의 결정에 영향을 미치는 모든 변동사항을 확인한다. 그리고 급여·서비스의 적정성을 관리하기 위해서 다음 사항들의 사후관리를 실시하고 있다. 첫째, 누락과 추가 서비스를 안내하며, 중복되는 급여·서비스를 방지하고 있다. 즉, 복지대상자의 자격정보와 급여·서비스 제공내역을 분석하여 누락서비스를 발굴·지원하고, 부정·중복 지원을 차단하고 있다. 둘째, 소득·재산, 근로능력, 부양 관계 등에 대한 추가 확인이 필요한 중점관리 대상을 통합 관리하여 정기적으로 확인하고 있다. 셋째, 부정수급자 등에 대한 보장비용 표준회수절차를 마련하고 보장비용의 반환과 징수를 실시한다. 사회보장정보시스템에서는 소득·재산 변동의 경우, 공적자료의 변동이 반영되고 담당 공무원 확인이 필요한 사항(예, 가액산정이 필요한 경우 등)은 확인 후에 반영하도록 하고 있다. 가구원들의 인적 변동의 경우, 거주지 변동, 가구원 변동(사망, 말소 등) 시 알림 기능으로 신속히 변동사항을 반영해 변경에 따른 보장가구를 재구성한다.

사회복지통합업무절차는 급여를 제공받기 위해서 필수적으로 수행되는 과정이며 이 중 ‘신청’ 단계에서 대상자와의 상담이 수행되는데, 이를 초기상담이라고 한다. 초기상담은 지원대상자의 주요한 문제와 욕구를 파악하여 복지서비스 제공의 필요성을 판단하고, 적합한 복지서비스를 안내하는 것이다. 주요상담 내용은 지원대상자 및 가구의 특성(인적사항, 가구원, 부양의무자 유무 등)과 개인 및 가구의 주요 문제와 복지욕구, 지원 가능한 복지서비스와 해당 복지서비스의 지원

기준과 신청 안내이다. 즉, 상담을 통해서 가구 특성, 주요 문제 및 욕구 등을 종합적으로 충분하게 파악하여 필요한 서비스 유형을 판단하고 제공 가능한 복지서비스를 확인하여 안내하는 것이다. 그리고 복지서비스 신청에 필요한 서류 및 조사 시 요구되는 서류를 정확히 안내하여 대상자의 방문을 최소화하여, 담당 공무원의 업무 효율화를 제고한다. 추가적으로 공공분야의 복지서비스 제공이 어려운 경우에 가능한 민간 서비스 연계한다.

‘복지 체감도 제고’ 측면에서의 초기상담은 지역주민과 행정기관이 처음으로 관계를 맺는 단계로 이 과정에서 신뢰가 형성된다면 기관에 대한 만족도가 높고 서비스에 대한 긍정적인 효과도 기대할 수 있을 것이다. 이와 함께 상담 대상자가 자신의 문제 해결에 대한 태도를 결정하는 중요한 시기로 현재 상태를 해결하기 위해 자발적인 노력이 가능하도록 할 수도 있다. 보다 내실 있는 초기상담을 위해서는 급여 신청을 위한 단순 상담이 아닌 주요 문제와 욕구를 정확히 파악하여 전체의 상황을 인지하고 적합한 서비스를 안내와 제공하여야 한다. 따라서 대상자가 현재의 상태를 적극적으로 바꿀 수 있도록 전문적인 능력을 갖춘 상담자의 체계적인 지원이 요구된다.

제2절 희망복지지원단과 통합사례관리

희망복지지원단은 복합적 욕구를 가진 대상자에게 통합사례관리를 제공하고, 지역 내의 자원 및 방문형서비스 사업 등을 관리함으로써 지역단위 통합서비스 제공의 역할을 수행하는 전담조직이다. 2012년 4월 희망복지지원단이 출범하였고, 전국 17개 시도 229개 시군구에 설치되었다. 또한 사례관리의 경우 2015년 시범사업을 거쳐 2016년

933개소(중심읍면동 포함), 2017년 2,100개소, 2018년 전국에서 시행 되도록 사업을 추진하고 있다⁵⁾.

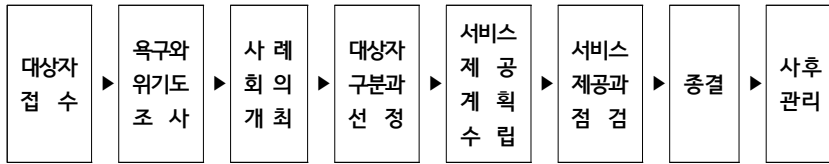
희망복지지원단은 기초수급자와 차상위 계층 등 빈곤층의 탈빈곤과 빈곤예방을 목표로 하고 지역주민의 다양한 복지욕구에도 대응하고 있다. 이는 민관협력을 통하여 지역단위의 통합 서비스제공 체계를 구축하고 운영하여 맞춤형 서비스의 제공과 지역주민의 복지체감도 향상을 도모하는 것이다.

주요 업무수행체계는 <표2-2>와 같이 ‘대상자 발굴’, ‘통합사례관리 실시’, ‘자원관리’ 등 이다. ‘대상자 발굴’은 읍면동 주민센터와 시군구 각 부서, 지역주민 및 관련 기관에서 선택된 대상자에 대해 읍면동 주민센터에서 초기상담을 수행하여 급여의 안내와 사례관리 대상자로 등록하는 것이다. ‘통합사례관리 실시’는 희망복지지원단을 중심으로 읍면동 주민센터, 지역사회보장협의체, 지역의 서비스제공기관과 연계 및 협력을 통해 대상자의 다양한 욕구를 충족시키는 맞춤형 서비스를 제공한다. 이를 위해서 대상자별로 서비스제공계획을 수립하여 통합 서비스를 제공하고, 점검 및 사후관리도 수행한다. ‘자원관리’는 지역 사회에서 활용 가능한 자원의 총괄 관리와 자원조사 및 발굴·배분을 통하여 원활한 지원과 공공 및 민간 복지통합관리시스템 운영을 통한 지역 자원의 효과적 관리 등을 의미한다.

<표 2-2> 희망복지지원단의 주요업무

업무	내용
대상자 발굴	통합사례관리의 대상자로 등록 및 발굴
통합사례관리 실시	복합 욕구를 충족하는 서비스 제공, 점검, 사후 관리
자원관리	지역 사회의 자원의 총괄 관리와 조사, 발굴 등

5) 보건복지부 보도자료(2016.5.31.)



[그림 2-3] 통합사례관리 절차

통합사례관리 사업은 희망복지지원단 업무의 핵심이라 할 수 있다. 사업을 상세하게 살펴보면, 지역사회 공공·민간자원에 대한 체계적인 관리·지원체계를 토대로 복합적이고 다양한 욕구를 가진 대상자에게 복지·보건·고용 등의 필요한 서비스를 통합적으로 연계 및 제공하고, 이를 지속적으로 상담·모니터링 하는 사업이다. 이는 주민의 다양한 욕구에 맞춤형 서비스를 연계 및 제공함으로써 주민의 안정적인 삶을 지원하여 복지제도의 효과성을 제고한다.

통합사례관리의 일반적인 절차는 [그림 2-3]과 같이 대상자 접수, 욕구 및 위기도 조사, 사례회의, 대상자 구분·선정, 서비스제공 계획 수립, 서비스제공 및 점검, 종결, 사후관리로 전체 8단계로 구성되어 있다. 절차의 시작인 대상자 접수에서 서비스제공계획 수립까지 30일 내에 처리를 원칙으로 하고 있다.

통합사례관리 절차에서 욕구와 위기도 조사의 내용은 [그림 2-4]의 ‘욕구조사관리’에서 확인할 수 있다. (1) 주요현상조회에서 대상자에게 적합한 욕구와 주요현상을 선택하여 입력 및 관리하고, (2) 주요원인 조회 및 관리로 원인과 원인제공자도 함께 등록한다. (3) 사회적지지 자원이 있는 경우 종합의견을 작성하여 입력하게 된다. 특히 욕구와 주요문제 영역은 10대 욕구⁶⁾를 기준으로 주요현상을 입력한다.

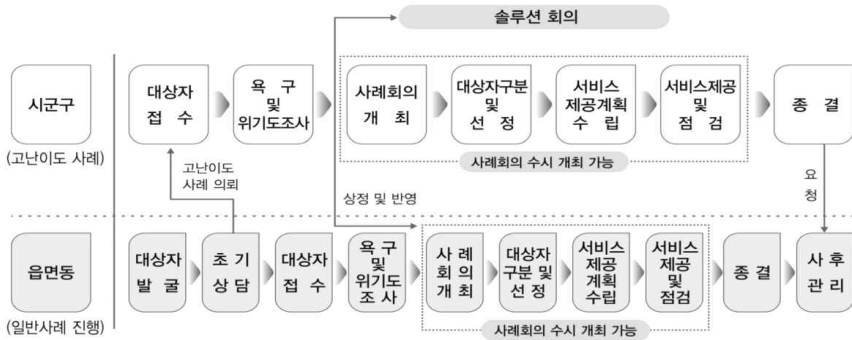
6) 10대 욕구 : 안전, 건강, 일상생활 유지, 가족 관계, 사회적 관계, 경제, 교육, 고용, 생활 환경, 법률 및 권익 보장

[그림 2-4] 사회보장정보시스템(행복e음) 읍구조사 관리화면⁷⁾

제3절 초기상담의 역할과 기능

‘복지허브화’로 기존 시군구의 통합사례관리가 읍면동을 중심으로 역할을 수행하게 되었다. 물론 통합사례관리 모델에 따라서 기능 및 역할의 차이가 있으나 [그림 2-5]처럼 ‘통합사례관리 기본형’을 중심으로 읍면동에서 사업을 추진하고 있다. 이와 함께 중심 읍면동에서 일반 읍면동의 통합사례관리를 수행하는 모델도 있다. 그리고 읍면동에서 통합사례관리가 불가능한 경우에는 기존의 방법과 같이 시군구 희망복지지원단에서 기능을 수행한다. 다만 사후관리는 읍면동의 역할이고, 사후관리 이행 사항 점검 등의 역할이 시군구에 마련되었다.

7) 보건복지부, 2016년 희망복지지원단 업무안내(2016)

[그림 2-5] 통합사례관리 기본형 흐름도⁸⁾

통합사례관리 기본형의 주요 특징으로 초기상담 수행 시 고난이도 사례로 판정될 경우, 시군구 희망복지지원단으로 의뢰해 시군구 희망복지지원단에서 통합사례관리를 수행하며, 읍면동의 자체 사례회의를 통해서도 문제가 해결되지 않는 경우도 역시 시군구 희망복지지원단에서 운영하는 ‘솔루션 회의’에 의뢰를 한다. 최종적으로 사후관리는 읍면동에서 모두 수행한다. 통합사례관리를 위해서 필수적인 자원에 대한 관리는 읍면동 단위 지역사회보장협의체 등을 통해 운영되고, 자체 발굴한 자원은 관할 읍면동 사례관리대상자에게 우선 연계하며, 추가 자원배분 등 조정이 필요한 경우는 시군구에서 지원한다.

초기상담의 단계에서 가구원별 주요문제 파악과 대상자가 생각하는 문제와 기대에 대하여 충분한 상담을 실시하여 핵심문제 등 근원적인 내용들을 파악한다. 주요문제 영역은 안전, 건강, 일상생활유지, 가족관계, 사회적 관계, 경제, 교육, 고용, 생활환경, 법률 및 권익보장 등 10개 영역의 복지욕구이다. 해당 욕구는 통합사례관리의 조사내용인 위기도 조사, 욕구조사 영역과 동일하게 구성되어 업무의 일관성을 유지하도록 체계가 구성되었다.

8) 보건복지부, 2016년 희망복지지원단 업무안내(2016)

[그림 2-6] 사회보장정보시스템(행복e음) 초기상담 기본화면⁹⁾

[그림 2-6]은 사회보장정보시스템의 초기상담 기본화면의 구성으로 (1) 대상자의 이름, 주민등록번호, 주소와 같은 기본정보를 조회하는 메뉴이고, (2) 대상자의 정보와 함께 현재까지 상담한 이력, 사례이력, 및 서비스 이력을 확인하여 초기상담에 참고하는 내용으로 구성되어 있고, (3) 초기상담 내용을 작성하는 곳으로 가구유형, 상담유형 등을 입력한다. 마지막으로 (4) 요청접수 할 경우에 활성화 되며 상담원의 판단에 따라 공적 서비스, 긴급복지, 사례관리 등을 선택하는 상담의 결과와 판정의견을 작성한다.

초기상담 절차는 ‘대상자발굴 → 방문상담계획수립 → 방문일정예약 → 방문상담실시 → 상담등록’으로 구성된다. 각 절차의 세부 내용으로 먼저 ‘대상자발굴’은 통합조사를 수행할 때 신규대상자의 가정방문의 동행, 수급신청 탈락자의 모니터링, 방문간호사의 만성질환자 명단, 인적안전망 등으로 연계 및 발굴하여 접수된다. ‘방문상담계획수립’은

9) 보건복지부, 2016년 희망복지지원단 업무안내(2016)

방문대상자의 수급이력을 사전에 파악하고, 전화 등을 통해 사전에 대상자와 방문일정을 예약한다. 이때, 공적급여의 경우 사회보장정보 시스템(행복e음)을 이용해 수급여부를 확인하고, 민간자원의 경우에는 서비스 제공이력 등을 해당 기관의 협조를 얻어 확인한다. ‘방문상담 실시’에서는 대상자의 생활실태 및 대상자의 욕구를 파악하고, 서비스 연계의 필요성을 확인하고 점검한다. 마지막으로 ‘상담등록’은 상담의 결과를 정리하게 되는데, 대상자의 정보는 기록으로 공유되어 향후 효과적으로 지원할 수 있는 자료로 활용되므로 약식 표현이나 축약을 하지 않고, 쉽게 이해할 수 있도록 구체적으로 작성한다. 초기상담의 서식에 표시하지 않고 직접 기술하는 내용은 대상자 지원에 참여하는 다른 업무관련자가 대상자에 대하여 이해할 수 있도록 상담 진행의 내용과 검토 사항 등을 상세하게 기록한다.

한편, 초기상담 이외에 ‘모니터상담’이라는 절차도 마련되어 있다. 모니터상담은 ‘대상자조회 → 모니터상담일정예약 → 모니터상담실시 → 상담등록’으로 구성되어 있다. 모니터상담은 기존대상자 모니터링으로 대상자의 건강·영양 상태, 주거환경 등 생활실태와 욕구파악과 제공된 급여·서비스가 제대로 전달되고 있는지를 파악한다. 그리고 신규 제도 및 서비스, 변화된 내용, 이용 가능한 자원 정보 제공 등이 주요 점검 사항이다. 이외의 절차는 초기상담과 유사하게 진행된다.

빅데이터 분석과 텍스트마이닝

제1절 빅데이터 분석

제2절 데이터마이닝과 텍스트마이닝

제3절 텍스트마이닝 활용사례

제3장 빅데이터 분석과 텍스트마이닝

제1절 빅데이터 분석

빅데이터(Big data)는 글자 그대로 ‘큰 자료’를 의미하기도 한다. 그러나 단순히 자료의 양만 방대한 것이 아니라 복잡성도 증가해서 기존의 자료 처리와 관리 도구로는 다루기 어려운 자료의 집합이라고 할 수 있다(McKinsey, 2011). 빅데이터의 출현은 산업계에서 다양한 고객의 정보가 축적되고 가치를 창출할 수 있는 기회가 생긴 것에서 시작된다. 그리고 학계에서 빅데이터를 다루는 연구가 증가하였는데, 대표적인 연구가 게놈 프로젝트(Genome Project)이다. 마지막으로 기술의 발전인데, 일상생활의 모든 것이 디지털화 되면서 축적되는 데이터의 양이 기하급수적으로 늘어났고, 이를 저장할 수 있는 도구 및 장치가 마련 된 것이다.

〈표 3-1〉 빅데이터를 분석하기 위한 시각의 변화

구분	내용
대상 측면	통계적 표본에서 전체 대상으로 확장
자료 측면	중요한 요인(변수) 이용에서 수집 가능한 모든 변수 활용
모델 측면	인과관계보다 현상의 설명 가능성 증명으로 전환

빅데이터와 함께 분석의 관점에서도 변화가 발생하게 되는데, 크게 3가지 측면이 있다. <표 3-1>처럼 표본에서 전수, 자료의 질보다 양, 인과관계보다 상관관계로의 변화가 특징이다. 빅데이터라는 용어에서 알 수 있듯이 자료의 양이 중요성을 갖게 되는데, 이는 결국 표본이 아닌 전체인 전수의 개념으로 확장과 현상을 설명하기 위한 자료로 요인의 질보다는 다수를 활용하는 양적인 증가로 나타났다. 통계학의 대표적인 모델인 회귀분석은 독립변수와 종속변수의 관계를 모형화한 것으로 원인과 결과의 설명이다. 그러나 빅데이터는 원인과 결과의 인과관계를 파악하기 보다는 현상을 설명할 수 있는 연관성에 관심을 두고 활용하게 된다.

빅데이터 활용 사례는 다양한 분야에서 계속해서 나타나고 있다. 구글(Google)은 사용자의 로그 자료를 활용하면서 기존의 페이지랭크(Page Rank) 알고리즘을 혁신했다. 월마트의 경우에는 고객의 구매 패턴을 분석해 상품 진열에 활용하는데, 예를 들어 허리케인이 올 때 손전등뿐 아니라 비상음식 구매가 증가하는 것을 파악하여 두 제품을 함께 진열해 매출을 증대시키기도 한다. 기업뿐만 아니라 정부에서도 대국민 서비스의 개선을 위하여 빅데이터를 활용한다. 복지사각지대 대상자 발굴을 위해 단전, 단수 등 다양한 정보를 이용¹⁰⁾하고, 실시간 교통정보 수집, 기후 정보, 각종 지질 활동, 소방 서비스를 모니터링 하여 상황분석 결과를 제공하기도 한다.

빅데이터 분석은 일반적으로 비즈니스 모델 개발에 활용되고 있고, 대표적인 기법으로 연관 규칙 학습(association rule learning), 유전 알고리즘(genetic algorithms), 회귀분석(regression), 소셜 네트워크 분석(social network analysis), 분류분석(classification), 기계학습(machine learning), 감성분석(sentiment analysis) 등이 있다. 최근

10) 김은하 외(2015), 사회보장정보시스템을 활용한 복지 사각지대 발굴방안 연구

컴퓨터의 성능 향상으로 인공지능망 기법을 확장한 딥러닝(deep learning)의 활용도 확대되고 있다.

기계학습은 예를 들어, ‘기존에 축적된 기록을 바탕으로 소비자가 현재 보유한 물품 중에서 어떤 것을 우선 구매 할 것인가?’와 같은 문제를 해결할 때 사용된다. 이를 해결하기 위해 기계학습은 자료를 이용해 스스로 학습하는 알고리즘을 포함하고 있다. 즉, 기계학습은 훈련 데이터로부터 학습한 특성을 활용해 ‘예측’과 ‘분류’하는 일에 초점을 둔다. 특히, 연관 규칙 학습은 어떤 변인들 간에 주목할 만한 상관관계가 있는지를 찾는 방법이다. 그리고 딥러닝은 여러 비선형 변환기법의 조합을 통해 높은 수준으로 추상화를 시도하는 기계학습 알고리즘의 집합으로 정의가 되며, 큰 틀에서는 사람의 사고방식을 컴퓨터에게 가르치는 인공지능의 한 분야이다.

이와 같은 방법들은 대부분 통계학과 전산학, 그리고 기계학습 및 데이터마이닝 분야에서 이용된 기법이며, 계속되는 비정형 데이터의 증가로 텍스트마이닝(text mining), 오피니언마이닝(opinion mining), 소셜 네트워크 분석(social network analysis)도 주목 받고 있다.

특히, 텍스트마이닝은 비·반정형 데이터에서 자연어 처리(natural language processing) 기술에 기반을 두어 유용한 정보를 추출·가공하는 것을 목적으로 하는 방법이다. 텍스트마이닝 기법으로 방대한 텍스트 데이터에서 의미 있는 정보를 추출하고, 텍스트 내 또는 다른 정보와의 연계성을 파악하며, 텍스트가 가진 범주를 찾아내는 등의 단순한 정보 검색 이상의 결과를 얻어낼 수 있다. 컴퓨터로 인간이 사용하는 언어를 탐색하고 그 안에 숨겨진 정보를 발굴하기 위하여 대용량 언어의 자원과 통계적 알고리즘이 활용된다.

오피니언마이닝은 텍스트마이닝의 관련 분야이다. 이 방법은 소셜 미디어 등의 정형·비정형 텍스트의 긍정, 부정, 중립의 선호도를 판별

하는 기술이다. 그리고 특정 서비스 및 상품에 대한 시장규모 예측, 소비자의 반응, 입소문(Viral Marketing) 분석 등에 활용되고 있다. 과학적인 오피니언마이닝을 위해서는 관련 분야의 전문가가 정의한 선호도를 나타내는 단어 자원의 축적이 필요하다.

웹 서비스가 다양한 분야로 확장되고 스마트 기기의 광범위한 보급으로 온라인 소셜 네트워크 서비스(SNS)가 사용자들에게 대중적으로 활용되고 있다. 온라인 소셜 네트워크 서비스는 온라인에서 사용자 간의 의사소통과 정보 공유, 인맥 확대 등을 위한 온라인 커뮤니티 서비스이다(Ellison et al., 2007). 온라인 소셜 네트워크 서비스에서 사용자간의 관계도는 노드(Node)와 링크(Link)로 연결하는 그래프로 표현되고, 온라인 소셜 네트워크 서비스의 사용자가 늘어남에 따라 이를 활용한 기업 마케팅과 사회현상 분석 등의 활용 가치 또한 확대되고 있다(Kwak et al., 2010, Sohn and Chung, 2011).

한편, 소셜 네트워크 분석은 사람들 사이의 관계를 여러 측면에서 해석하기 위해서 노드와 링크로 연결된 사회망 그래프를 생성하고, 이를 분석하는 방법들을 통칭한다. 이 방법으로 사람과 사람 사이의 관계, 친밀도, 그룹 분류, 연결 강도 등을 측정하는 것이 가능하다. 소셜 네트워크 분석(SNA)의 대표적인 방법으로 차수 중심성(Degree Centrality), 매개 중심성(Betweenness Centrality), 근접 중심성(Closeness Centrality)이 있다. 중심성 분석에는 노드와 노드 사이의 최단 경로가 고려되지 않으나, 매개 중심성과 근접 중심성 등을 포함하는 많은 네트워크 분석 방법에서 노드와 노드 사이의 최단 경로가 중요한 지표로 사용되고 있다(Otte and Rousseau, 2002).

제2절 데이터마이닝과 텍스트마이닝

일리노이 대학의 Jiawei Han 교수는 데이터마이닝(Datamining)을 ‘데이터로부터 지식(Knowledge)의 발견’이라 정의하고, 방대한 양의 데이터로부터 흥미로운 패턴을 도출하는 것이라 하였다. 이는 데이터 마이닝이 광산(mine)에서 광물을 캐내는 것(mining)에 비유한 것으로 수많은 데이터로부터 유용한 정보를 찾아내는 것이다. 즉, 대용량의 데이터로부터 이들 데이터 내에 존재하는 관계, 패턴, 규칙 등을 탐색하고 모델링하여 유용한 지식을 추출하는 일련의 과정이다.

데이터마이닝은 통계학과 기계학습으로 알려진 두 분야의 공통영역에서 발생했다고 할 수 있다. 즉, 데이터를 탐색하고 모델을 구축하는 다양한 기법은 통계학에 존재하였다. 그러나 방대한 데이터와 계산이 요구되는 데이터마이닝의 응용분야에서는 계산의 복잡성 등 통계학을 적용하기에 무리가 있었다. 더불어 전통적인 통계학은 추론에 관심을 두고 있지만, 데이터마이닝은 다양한 방식으로 대량의 데이터 집합을 다루기 때문에 일반화 또는 예측에 관심을 갖고 있다.

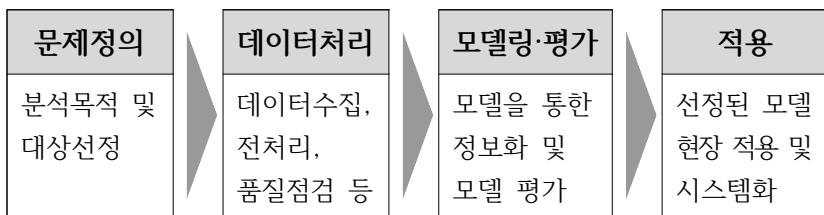
빅데이터 시대에 앞서 데이터마이닝이 알려졌으며, 데이터마이닝을 효과적으로 수행하기 위해서는 기존에 이용하던 데이터보다 더 많은 데이터가 필요로 하게 되어 결국 빅데이터 시대는 더 빠르게 다가왔다고 할 수도 있다. 즉, 데이터가 방대해지고 업무의 복잡도가 증가해 전문가의 통찰력에도 한계가 발생하여, 결국 데이터에 숨겨진 유용한 지식과 패턴을 파악하는 것을 사람이 직접 수행하기 어려워져 데이터 마이닝의 개념을 활용하게 되었고 빅데이터는 발전하였다.

일반적으로 통계적 분석은 가설을 바탕으로 검증할 것의 데이터를 수집하고 분석 하는 것으로 연구자의 현상에 대한 통찰이 중요하다. 그러나 데이터마이닝은 데이터에 기반을 두어 지식 패턴을 추출하기

때문에 연구자가 파악하지 못한 추가적인 현상들을 탐색 할 수 있게 된다. 탐색의 결과 중 논리적인 설명이 어려운 경우도 종종 발생하며 연구자들에 의해 간과되기도 한다. 그러나 이런 결과도 현실적으로 활용성이 높아 실생활에 적용되기도 한다.

데이터마이닝 절차는 [그림 3-1]과 같이 ‘문제정의’, ‘데이터처리’, ‘모델링·평가’, ‘적용’으로 나눈다. 데이터마이닝의 1단계 문제정의는 분석 목적 및 대상을 선정하고 전체계획을 수립하는 단계이다. 2단계 데이터처리는 데이터수집, 품질점검, 샘플링, 데이터변환, 데이터분할 등을 통해 분석에 맞도록 데이터를 준비하는 단계이고, 3단계 모델링·평가는 여러 가지 분석기법으로 모델링을 실행하고, 평가를 통해서 설명력이나 예측력이 높은 모델을 선정하는 것이다. 4단계인 적용은 선정된 모델을 관심 있는 새로운 데이터에 실제로 적용 및 시스템화 하는 것이다.

데이터마이닝의 특징으로 첫째, 대용량의 관측된 데이터를 다루고, 둘째, 이론보다는 실무와 경험 데이터에 기반을 둔 시스템을 활용한 업무 중심의 방법이고, 셋째, 개발되는 모델이 새로운 데이터에 적용 가능성의 평가이다. 마지막으로 다양한 상황에서 의사결정을 지원하기 위해서 활용된다.



[그림 3-1] 데이터마이닝 기본절차

데이터마이닝은 다양한 분야에 활용 가능한데, 기업에서는 고객의 소비 및 이용 패턴을 분석하여 마케팅 전략을 계획한다. 예를 들어, 아마존은 기존의 구매 내역을 분석하여 고객에게 필요한 책, 물품을 추천하는 메일을 보내고 있다. 그리고 신용평가, 부정·적정 행위 판단 등에도 활용하고, 병원에서는 환자 데이터를 이용해 발생 가능성이 높은 질병을 예측하기도 한다.

통계학의 기본적인 분석 데이터가 정형화된 수치 및 범주를 다루고 있고, 기본적인 데이터마이닝도 데이터적인 측면에서는 다르지 않다. 그러나 비정형데이터마이닝은 정형화된 데이터가 아닌 텍스트, 영상, 음성 등의 데이터를 다룬다. 텍스트마이닝과 소셜 네트워크 분석이 대표적인 분석 방법이다.

텍스트마이닝은 구조화되지 않은 문서에서 자동적으로 정보를 추출하고 패턴을 분석하는 것으로 결국, 텍스트마이닝의 목적은 알려져 있지 않은 정보나 패턴을 도출하는 것이다(Feldman and Dagan, 1995). 즉, 방대한 양의 데이터에서 유용한 정보를 이끌어내는 데이터마이닝과 유사하게 텍스트마이닝은 구조화되지 않은 문서에 동일한 개념을 적용한다. 수집된 텍스트 문서의 탐색이 실행된 후에 문서들 사이에 존재하는 복잡한 연관관계를 찾게 된다.

텍스트마이닝의 핵심은 구조화와 정보추출인데, 구조화란 비정형 텍스트를 정형화하는 과정을 포함하며 전처리 단계라고 할 수 있다. 모델에 이용되는 데이터는 전처리 과정이 필수적인데, 텍스트의 경우 입력 방법의 다양성으로 수정·변환 등의 전처리에 많은 노력이 요구된다. Feldman과 Dagan은 텍스트 형식의 데이터베이스에서 지식의 발견이라는 개념을 제시하여, 텍스트에서 의미 있는 정보의 집단을 분류하고, 이들을 계층구조로 정리하였으며, 문서들 간의 연관관계에 대해 해석하였다.

텍스트란 언어를 문자로 표현한 것으로 언어가 갖고 있는 특성이 문화, 관습에 따라 다르게 표현된다. 즉, 한국어를 모국어로 사용하는 분석전문가는 영어문화권의 텍스트를 분석하던 사람에 비해서 분석의 성과가 우위에 있기 쉽지 않다. 또한 미국과 영국이 영어를 모국어로 사용하지만, 문화와 관습의 차이가 텍스트에 반영되기 때문에 한쪽 국가에서 모델링한 결과의 정확도가 높았다고 이를 다른 쪽 국가에 동일하게 적용할 수 있는 것은 아니다. 결국 국가별로 상황과 현상에 맞는 분석 및 기법이 적용되는데, 이와 같은 언어, 문화 등의 차이로 텍스트마이닝의 어려움은 가중된다.

이외에도 텍스트마이닝 기법으로 ‘동시단어’ 분석은 문헌의 키워드 또는 용어 간의 관계를 파악하는 방법으로 연구동향 등을 분석하는데 용이하다(Choi et al., 2001; Yoon and Park, 2007). ‘문헌키워드’ 분석은 관련 학문 분야의 연구 현황들을 파악하는 유용한 방법으로 관련 연구자에 의해 많이 보고되어 왔다(Qin He, 1999). 그리고 연구 논문, 보고서 등의 주요 키워드를 활용하여 학문의 연구 동향을 분석하는 방법으로 ‘계량서지학’ 분석 방법이 활용되고 있다. 전통적으로 계량서지학 방법에는 논문 수, 공저자, 인용 분석, 동시출연 단어분석 등이 있다(Kim, 2007).

텍스트마이닝은 텍스트로부터 고품질의 정보를 도출하는 과정으로 입력된 텍스트를 구조화해 데이터에서 패턴을 도출한 후 결과를 평가 및 해석하는 일련의 과정이다. 즉, 빅데이터 시대에 맞게 여러 종류의 문서로부터 데이터를 수집하고 이를 문서별 단어 등을 추출해 연관성 등의 결과를 얻거나 의사결정을 지원하는 방법이다. 알려진 예로서 트위터는 자사 트위터 계정에서 고객이 주고받는 문장을 API로 읽어 분석하고 있다. 사용자별, 시간별로 도출 단어, 긍정·부정 언어 사용, 시간의 흐름에 따른 키워드 변화 등을 분석한다.

최근 인터넷에 있는 빅데이터의 대부분은 텍스트, 동영상 형태의 비정형, 비구조화된 데이터의 형태로 존재한다. 따라서 대용량 텍스트에서 의미 있는 패턴을 발견하기 위해서 전처리 과정이 필수적이며, 데이터베이스 구조화 및 수학적 알고리즘 적용이 요구된다. 분석의 목적과 데이터의 조건, 연구 환경에 따라 다양한 정보 추출 방법과 전처리 및 분석 알고리즘도 필요하다. 텍스트의 시간정보를 이용한 종단적 분석으로 시계열적 변화를 추적하여 트렌드(Trend)의 분석과 기술 예측 등 미래예측과 전망에 활용하고, 최근에는 단순히 텍스트만이 아니라 그림, 음성, 영상 등 멀티미디어의 비텍스트형 데이터도 구조화된 처리를 통해 활용한다.

빅데이터를 기반으로 한 텍스트마이닝의 정보 추출 방법인 키워드 추출(Keyword Extraction)은 텍스트를 이용한 여러 분석에 활용되는 기본적인 시작으로 문서의 핵심어를 추출하거나 문서 사이의 비슷함 정도를 구하여 핵심 키워드를 도출하는 과정이다. 여러 키워드 추출 알고리즘을 사용하며, 도출된 키워드는 연관성을 고려하여 군집화하여 네트워크 분석을 하거나 등장 빈도 등을 고려한 가중치를 산정하여 워드클라우드 등의 형태로 시각화해 직관적으로 표현한다.

다른 방법인 역문헌 빈도수(TF-IDF, Term Frequency-Inverse Document Frequency)는 특정 단어가 수집한 문서들에서 계속해서 나타나는 경우 중요한 의미를 가질 것이라는 전제로 단어들의 중요도(가중치)를 계산하는 방법이다. 분석 대상 문서의 핵심어를 추출하고 검색 엔진에서 검색어 순위, 문서 사이의 유사도 검증 등에 사용 할 수 있다. TF(Term Frequency)는 특정 단어가 분석 대상 문서 집합 내의 등장한 빈도수이며, DF(Document Frequency)는 특정 단어가 나타난 문서의 수를 의미한다. IDF(Inverse Document Frequency)는 DF의 역수로 TF와 IDF의 곱이 특정 단어의 전체적인 중요도를 계산

하는 가중치로 정의된다. 그리고 다른 기법인 토픽 모델링(topic modeling)이 있는데, 이는 단어들이 동시에 출현 하는 빈도와 거리 등을 이용해 비슷한 의미를 가진 단어들을 묶어주는 클러스터링 방식으로 텍스트의 주제를 파악하는 방법이다. 이런 방법들은 빅데이터를 변환 및 처리하는 기술들의 발달에 따라 대용량의 데이터에 내포된 주제와 핵심어 등을 분석할 수 있게 기회를 제공한 것이다.

제3절 텍스트마이닝 활용사례

복지급여의 접점에는 국민이 있으며, 국민의 소리(VOC)는 정책변화 및 서비스 만족도 향상 등의 의사결정에 중요한 데이터로 활용된다. 그래서 최근 공공기관뿐만 아니라 의료, 금융, 교육기관 등의 대부분 조직에서 고객의 소리를 수집하여 이용한다. 비정형인 고객의 소리는 SNS, 웹 Q&A, 방문, 우편 등 소셜미디어의 발전으로 다양한 수단에 의해 적재된다. 그러나 텍스트인 비정형 자료에 대한 처리와 분석의 어려움으로 여전히 연구 내용의 수준이 낮다고 할 수 있다. 그러나 텍스트 형식의 데이터를 활용한 분석은 계속되고 있으며 이를 활용한 사례의 영역도 확대되고 있다.

국민건강보험공단의 ‘국민건강주의 예보 시범서비스 구축’은 기존에 전염병이 유행하면 사후 대처로 사회경제적 비용이 발생함을 최소로 하고자 마련되었다. 따라서 빅데이터 분석을 활용하여 예방과 예측이 가능한 의료서비스 제공을 시도하여, 기상청의 일기예보와 같이 각종 전염 질환의 위험도를 예보하는 체계를 구축하는 것이다. 주요 개발 내용은 유행성 질병에 대한 위험도 예보와 지역·연령별 위험도 예보 등 유행할 가능성을 예보한다. 소셜미디어를 통해 수집되는 사람의

추이를 파악하고, 건강보험정보와 소셜미디어 정보를 융합하여 질병을 예측하는 모델을 개발한다.

국민권익위원회의 ‘민원정보분석 시스템을 통한 국민과 정보의 소통 활성화’는 통합서비스를 구축하여 기관 간 유기적으로 연계된 표준화된 상담 서비스의 개발과 공급으로 지속적으로 유지할 수 있는 지원 체계 구축이 필요해 시작되었다. 그리고 전화 상담, 민원 제기 등의 절차가 일반인에게 까다로워 편의성을 제고하는 방안으로 개선하였다. 이를 해결하기 위해서 홈페이지 민원, 제안, 콜센터 상담 등을 통해 축적된 민원 데이터를 종합적·체계적으로 분석하여 지원정보의 획득과 민원처리 시간 단축 등 업무를 효율화하였다. 그리고 국민의 요구와 만족을 근거로 하는 서비스 체계의 수립과 상호작용을 통한 지속적 서비스 개선 기반을 설계하였다.

한국도로공사의 ‘고객 목소리 분석 시스템을 통한 서비스 혁신’은 기존에는 고객의 목소리를 직접 들을 수 있는 수단이 한정적이었으나 소셜미디어의 발달로 고객의 의견을 접할 수 있는 여러 통로가 개설되어 실시간 고객 응대 등 효율적인 관리가 가능하다. 고도화된 언어 처리 기법으로 콜 상담서비스, 민원관리 시스템, 채팅상담 시스템을 분석하여 고객만족도에 도움이 될 수 있는 지표와 이슈를 도출한다. 분석의 주제를 중심으로 핵심어, 토픽을 추출하고 유형·시간별 분석과 연관도와 추이 분석을 한다. 따라서 주요 이슈나 불만을 사전에 파악하여 대응할 수 있고, 서비스 전략 수립이나 정책 수립을 위한 의사결정 등의 지원이 가능하다.

이외에도 SK텔레콤의 ‘소셜 네트워크에서의 여론분석을 위한 스마트 인사이트 시스템’은 소셜 네트워크의 여론을 분석하여 기업에 필요한 콘텐츠를 제공한다. 즉, 관심 있는 핵심어를 중심으로 온라인 동향을 분석하여 소비자의 반응을 빠르고 정확하게 파악하는 기능을 한다.

국외에서 ‘SNS를 활용한 할리우드 흥행 수익 예측’은 SNS에서의 사용자들의 관심사를 분석함으로써 비즈니스 시장을 예측하는 것으로 SNS 중 트위터(Twitter)의 내용을 활용한 것이다. 소셜 네트워크에서 표현된 개인들의 의견을 종합적으로 분석하여 대중의 주요 관심사와 흐름을 실시간으로 파악하는 것이다. 코카콜라의 ‘SNS 데이터 활용을 통한 가치 향상’은 트위터와 페이스북(Facebook)에서 자사와 관련된 데이터를 수집하여 분석한 후 제품의 판매와 관련된 의사결정 정보를 반영한다. 즉, 제품에 부정적인 정보가 증가하는 국가나 지역을 대상으로 홍보를 강화하는 등 실시간으로 대응하고, 프로세스 개선, 고객 만족도, 이슈 도출 등을 실시하여 최적의 서비스 제공한다.

앞에서 기술한 텍스트마이닝 사례뿐만 아니라, 빅데이터의 활용은 다양한데, 주요 국가의 정부는 빅데이터 시장의 활성화를 위한 공공 정보 공개, 빅데이터 기반 공공서비스 제공 등의 다양한 사업을 추진하고 있다. 예를 들어 미국은 2억 달러 규모의 빅데이터 연구개발에 이니셔티브를 주도하기 위해 기술력 확보, 사회 각 영역의 활용, 인력 양성의 3가지 측면에 집중적으로 투자하고 있다. 영국은 2012년 6월 발표된 오픈 데이터 전략에 따라 각 부처는 의료, 고용, 교육, 세금, 지리 데이터 등에 대해 2015년까지 순차적으로 개방한다.

기업 중심의 빅데이터 시장의 경우, Ford는 차량에 설치된 센서를 이용해 운전자의 주행 습관뿐만 아니라 주행 환경에 대한 데이터를 수집하고, 수집된 데이터를 분석해 고객의 잠재된 요구사항을 파악해 신제품에 반영한다. Zara는 빅데이터 분석을 통하여 현재 유행하는 패션 경향을 즉각 반영한 다품종 소량 생산전략을 통해 급성장했다. 특히 상품수요의 예측, 매장별 적정 재고 산출, 상품별 가격 그리고 운송계획까지 모두 실시간으로 수집되는 빅데이터 분석을 통해 의사결정을 내린다.

이와 함께 빅데이터 기반의 인공지능(Artificial Intelligence) 자기 학습 컴퓨터도 지속적으로 구현되고 있다. 대표적으로 2005년 개발을 시작한 IBM의 인공지능 플랫폼인 왓슨(Watson)이 언급된다. 왓슨은 인간의 언어를 이해하고 해석하며 빅데이터를 이용해 고도의 지능적 문제를 분석하여 해답을 찾아내는 수준에 도달하였다. 결국 사람과 비슷한 형식으로 보다 빠르게 빅데이터를 분석하고 처리할 수 있으며 자연어를 이해하여 데이터 기반의 가설을 제안하고 학습을 수행한다. 왓슨(Watson)은 2011년 미국 ABC 텔레비전 퀴즈쇼에서 우승하면서 크게 주목받았으며, 최근에는 클라우드 기반의 제품과 서비스를 제공하는데 의료, 은행, 보험, 질병의 진단 및 치료, 유통, 교육까지 다양하게 활용된다. 특히 왓슨을 이용해 암 진단 서비스를 제공하는 미국 앤더슨 암센터에 따르면 왓슨의 평균적인 진단율은 약 96%로 전문의 보다 정확도가 높다고 한다.

아울러 최근에 국내외의 인공지능에 대한 인식을 새롭게 한 구글 딥마인드가 개발한 인공지능 바둑 프로그램 알파고(alphago)가 있다. 알파고의 알고리즘은 머신 러닝과 트리 순환 기술을 조합해 인간과 컴퓨터 모두와 학습하고 결과를 반영해 문제의 해결능력을 향상한다. 핵심 알고리즘인 딥러닝은 인신경망 기술로 구현한 ‘가치 네트워크’와 ‘정책 네트워크’에 기인한 몬테칼로 트리 순회를 사용한다. 알파고는 처음에는 3천 만 정도의 데이터베이스를 이용해 기존의 바둑을 흉내 내어 학습하였고, 일정한 수준에 도달하자 강화된 학습 방법으로 또 다른 자신과 훈련을 통해 계속 진화하도록 개발하였다.

복지영역별 토픽 분석

제1절 토픽 분석의 방법과 절차

제2절 복지영역별 텍스트마이닝 분석결과

제3절 초기상담 기능강화 방안

제4장 복지영역별 토픽 분석

제1절 토픽 분석의 방법과 절차

복지영역별 핵심어들의 분석은 ‘임신·출산’, ‘아동청소년’, ‘건강’, ‘서민금융’, ‘문화’, ‘고용’ 등 [그림 4-1]과 같이 복지로 ‘한눈에 보는 복지정보’의 16개 영역별로 분류해 단어의 다빈도, 중요도 및 토픽을 분석하여 복지영역별로 상담에서 관심 사항을 판단할 수 있는 정보를 구축하고자 한다. 정의된 16개 복지영역은 생애주기별, 인구집단별, 생활영역별로 범주화할 수 있다.

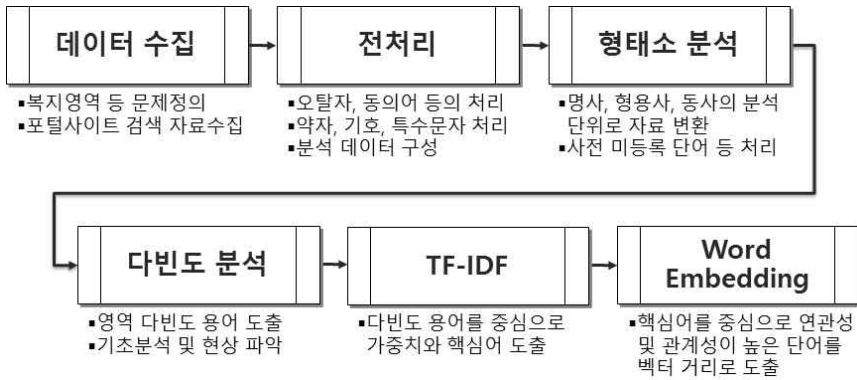


[그림 4-1] 복지로 ‘한눈에 보는 복지정보’ 복지영역별 분류

텍스트마이닝을 위한 데이터 수집은 포털사이트¹¹⁾ 상세검색 결과를 웹 크롤링(Web Crawling)을 통해 수행하였다. 복지영역별로 수집된 텍스트를 단어의 다빈도와 중요도 등의 분석을 위해 전처리 과정으로 정형화된 텍스트로 변환한다. 다음으로 정형화된 단어를 이용 가능한 단어로 결정하기 위해 오피니언마이닝으로 각 영역별로 관련이 적은 단어와 텍스트의 특성상 구조화되지 않는 등의 분석이 어려운 단어를 제외하였다. 그리고 각 영역별로 연관성 분석을 통하여 초기상담에 활용할 수 있는 관심단어 및 영역 등을 제시한다.

일반적으로 텍스트마이닝은 데이터의 수집과 이를 분석용 자료로 변환하는 전처리 과정에 상당한 시간이 소요된다. 그리고 정형화된 자료의 범주가 성별(남, 여), 연령대(10대, 20대, 30대 등)처럼 2수준 또는 10수준 내외의 정보를 이용하지만, 텍스트는 일반적으로 언어의 다양성만큼 인지하고 있는 수준의 복잡성이 높다. 따라서 일반화된 결론을 도출하기 위해서는 수집과 분석에 기존보다 많은 양의 자료가 필요하다. 그런데, 일부 복지영역의 경우 토픽 도출의 분석 및 결론에 충분하지 못한 자료가 수집되기도 하였다. 이런 한계로 복지사업의 연계 상담에 필요한 토픽 분석에는 앞서 언급한 것처럼 관심 영역의 단어를 제시만 하고자 한다. 일반적으로 모델링에서 훈련용 데이터(training), 검정용 데이터(validation)로 구분하여 연구를 수행하나, 본 연구에서는 훈련 데이터를 이용한 모델링 과정만 수행하여 결과를 제시하고자 한다. 따라서 본 연구를 수행함에 있어 결과에서 재현성, 과적합(overfitting) 등의 문제가 발생할 수 있으나, 방법론의 적용과 수행의 과정은 적절하다고 판단된다.

11) 국내의 포털사이트 중 하나인 네이버지식IN으로 일반 국민이 생각하는 복지영역의 관심 사항 등의 데이터를 수집할 수 있으며 이를 상담, 의사결정 등에 활용할 수 있을 것임



[그림 4-2] 복지영역별 토픽 분석 수행 절차

본 연구의 분석 절차는 [그림 4-2]와 같이 ‘자료수집’ → ‘전처리’ → ‘형태소분석’ → ‘다빈도분석’ → ‘TF-IDF’ → ‘Embedding’으로 앞에서 설명했던 [그림 3-1]의 데이터마ining 절차 중 ‘데이터처리’와 ‘모델링·평가’의 단계가 세분화된 것이다. 즉, 자료수집·전처리·형태소 분석은 ‘데이터처리’의 단계이고, 다빈도분석·TF-IDF·Embedding은 ‘모델링·평가’의 단계라 할 수 있다.

각 단계별로 내용을 살펴보면, ‘자료수집’ 단계는 문제정의와 함께 해결을 위한 데이터를 수집하는 과정으로 연구에 적합한 자료의 수집이다. 이 단계에서 웹 크롤링을 통해 약 10년(2006.1.~2016.6.)간의 포털사이트의 상세검색 데이터인 텍스트를 수집 처리하였다. 다음으로 ‘전처리’ 단계에서는 수집된 텍스트를 분석 처리할 수 있는 텍스트로 변환하는 과정으로 텍스트의 특성상 오타자의 점검(맞춤법), 유의어, 동의어 등 정리의 기초적인 수정을 거친다. 다음으로 텍스트마ining 알고리즘을 적용하기 위한 텍스트 분해의 단계로 텍스트를 형태소의 단위로 분해한다. 특히, 명사, 형용사, 동사가 연구를 위한 기본적인 형태소로 정의되어 문서내의 텍스트를 나눈다. 이 단계에서 분해된

단어를 오피니언마이닝을 통하여 전처리 과정에서 수행한 기초적인 점검을 다시 수행하여 유의어, 동의어 등을 점검한다.

위의 데이터처리 단계를 거쳐서 모델링·평가의 단계가 진행되는데, 이 과정에서 ‘다빈도분석’이 먼저 수행된다. 이 과정은 텍스트마이닝의 기초분석으로 각 단어의 빈도분석이라고 할 수 있다. 즉, 빈도를 통해 정보의 경향성 등을 판단하여 연구의 내용을 파악하는 과정이 된다. 더불어 정형화된 정보의 도출이 시작되는 단계이다. 다음은 ‘TF-IDF’ 값을 계산하여 정보의 가치와 연관성 등을 파악하게 된다. TF-IDF는 텍스트마이닝에서 이용하는 가중치로 정보 검색을 위해 여러 문서로 나누어진 문서 집합에서 어떤 단어가 특정 문서 내에서 어느 정도로 중요한지를 나타내는 통계적 수치이다. 일반적으로 문서의 핵심어를 추출하거나, 검색 엔진에서 검색 결과의 순위를 결정할 때 이용된다. 즉, 유사한 문서들 사이에서 특정 분야에 집중하여 발생하는 단어와 문서를 추출하는 것이다. 마지막으로 ‘word embedding’은 사람들의 언어를 컴퓨터가 인식할 수 있도록 변환한 것으로 알려진 알고리즘이 Word2Vec이다. 이 방법은 텍스트를 처리하는 인공지능망이 두 개의 층으로 구성되어 있으며, 말뭉치(corpus)를 초기값으로 해 말뭉치의 단어를 벡터로 표현하는 방법을 찾는데, 이 벡터의 값은 말뭉치에서 단어가 가지는 의미나 역할을 표현해주는 값이다. 이렇게 단어들의 의미와 맥락을 고려해 단어를 벡터로 표현한 것을 word embedding 이라고 한다.

여기서 먼저 단어의 중요도를 표현하는 TF-IDF를 상세히 소개하면 다음과 같다. 즉, TF(Term Frequency)는 단어 빈도로 특정 단어가 문서 내에 얼마나 빈번하게 나타나는지를 측정하는 값이며, 그 값이 클수록 문서에서 중요성이 높다고 정의한다. 그러나 여러 문서에서 계속해서 사용되는 경우, 그 단어는 일반적으로 이용되는 단어로서

정의해 중요도를 낮추게 된다. 이것을 DF(Document Frequency)인 문서의 빈도로 정의하고 이 값의 역수를 IDF(Inverse Document Frequency)로 정의한다. 따라서 TF-IDF는 TF와 IDF를 곱한 단어 빈도와 역문서 빈도의 곱이다.

<표 4-1> TF(단어 빈도)의 수학적 측정 방법

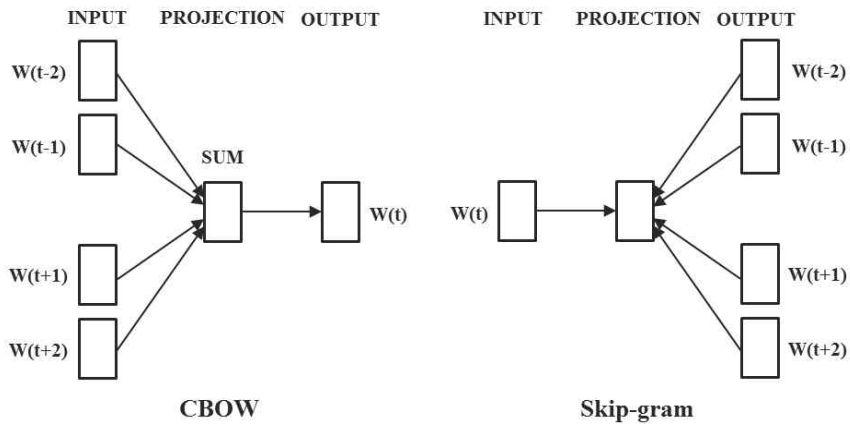
구분	TF 정의
단어출현 유무	0, 1
단순 빈도	$f_{t,d}$
로그 표준화	$1 + \log(f_{t,d})$
이중 표준화 0.5	$0.5 + 0.5 \times \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}}$
이중 표준화 K	$K + (1 - K) \times \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}}$

f : 빈도, t : 관심 단어, d : 문서, t' : 전체 대상 단어, $K \in (0, 1)$: 임의의 상수

<표 4-2> IDF(역문서 빈도)의 수학적 측정 방법

구분	IDF 정의
문서에 단어 포함	1
역문서 빈도	$\log \frac{N}{n_t}$
역문서 빈도 평탄화	$\log \left(1 + \frac{N}{n_t} \right)$
역문서 빈도 최대화	$\log \left(1 + \frac{\max_{\{t' \in d\}} n_{t'}}{n_t} \right)$
확률적 역문서 빈도	$\log \frac{N - n_t}{n_t}$

t : 관심 단어, d : 문서, t' : 전체 대상 단어, N : 문서 수, n_t : t 를 포함한 문서 수



[그림 4-3] Word2Vec 알고리즘의 아키텍처

TF와 IDF의 개념적 정의를 수학적으로 측정하는 식은 <표 4-1>과 <표 4-2>처럼 다양하게 소개되고 있다. 일반적으로 알려진 TF와 IDF 정의 중에서 연구에 적용은 분석의 상황에 적합하게 선택되어 핵심 단어를 탐색하게 된다. 선별된 단어 중심으로 word embedding으로 단어와의 관계를 분석하여 각 복지영역별 상담과 관련된 영역, 관심 단어 등을 도출하게 된다. 본 연구에서는 사용한 TF-IDF의 측정방법으로 '이중 표준화 0.5'와 '역문서 빈도 평탄화'의 곱을 이용하였으며 추가적으로 단어의 빈도가 일정 수준 이상 된 것에 의미를 부여하기 위해, 복지영역별 단어의 빈도가 상위 500 단어의 TF-IDF의 값으로 정렬하는 수정된 TF-IDF로 중요도를 측정하였다.

다음으로 단어들 간의 연관성을 측정하는 Word2Vec 알고리즘의 기본 구조는 Hidden Layer 없이 뉴럴 네트워크(Neural Network)를 이용하여 연산의 양을 최적화하였다. 내부적으로 [그림 4-3]과 같이 CBOW(Contiguous Bag Of Words)와 SGNS(Skip Gram model with Negative Sampling)인 두 개의 신경망 모델을 이용하여 문장을

학습해 유의어를 가까운 벡터공간에 표현한다(Tomas et al. 2013). 이는 단순히 한 단어의 앞뒤에 위치한 단어들이 유사한 의미를 갖는 단어로만 학습하는 것이 아닌 전후의 일정 범위에 나타나는 단어로 확장하여 유사도를 측정하는 개념이다.

CBOW 모델은 input이 t-2, t-1, t+1, t+2의 주변 단어들이고, output으로 t번째 단어를 예측하는 모델이다. SGNS보다 상대적으로 속도가 빠르다고 알려져 있다. 반대로 SGNS 모델은 input이 1번째 단어로 1개이고, output이 t-2, t-1, t+1, t+2의 주변 단어를 예측하는 모델이다. 상대적으로 빈도가 적은 단어를 잘 학습해서 성능이 CBOW 보다 좋다고 알려져 있다.

단어들 사이의 유사도의 측정은 코사인 유사도(Cosine Similarity)식을 이용한다. 코사인 유사도는 내적공간의 두 벡터 사이 각도를 cosine 값으로 변환하여 두 단어 간의 유사도를 표현한다. 이 값은 벡터의 크기와 관계없이 방향만을 이용하여 유사도를 계산하게 되며, 측정값은 0~1 사이의 양의 실수로 정의한다. 코사인 유사도는 아래의 식과 같이 계산되며, 여기서 A 와 B 는 벡터 A 와 B 의 성분이다.

$$\text{Cosine Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2 \times \sum_{i=1}^n (B_i)^2}}$$

제2절 복지영역별 텍스트마이닝 분석결과

1. 텍스트마이닝 분석결과

1) 분석데이터의 특징 및 분석 방법

초기상담을 위한 텍스트마이닝을 수행하기 위해서 일반인이 복지와 관련한 어떠한 내용들을 질의하는지에 대한 자료를 수집하였다. 현재 복지포(http://www.bokjiro.go.kr)에서 구분하여 제공하는 16가지의 서비스 유형을 초기 검색어로 설정하고, 포털사이트에서 검색결과를 수집하였다. 검색기간은 2006년 1월 1일부터 2016년 6월 9일까지로, 포털사이트 검색은 일반검색이 아닌 상세검색을 통해서 16개의 복지 영역별로 ‘복지’ 검색어를 반드시 포함하도록 검색하도록 설정하였다. 관련 검색어와 검색결과 건수는 <표 4-3>과 같다.

<표 4-3> 복지영역별 포털사이트 검색 현황

연번	영역	검색어	검색결과 수 ¹²⁾ (단위:건)
01	임신출산	임신 출산 + 복지 ¹³⁾	15,401
02	영유아	영유아 + 복지	14,068
03	아동청소년	아동 청소년 + 복지	121,626
04	청년	청년 + 복지	9,846
05	중장년	중년 장년 + 복지	2,163
06	노년	노인 노년 기초연금 + 복지 ¹⁴⁾	66,338
07	장애인	장애인 + 복지	69,816
08	한부모	한부모 + 복지	48,316
09	다문화(새터민)	다문화 새터민 + 복지	6,705
10	저소득층	저소득층 + 복지	28,816
11	교육	교육 + 복지	176,342

12) 각 영역에서의 중복된 검색결과를 삭제한 수치이다.

연번	영역	검색어	검색결과 수 (단위: 건)
12	고용	고용 + 복지	35,217
13	주거	주거 + 복지	13,310
14	건강	건강 + 복지	80,698
15	서민금융	생계자금 주거자금 고용지원 학자금 + 복지 ¹⁵⁾	20,897
16	문화	문화 + 복지	58,565

검색결과(768,124건)에 대한 텍스트마이닝은 형태소, 다빈도 분석과 단어별 유사도 분석을 실시하였다.¹⁶⁾ 형태소란 일정한 음성에 일정한 뜻이 결합되어 있는 말의 가장 적은 단위(박용욱, 2012)이다. 본 연구에서는 포털사이트의 검색결과에 대하여 위에서 정의한 각 복지영역별로 주요 키워드를 도출하고, 단어들 간의 관계성을 살펴봄으로써 각 복지영역별 주요 토픽(topic)을 찾아내고자 한다. 형태소 분석기는 오픈소스인 코엔엘파이(KoNLPy)¹⁷⁾를 사용하였고, 관련 패키지는 ‘Twitter Class’를 적용하였으며, 단어들 간의 유사도 분석을 위해서 ‘Word2Vec’¹⁸⁾을 이용하였다. 코엔엘파이는 문서들 내의 형태소들을 품사로 태깅(tagging)하는 역할을 수행하며, 본 연구에서는 한국어의 품사 중 명사, 동사, 형용사만을 대상으로 하여 추출하였다¹⁹⁾.

13) 포털사이트 검색 규칙에 따라 해당 검색어는 ‘임신’, ‘출산’을 하나이상 포함한 검색결과 중 ‘복지’를 포함한 상세검색결과

14) 복지로에 구분된 ‘노년’이라는 키워드는 포털사이트에서 많이 사용하고 있는 키워드가 아니므로 이와 유사한 단어인 ‘노인’이라는 키워드를 함께 넣었음

15) ‘서민금융’과 매칭되는 포털사이트 검색결과가 매우 적으므로, 복지로의 ‘서민금융’ 하위 카테고리인 ‘생계자금’, ‘주거자금’, ‘고용지원’, ‘학자금’의 키워드를 매칭

16) 분석을 위한 language는 python 3.5를 활용하였음

17) <http://www.konlpy.org>

18) Word2Vec는 구글의 연구원인 Tomas Mikolov와 Kai Chen, Greg Corrado, Jeffrey Dean에 의해 쓰인 논문인 "Efficient Estimation of Word Representations in Vector Space(벡터공간상에서 단어 의미의 효율적인 추정)"에서 제안된 방법을 구현한 알고리즘

19) 그 이유는 한국어에서의 의미전달에 있어서 가장 기본이 되는 품사이기 때문임

Word2Vec는 인공 신경망에서 시작되었는데 동일한 맥락을 갖는 단어는 가까운 거리에 있으며 비슷한 의미를 갖는다는 전제에서 개발되었다. 즉, 텍스트들 내의 단어들을 이용하여 학습을 수행하며, 기준 단어를 중심으로 근처(전후 5-10단어 수준)에 위치한 다른 단어와의 관계를 인공 신경망을 통해 학습한다. 결국 비슷한 성격의 단어들은 텍스트 내에서 가까운 곳에 위치할 가능성이 높기 때문에 반복되는 학습의 수행으로 두 단어는 점차 가까운 벡터로 인식되게 된다. 이를 이용해 형태소들 간의 유사성을 구하게 되며, 유사도는 앞에서 설명하였듯이 코사인 유사도를 이용한다. 코사인 유사도는 두 벡터 간의 상대적인 각도를 이용하여 계산하는데, 즉 두 벡터의 사잇각이 0도인 경우엔 유사도가 최댓값인 1로 사잇각이 90도가 되는 경우 유사도가 최솟값인 0으로 가장 낮은 유사도로 정의된다.

2) 키워드별 형태소의 빈도 및 유사도 분석결과

가. ‘임신출산’ 키워드 분석결과

다음 <표 4-4>는 임신출산과 관련하여 단어의 가중치(TF-IDF)를 기준으로 다빈도 순으로 정렬한 것이다. 결과는 ‘시댁’, ‘와이프’ 등과 같이 가족과 관련된 형태소들이 높은 순위로 나타났다. 또한, 양육 및 육아에 따른 휴직 등과 같은 복지정책의 이슈와 관련된 형태소들도 나타나고 있다. 한편, 이혼(위자료), 미혼모 등과 같은 사회적 이슈와 관련된 형태소들도 나타났다.

<표 4-4> ‘임신출산’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	시댁/Noun	1,193	0.6241
2	와이프/Noun	1,073	0.6232
3	남자친구/Noun	1,034	0.6230
4	생리/Noun	1,018	0.6228
5	둘째/Noun	828	0.6214
6	낳다/Verb	2,328	0.4906
7	신랑/Noun	2,068	0.4891
8	양육/Noun	1,340	0.4849
9	양육비/Noun	1,254	0.4844
10	출산휴가/Noun	1,063	0.4833
11	육아휴직/Noun	1,056	0.4832
12	만나다/Verb	1,056	0.4832
13	친정/Noun	1,048	0.4832
14	싸우다/Verb	1,027	0.4831
15	혼인신고/Noun	1,000	0.4829
16	위자료/Noun	992	0.4829
17	휴가/Noun	914	0.4824
18	미혼모/Noun	843	0.4820
19	육/Noun	783	0.4817
20	데리다/Verb	764	0.4816

한편, 앞에서 데이터의 특성을 설명하는 부분에서도 언급하였지만, 위의 검색결과는 다빈도 현황이다. 각각의 형태소가 구체적으로 어떤 이슈를 포괄하고 있는지에 대하여 각 형태소별 유사도 분석을 실시한 결과는 다음의 <표 4-5>와 같다. 이를 통하여 앞에서 살펴본 다빈도 형태소 분석에서 살펴본 이슈를 보다 구체적으로 살펴볼 수 있는데 예를 들어, ‘출산휴가’, ‘육아휴직’과 관련하여 휴가나 휴직이외에도 퇴직이나 권고사직 등의 이슈 등이 연관되어서 나타난다. 즉, 제도적으로 출산 후, 출산휴가나 육아휴직 등을 이용할 수 있으나 퇴직이나 권고사직 등의 문제도 함께 공존하고 있는 것이다. 미혼모의 경우는 보호시설이나 미혼모센터 등과 같은 형태소들과 한부모가정, 기초생활수급 등과 같은 복지서비스 지원과 관련되는 형태소들이 나타났다.

그 외에도 미숙아, 시험관 등 임신출산 과정에서의 제도적 지원, 정책 등의 형태소들도 나타난다.

한편, 낙태, 지우다, 원치않다, 피임, 콘돔 등 사회적 이슈와 관련된 형태소들도 나타났고, 이혼, 위자료와 관련하여 이혼, 소송, 양육비 등과 같은 형태소들도 나타났다.

<표 4-5> '임신출산' 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	시댁 /Noun	시부모(0.8354) ²⁰⁾ , 시어머니(0.8006), 친정(0.8000), 식구(0.7006), 시아버지(0.6958), 시어머니(0.6769), 분가(0.6721), 어머니(0.6687), 친정집(0.6476), 시누이(0.6356), 시아버님(0.6139), 아버님(0.6136), 제사(0.6103), 친정엄마(0.6062), 신랑(0.6039), 신혼집(0.6002), 남편(0.6000), 어른(0.5977), 처가(0.5924), 맥(0.5821)
2	와이프 /Noun	집사람(0.6844), 아내(0.6605), 장모님(0.6338), 누나(0.5427), 처가(0.5311), 장인(0.5235), 신혼(0.5215), 다툼(0.5135), 신랑(0.5091), 동생(0.5025), 어머니(0.4948), 남편(0.4924), 어머님(0.4869), 시어머니(0.4864), 처남(0.4737), 화해(0.4639), 형부(0.4630), 장인이름(0.4618), 그동안(0.4615), 분가(0.4475)
3	남자친구 /Noun	여자친구(0.7720), 헤어지다(0.6491), 미성년자(0.6255), 사귀다(0.5865), 알리다(0.5726), 애(0.5699), 지우다(0.5674), 여친(0.5653), 부모님(0.5623), 미혼(0.5412), 양쪽(0.5305), 돌싱(0.5221), 거들다(0.5216), 미혼모(0.5211), 자취(0.5177), 원치않다(0.5106), 모이다(0.5101), 지금(0.5035), 친구(0.5000), 사정(0.4929)
4	생리 /Noun	배란(0.7430), 질내사정(0.7276), 피임(0.7270), 불규칙(0.7103), 쿠퍼액(0.6813), 피임약(0.6694), 테스트(0.6666), 콘돔(0.6656), 질외(0.6646), 착상(0.6618), 혈(0.6554), 테스트기(0.6470), 갈색(0.6379), 임신증상(0.6243), 성관계(0.6152), 생리통(0.6143), 멧다(0.6125), 불안하다(0.6084), 설마(0.5937)
5	둘째 /Noun	첫째(0.8947), 첫째(0.7734), 쌍둥이(0.7536), 딸아이(0.6862), 아이(0.6481), 애가(0.6072), 첫(0.5878), 아가(0.5860), 여자아이(0.5739), 몸조리(0.5660), 신생아(0.5514), 예정일(0.5488), 산후조리(0.5414), 제왕절개(0.5358), 무렵(0.5207), 딸(0.5179), 미숙아(0.5168), 뱃속(0.5159), 무사하다(0.5153), 아들(0.4964)
6	낳다 /Verb	지우다(0.7049), 뱃속(0.6453), 태어나다(0.6453), 키우다(0.6135), 백일(0.6082), 책임지다(0.5695), 출산(0.5636), 돌보다(0.5465), 생기다(0.5394), 이쁘다(0.5192), 가지다(0.5173), 기르다(0.5011), 낙태(0.4902), 원치않다(0.4896), 시험관(0.4881), 첫째(0.4671), 헤어지다(0.4637), 출생신고(0.4592), 데려오다(0.4547), 입양(0.4531)
7	신랑 /Noun	어머님(0.6316), 시어머니(0.6278), 남편(0.6278), 시댁(0.6039), 시어머니(0.5750), 시아버지(0.5656), 시부모(0.5642), 아버님(0.5595), 집사람(0.5414), 와이프(0.5091), 지금껏(0.4997), 시아버님(0.4848), 식구(0.4797), 친정(0.4749), 집다(0.4724), 친정엄마(0.4719), 장사(0.4662), 거짓말(0.4661), 시누이(0.4519), 시누(0.4514)

순위	단어	단어유사도(word2vec)
8	양육 /Noun	친권(0.8745), 포기(0.7058), 양육비(0.6967), 재산분할(0.6854), 이혼 소송(0.6839), 위자료(0.6682), 친권자(0.6545), 소송(0.6369), 가져 오다(0.6337), 친권양육(0.6279), 이혼(0.6244), 키우다(0.5908), 면접 교섭권(0.5830), 협의(0.5546), 넘겨주다(0.5465), 뺏다(0.5431), 청구 소송(0.5363), 데려오다(0.5351), 승소(0.5318), 합의(0.5279)
9	양육비 /Noun	위자료(0.7931), 친권(0.6982), 양육(0.6967), 친권양육(0.6771), 재산 분할(0.6546), 소송(0.6183), 청구(0.5878), 청구소송(0.5819), 합의 (0.5767), 이혼(0.5670), 공증(0.5530), 협의(0.5451), 키우다(0.5364), 포기(0.5342), 각서(0.5329), 생활비(0.5308), 법적(0.5275), 추후 (0.5230), 친권자(0.5122), 넘겨주다(0.5034)
10	출산휴가 /Noun	육아휴직(0.8957), 복직(0.7994), 휴가(0.7954), 휴직(0.7920), 실업 급여(0.7360), 퇴사(0.7110), 무급(0.6767), 급여(0.6432), 고용보험 (0.6336), 산전(0.6289), 권고사직(0.6195), 회사(0.6164), 근로계약 서(0.6137), 대체인력(0.6124), 복귀(0.6078), 퇴직(0.6063), 사업주 (0.6018), 고용주(0.6015), 계약직(0.5979), 유급(0.5907)
11	육아휴직 /Noun	출산휴가(0.8957), 복직(0.8416), 휴직(0.8302), 휴가(0.7371), 실업 급여(0.7262), 퇴사(0.6944), 권고사직(0.6757), 복귀(0.6633), 육아휴직 급여(0.6492), 퇴직(0.6475), 대체인력(0.6465), 근로계약서(0.6416), 급여(0.6369), 무급(0.6344), 사업주(0.6330), 산전(0.6296), 고용주 (0.5984), 사업장(0.5929), 계약직(0.5905), 유급(0.5892)
12	만나다 /Verb	사귀다(0.7858), 친구(0.7160), 헤어지다(0.6718), 동거(0.6254), 채 팅(0.5993), 만남(0.5967), 주고받다(0.5776), 애인(0.5762), 약속 (0.5726), 연애(0.5697), 지인(0.5644), 교제(0.5591), 유부(0.5591), 떠나다(0.5585), 지내다(0.5540), 친하다(0.5537), 사랑(0.5414), 바람 (0.5408), 연락(0.5347), 소식(0.5279)
13	친정 /Noun	시댁(0.8000), 친정엄마(0.7967), 친정집(0.7845), 시부모(0.7814), 시어머니(0.7474), 장모님(0.7197), 시누이(0.6941), 맥(0.6926), 시 어머니(0.6925), 시아버님(0.6884), 시아버지(0.6859), 식구(0.6849), 어머니(0.6812), 집(0.6595), 제사(0.6496), 몸조리(0.6458), 분가 (0.6311), 처가(0.6201), 이모(0.6186), 아버지(0.61710)
14	싸우다 /Verb	다투다(0.8018), 싸움(0.7930), 말다툼(0.7598), 부부싸움(0.7113), 때리다(0.6944), 자주(0.6836), 손찌검(0.6705), 잠자리(0.6679), 반복 (0.6504), 사소하다(0.6470), 헤어지자(0.6420), 툴(0.6238), 다툼(0.6196), 폭언(0.6130), 몸싸움(0.6087), 외박(0.5974), 화나다(0.5951), 화해 (0.5951), 욕설(0.5910), 거짓말(0.5884)
15	혼인신고 /Noun	결혼식(0.7522), 출생신고(0.7390), 동거(0.6738), 호적(0.6312), 전입 신고(0.6250), 혼전임신(0.6119), 미루다(0.5887), 헤어지다(0.5793), 전남편(0.5782), 별거(0.5746), 파혼(0.5618), 분가(0.5548), 결혼 (0.5500), 혼인(0.5359), 미성년자(0.5348), 재혼(0.5163), 년도(0.5137), 상견례(0.5107), 주소이전(0.5063)

순위	단어	단어유사도(word2vec)
16	위자료 /Noun	재산분할(0.8266), 양육비(0.7931), 청구(0.7049), 청구소송(0.6806), 소송(0.6792), 양육(0.6682), 이혼소송(0.6585), 친권(0.6572), 친권양육(0.6555), 분할(0.6520), 이혼(0.6283), 합의(0.6221), 공증(0.6172), 승소(0.6011), 가져오다(0.5942), 이혼사유(0.5762), 법적(0.5730), 재산(0.5656), 파기(0.5655), 간통(0.5635)
17	휴가 /Noun	무급(0.8384), 휴직(0.8081), 출산휴가(0.7954), 유급(0.7751), 육아휴직(0.7371), 복직(0.7319), 산전(0.7021), 복귀(0.6670), 고용주(0.6667), 퇴직(0.6580), 병가(0.6561), 대체인력(0.6519), 육아휴직급여(0.6499), 육아휴가(0.6436), 휴무(0.6408), 실업급여(0.6335), 계약직(0.6327), 급여(0.6241), 사업주(0.6220), 퇴직금(0.6183)
18	미혼모 /Noun	미혼모시설(0.8150), 미혼모센터(0.8008), 시설입소(0.7745), 미혼(0.7527), 미성년자(0.7264), 입소(0.7160), 한부모가정(0.7121), 모이다(0.6935), 모자(0.6876), 시설(0.6630), 기초생활수급자(0.6537), 기초수급(0.6436), 모자가정(0.6420), 입양(0.6280), 헛터(0.6118), 한부모(0.5991), 미혼부(0.5905), 보호시설(0.5726), 보육료(0.5720), 대구(0.5680)
19	육 /Noun	막말(0.7795), 화나다(0.7396), 때리다(0.7282), 붙잡다(0.7113), 욕설(0.7087), 쌍욕(0.7084), 툽(0.7055), 말대꾸(0.6975), 온갖(0.6907), 이야기(0.6898), 성질(0.6878), 깃(0.6846), 말리다(0.6740), 말투(0.6708), 무시(0.6697), 소리(0.6534), 버릇(0.6533), 반말(0.6468), 달래다(0.6434), 손찌검(0.6430)
20	데리다 /Verb	데려가다(0.7857), 데려오다(0.7514), 맡기다(0.6417), 친절집(0.6197), 뺏다(0.6139), 고아원(0.6101), 낚두다(0.6037), 내려오다(0.5906), 도망(0.5892), 돌보다(0.5689), 처갓집(0.5534), 친절(0.5524), 백일(0.5512), 찾아오다(0.5475), 대리(0.5468), 간다(0.5432), 몸조리(0.5367), 키우다(0.5275), 짐(0.5170), 집(0.5084)

나. ‘영유아’ 키워드 분석결과

‘영유아’에 대한 검색결과는 많지 않았기 때문에 형태소를 ‘영아’, ‘유아’로 나누어 검색하였는데 주로 학과, 취업 등과 관련한 진로에 대한 형태소들이 대부분이었다.

〈표 4-6〉 ‘영유아’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	보육학/Noun	170	0.6250
2	정신건강론/Noun	142	0.6234

순위	단어	빈도(개)	TF-IDF
3	교수/Noun	137	0.6231
4	아동복지론/Noun	134	0.6229
5	아동발달/Noun	132	0.6228
6	유아교육/Noun	2,488	0.4942
7	유아교육학/Noun	341	0.4924
8	유치원교사/Noun	266	0.4890
9	발달/Noun	150	0.4838
10	학과/Noun	130	0.4829
11	보육/Noun	1,033	0.4396
12	수강/Noun	243	0.4099
13	평가/Noun	176	0.4074
14	아동복지/Noun	169	0.4072
15	평생교육/Noun	168	0.4071
16	전문학교/Noun	132	0.4058
17	아동복지학과/Noun	132	0.4058
18	보육교사/Noun	2,334	0.3797
19	유치원/Noun	554	0.3676
20	교사/Noun	518	0.3664

앞의 ‘임신출산’관련 유사도 분석과 마찬가지로 ‘영아’와 ‘유아’의 형태소의 유사도 분석을 실시하였는데 그 결과는 ‘영유아’의 형태소 빈도분석과 크게 다르지 않았다.

<표 4-7> ‘영유아’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	보육학 /Noun	개론(0.9434), 아동발달(0.9221), 과정(0.8957), 아동복지론(0.8857), 교수(0.8782), 방법론(0.8634), 필수(0.8598), 아동건강(0.8510), 교육실습(0.8469), 교직(0.8442), 이미지(0.8377), 교육학(0.8346), 아동문학(0.8341), 발달(0.8318), 아동발달론(0.8280), 교육(0.8130), 영양학(0.7977), 지도(0.7975), 실기(0.7941), 보육교사론(0.7933)
2	정신 건강론 /Noun	아동복지론(0.9765), 아동발달론(0.9765), 가족복지론(0.9750), 일선(0.9744), 지역사회복지론(0.9726), 교육사회학(0.9653), 방법론(0.9626), 자원봉사론(0.9527), 건강(0.9506), 아동발달(0.9503), 응급처치(0.9501), 스포츠(0.9500), 전선(0.9468), 정론(0.9440), 현장실습(0.9431), 부모교육론(0.9429), 인간관계(0.9425), 기초(0.9385), 실천기술(0.9377), 행동(0.9376)

순위	단어	단어유사도(word2vec)
3	교수 /Noun	프로그램(0.9410), 아동문학(0.9275), 방법론(0.9270), 아동발달(0.9206), 개발(0.9184), 아동건강(0.9050), 영양학(0.8921), 평가(0.8872), 아동복지론(0.8860), 아동발달론(0.8814), 발달(0.8795), 보육학(0.8782), 언어(0.8775), 보육교사론(0.8715), 지도(0.8711), 정신건강론(0.8657), 개론(0.8526), 교육실습(0.8457), 임상(0.8372), 교직(0.8286)
4	아동 복지론 /Noun	정신건강론(0.9765), 아동발달(0.9688), 개론(0.9623), 아동발달론(0.9584), 방법론(0.9565), 일선(0.9516), 지역사회복지론(0.9501), 가족복지론(0.9445), 이미지(0.9368), 현장실습(0.9360), 응급처치(0.9322), 교육사회학(0.9177), 보육교사론(0.9166), 인간(0.9147), 스포츠(0.9145), 전선(0.9133), 아동건강(0.9128), 행동(0.9118), 기초(0.9105), 리다십(0.9029)
5	아동발달 /Noun	방법론(0.9735), 아동복지론(0.9688), 아동발달론(0.9669), 개론(0.9632), 아동건강(0.9613), 정신건강론(0.9503), 교직(0.9457), 이미지(0.9367), 아동문학(0.9363), 영양학(0.9316), 보육교사론(0.9238), 보육학(0.9221), 교수(0.9206), 발달(0.9203), 교육실습(0.9148), 교육학(0.9143), 일선(0.9140), 프로그램(0.9002), 응급처치(0.8971), 임상(0.8962)
6	유아교육 /Noun	유아교육학(0.9160), 아동복지학과(0.9118), 학과(0.9092), 대학(0.8910), 대학교(0.8904), 전문대(0.8887), 진학(0.8840), 학과(0.8814), 사회복지학과(0.8479), 택하다(0.8472), 사회복지과(0.8462), 사범(0.8386), 가다(0.8375), 실업계(0.8369), 유명하다(0.8345), 고등학교(0.8242), 희망(0.8122), 학교(0.8048), 년제(0.8046), 생(0.8001)
7	유아 교육학 /Noun	실업계(0.9192), 유아교육(0.9160), 가기(0.8972), 진학(0.8953), 학과(0.8949), 대학교(0.8834), 가다(0.8814), 전문대(0.8769), 희망(0.8769), 유명하다(0.8758), 부산(0.8740), 중앙대(0.8610), 사범(0.8584), 한양(0.8551), 둘(0.8548), 택하다(0.8545), 전문학교(0.8536), 지방(0.8528), 인문(0.8525), 유치원선생님(0.8494)
8	유치원 교사 /Noun	어린이집교사(0.9710), 유치원선생님(0.9590), 어린이집선생님(0.9161), 임용고시(0.8956), 나중(0.8887), 선생님(0.8871), 병설유치원(0.8778), 목표(0.8606), 차이점(0.8596), 다르다(0.8517), 취직(0.8494), 차이(0.8477), 유치원(0.8452), 꿈(0.8350), 호봉(0.8346), 미래(0.8323), 남자(0.8323), 싫다(0.8273), 사립(0.8212), 갓다(0.8188)
9	발달 /Noun	보육교사론(0.9463), 영양학(0.9403), 아동건강(0.9359), 아동발달(0.9203), 아동발달론(0.9196), 지도(0.9165), 생활지도(0.9149), 아동문학(0.9094), 프로그램(0.8965), 방법론(0.8957), 안전(0.8951), 평가(0.8947), 영역(0.8941), 개발(0.8921), 아동복지론(0.8897), 언어(0.8874), 영양(0.8865), 정신건강론(0.8856), 놀이(0.8853), 부모(0.8838)
10	학과 /Noun	진학(0.9629), 유아교육학(0.9621), 유아교육(0.9392), 전문대(0.9151), 년제(0.8886), 둘(0.8861), 대학교(0.8813), 실업계(0.8765), 가다(0.8730), 사범(0.8697), 유명하다(0.8672), 사범대(0.8522), 들어가다(0.8518), 대학(0.8470), 아동복지학과(0.8416), 고등학교(0.8354), 택하다(0.8344)
11	보육 /Noun	영유아(0.8361), 아동복지(0.7598), 보육학(0.7445), 복지학과(0.7131), 보육학과(0.6959), 사회복지과(0.6858), 사회복지학과(0.6796), 유아(0.6630), 교수(0.6569), 교육실습(0.6489), 아동(0.6466), 심리학과(0.6446), 아동복지학과(0.6439), 교육(0.6412), 나뉘다(0.6394), 교직(0.6363), 복지과(0.6363), 특수교육(0.6321), 방법론(0.6307), 미술(0.6298)

순위	단어	단어유사도(word2vec)
12	수강 /Noun	교양(0.9355), 교양과목(0.9229), 독학사(0.9047), 개(0.9031), 일반(0.8915), 과목(0.8914), 필수과목(0.8833), 필수(0.8811), 학기(0.8809), 완료(0.8740), 이수(0.8702), 나누다(0.8634), 단계(0.8613), 가정학(0.8583), 나머지(0.8514), 총(0.8504), 남다(0.8412), 유사(0.8379), 개설(0.8313), 전필(0.8184)
13	평가 /Noun	개발(0.9776), 프로그램(0.9721), 영양학(0.9266), 아동건강(0.9231), 보육교사론(0.9214), 아동문학(0.9169), 아동발달론(0.9131), 방법론(0.9121), 부모(0.9111), 정신건강론(0.9071), 언어(0.9018), 발달(0.8947), 건강(0.8905), 안전(0.8890), 임상(0.8889), 표(0.8885), 교수(0.8872), 생활지도(0.8841), 택(0.8837), 영양(0.8822)
14	아동복지 /Noun	아동(0.8919), 복지학과(0.8777), 심리(0.8764), 미술(0.8671), 특수교육(0.8564), 복지과(0.8561), 보육학과(0.8174), 특수(0.8078), 교육평가(0.7952), 영유아(0.7866), 지도자(0.7718), 제목(0.7714), 유아(0.7692), 대구(0.7667), 교육학(0.7625), 보육(0.7598), 보건(0.7544), 유아(0.7526), 교육과정(0.7485), 미술심리치료사(0.7399)
15	평생교육 /Noun	원격(0.8748), 사이버(0.8175), 사회복지(0.7966), 과정(0.7829), 아동학(0.7795), 건강가정사(0.7769), 병행(0.7674), 평생교육사(0.7618), 완료(0.7613), 교과목(0.7603), 학점은행제(0.7562), 이수(0.7559), 사회복지사(0.7314), 사회복지학(0.7312), 온라인(0.7309), 시간제(0.7270), 강의(0.7258), 설계(0.7230), 필수과목(0.7185), 학점이수(0.7138)
16	전문학교 /Noun	한양(0.9761), 지방(0.9373), 수도권(0.9338), 가기(0.9192), 경기도(0.9092), 부산(0.9090), 희망(0.9064), 알아주다(0.9022), 서울(0.9021), 취업률(0.8962), 가다(0.8926), 경기(0.8898), 실업제(0.8842), 학과(0.8777), 중앙대(0.8672), 유교(0.8671), 특별전형(0.8616), 유명하다(0.8615), 어디(0.8608), 낮다(0.8606)
17	아동복지 학과 /Noun	보육학과(0.9504), 학과(0.9387), 유아교육(0.9118), 복지학과(0.8901), 진로(0.8654), 사회복지과(0.8360), 유아교육학(0.8195), 사회복지학과(0.8178), 복지과(0.8120), 택하다(0.8120), 학과(0.8009), 사범대(0.8000), 꿈꾸다(0.7958), 차이점(0.7893), 전주(0.7873), 한양(0.7806), 진학(0.7802), 국립(0.7756), 심리학과(0.7706)
18	보육교사 /Noun	자격증(0.9233), 따다(0.9056), 취득(0.8555), 급(0.8522), 수료(0.8503), 동시(0.8363), 고졸(0.8337), 사회복지사(0.8266), 유치(0.8239), 특수교사(0.8227), 교사자격증(0.8197), 필요하다(0.8111), 독학(0.7967), 청소년상담사(0.7956), 우선(0.7943), 원정(0.7940), 전문대졸(0.7914), 주어진다(0.7905), 사이버(0.7745)
19	유치원 /Noun	교사(0.9278), 사립(0.9088), 취직(0.9053), 병설유치원(0.9003), 정(0.8862), 선생님(0.8859), 차리다(0.8750), 어린이집(0.8749), 차이(0.8741), 원장(0.8693), 원정(0.8653), 국공립(0.8607), 쌀다(0.8563), 유치(0.8464), 유치원교사(0.8452), 교사자격증(0.8421), 임용고시(0.8416), 어린이집선생님(0.8366), 어린이집교사(0.8349), 나중(0.8274)

순위	단어	단어유사도(word2vec)
20	교사 /Noun	유치원(0.9278), 원정(0.9199), 유치(0.9082), 교사자격증(0.8820), 주 어지다(0.8682), 중등(0.8593), 차리다(0.8342), 원장(0.8207), 어린이 집(0.8147), 쌀다(0.8106), 사립(0.8017), 취직(0.7995), 유보(0.7900), 근무(0.7777), 국공립(0.7707), 나중(0.7673), 추후(0.7655), 학습지 (0.7604), 병설유치원(0.7493), 차이(0.7473)

‘영유아’라는 형태소가 전문분야를 제외하고, 일반적으로 많이 사용
되지 않을 수 있기 때문에 이와 유사한 단어인 ‘아기’ 또는 ‘아이’ 등
과 같은 단어의 조합을 통하여 연관된 형태소를 추출하였다.

<표 4-8> ‘아기’, ‘아이’ 관련 키워드 유사도

단어	단어유사도(word2vec)
아기 /Noun	다루다(0.7220), 미술(0.7180), 전담(0.7176), 돌보미(0.7171), 키즈 (0.7140), 직업(0.7115), 봉사활동(0.7068), 센터(0.7034), 갑다(0.7030), 보 호소(0.6999), 보탬(0.69733), 복지시설(0.6940), 유아원(0.6929), 아동센터 (0.6891), 보육시설(0.6877), 교구(0.6850), 한글(0.6842), 초등(0.6837), 복 지학과(0.6833), 지내다(0.6805)

앞서 ‘영유아’에 대한 분석결과가 주로 학업이나 진로에 대한 형태소가
유사도가 높은 것으로 나타난 반면에 키워드를 ‘아기’, ‘아이’로 설정
하였을 때, 육아, 복지시설과 관련한 형태소들이 높은 벡터값을 갖는
것으로 나타났다. 이는 관련된 질의를 할 경우 보편적으로 사용하는
형태소를 이용해야 보다 정확한 결과를 유도할 수 있음을 의미한다.
즉, 전문적인 형태소의 결과와 다른 것을 확인할 수 있었는데, 인터넷
상의 불특정 다수를 대상으로 하는 텍스트 데이터를 수집하여 마이닝
하는 경우 전문적인 용어보다는 일반적으로 많이 사용하는 형태소를
중심으로 문서를 수집하는 것이 중요하다.

다. ‘아동청소년’ 키워드 분석결과

‘아동청소년’ 키워드 분석결과는 앞의 ‘영유아’ 분석결과와 유사하게 학과, 취업 등과 관련한 진로에 관련된 형태소가 다빈도로 나타났다.

<표 4-9> ‘아동청소년’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	보육학/Noun	170	0.6250
2	정신건강론/Noun	142	0.6234
3	교수/Noun	137	0.6231
4	아동복지론/Noun	134	0.6229
5	아동발달/Noun	132	0.6228
6	유아교육/Noun	2,488	0.4942
7	유아교육학/Noun	341	0.4924
8	유치원교사/Noun	266	0.4890
9	발달/Noun	150	0.4838
10	학과/Noun	130	0.4829
11	보육/Noun	1,033	0.4396
12	수강/Noun	243	0.4099
13	평가/Noun	176	0.4074
14	아동복지/Noun	169	0.4072
15	평생교육/Noun	168	0.4071
16	전문학교/Noun	132	0.4058
17	아동복지학과/Noun	132	0.4058
18	보육교사/Noun	2,334	0.3797
19	유치원/Noun	554	0.3676
20	교사/Noun	518	0.3664

‘아동’, ‘청소년’ 관련된 키워드의 유사도 분석결과는 <표 4-10>과 같다. 유사도 분석결과 역시 다빈도 형태소 분석결과와 마찬가지로 학과, 직업 등의 진로와 관련된 단어들이 대부분이었다.

<표 4-10> '아동청소년' 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	심리 상담사 /Noun	미술심리치료사(0.8074), 심리치료사(0.7842), 상담사(0.6609), 심리 상담가(0.6408), 심리상담(0.6137), 지도사(0.6081), 상담심리사 (0.6043), 미술치료(0.5837), 청소년상담사(0.5624), 음악치료(0.5485), 미술심리(0.5443), 방과후지도사(0.5370), 미술(0.5295), 심리(0.5218), 복지사(0.5055), 심리치료(0.5028), 사회복지사(0.4803), 심리상담지 도사(0.4741), 민간(0.4694), 심리학(0.4663)
2	지도사 /Noun	상담사(0.6893), 지도자(0.6823), 방과후지도사(0.6635), 청소년지도 사(0.6583), 심리상담사(0.6081), 미술심리치료사(0.5619), 건강가정 사(0.5531), 평생교육사(0.5415), 레크레이션(0.5333), 자격증(0.5257), 사회복지사(0.5131), 수련(0.5105), 청소년상담사(0.4986), 지도 (0.4887), 복지사(0.4886), 상담심리사(0.4868), 종이접기(0.4678), 직 업상담사(0.4676), 급(0.4614), 교관(0.4569)
3	심리학 /Noun	심리학과(0.7785), 심리(0.6923), 상담심리학(0.6696), 상담심리(0.6156), 심리치료(0.5844), 사회복지학(0.5831), 교육학(0.5753), 임상심리학 (0.5474), 심리치료사(0.5433), 대학원(0.5303), 심리상담(0.5242), 사 회복지학과(0.5146), 음악치료(0.4867), 범죄심리학(0.4856), 아동복지 학(0.4851), 심리상담가(0.4839), 미술치료(0.4773), 학문(0.4744), 상 담(0.4642), 발달심리학(0.4636)
4	청소년 상담사 /Noun	청소년지도사(0.7730), 상담사(0.7382), 상담심리사(0.6907), 직업상 담사(0.6647), 임상심리사(0.6555), 응시자격(0.6184), 사회복지사 (0.6026), 응시(0.5880), 요건(0.5849), 심리상담사(0.5624), 사회복지 공무원(0.5191), 자격(0.5078), 지도사(0.4986), 조건(0.4980), 눈길 (0.4941), 시험(0.4897), 평생교육사(0.4865), 건강가정사(0.4752), 사 회복지학(0.4620), 자격증(0.4611)
5	상담사 /Noun	청소년상담사(0.7382), 지도사(0.6893), 심리상담사(0.6609), 사회복 지사(0.6250), 상담심리사(0.6226), 청소년지도사(0.5964), 심리상담 (0.5702), 상담(0.5677), 심리상담가(0.5485), 미술심리치료사(0.5366), 임상심리사(0.5347), 심리치료사(0.5270), 심리(0.5189), 복지사 (0.5141), 요건(0.4838), 응시자격(0.4772), 자격증(0.4737), 심리학과 (0.4651), 직업상담사(0.4585), 응시(0.4558)
6	심리 /Noun	심리치료(0.7610), 미술심리(0.7031), 심리상담(0.7027), 미술치료 (0.7018), 심리학(0.6923), 심리학과(0.6756), 음악치료(0.6496), 치료 (0.6488), 임상(0.6366), 상담심리(0.6219), 심리치료사(0.6172), 심리 검사(0.5711), 상담(0.5707), 사나(0.5601), 심리상담가(0.5434), 상담 심리사(0.5405), 상담심리학(0.5401), 임상심리사(0.5219), 심리상담 사(0.5218), 상담사(0.5189)
7	아동복지 /Noun	사회복지(0.5424), 복지(0.5405), 아동복지학과(0.5130), 사회복지과 (0.5098), 아동(0.4935), 노인복지(0.4751), 유아교육(0.4688), 청소년 복지(0.4665), 아동복지학(0.4594), 보육원(0.4299), 보육(0.4288), 유 치원교사(0.4189), 유아(0.4084), 여러가지(0.4080), 고아원(0.3862), 장애인복지(0.3836), 유치원선생님(0.3719), 아동학(0.3690), 사복 (0.3650), 여기(0.3640)

순위	단어	단어유사도(word2vec)
8	아동복지 학과 /Noun	유아교육학(0.7360), 사회복지학과(0.6987), 유아교육(0.6279), 유치원교사(0.6090), 아동복지학(0.5989), 보육학과(0.5642), 유치원선생님(0.5641), 학과(0.5462), 과(0.5403), 충북대(0.5306), 심리학과(0.5254), 전문학교(0.5235), 중앙대(0.5145), 아동복지(0.5130), 사회복지과(0.4991), 유교(0.4906), 숙명여대(0.4885), 유치원(0.4752), 전문대(0.4742), 충북대학교(0.4653)
9	유아교육 /Noun	유아교육학(0.7742), 유교(0.7183), 보육학과(0.6370), 아동복지학과(0.6279), 유치원교사(0.6144), 사회복지과(0.6044), 보육(0.6029), 복지과(0.5879), 유치원(0.5829), 유아(0.5818), 유치원선생님(0.5588), 전문대(0.5587), 과(0.5386), 숙대(0.5349), 사회복지학과(0.5179), 교유(0.5084), 한양여대(0.4906), 숙명여대(0.4895), 과중(0.4857), 교사자격증(0.4788)
10	보육 /Noun	복지과(0.6609), 보육학과(0.6605), 보육학(0.6149), 유아교육(0.6029), 사회복지과(0.5793), 영유아(0.5356), 보육교사(0.5269), 유아(0.4960), 유교(0.4846), 전문대(0.4838), 생활지도(0.4825), 동서울대(0.4821), 영아(0.4805), 보육자격증(0.4734), 영양학(0.4684), 수원여대(0.4655), 보육시설(0.4581), 어린이집(0.4488), 신구대(0.4473), 사회복지(0.4403)
11	평생교육 /Noun	사이버(0.7380), 원격(0.6933), 학점은행(0.6884), 에듀(0.6564), 사이버대(0.6369), 학점은행제(0.6340), 사이버대학(0.6302), 오프라인(0.6265), 사이버대학교(0.6162), 온라인(0.6098), 시간제(0.5952), 강의(0.5910), 교육기관(0.5863), 졸업장(0.5350), 수강(0.5304), 방송통신대학(0.5276), 학점이수(0.5146), 방통대(0.5145), 사이버강의(0.5127), 디지털대학교(0.5086)
12	사회 복지과 /Noun	복지과(0.7138), 사회복지학과(0.6835), 전문대(0.6652), 과(0.6449), 유아교육(0.6044), 년제(0.5991), 보육(0.5793), 대학(0.5508), 보육학과(0.5483), 청소년복지(0.5441), 복지학과(0.5276), 학과(0.5160), 아동복지(0.5098), 아동복지학(0.5090), 심리학과(0.5061), 전문대학교(0.5027), 노인복지(0.5007), 아동복지학과(0.4991), 수원여대(0.4986), 전문대학(0.4896)
13	학위 /Noun	학사(0.6870), 학점은행(0.6598), 학점은행제(0.6409), 아동학(0.6377), 졸업장(0.6157), 학점(0.6011), 전문(0.5893), 학점이수(0.5745), 석사(0.5727), 전문대졸(0.5632), 급(0.5591), 자격증(0.5540), 박사학위(0.5275), 박사(0.5043), 일반편입(0.4999), 수여(0.4984), 과목(0.4967), 단기간(0.4872), 시간제(0.4792), 대졸(0.4786)
14	교사 /Noun	유치원(0.7239), 교사자격증(0.6576), 특수교사(0.5972), 병설유치원(0.5954), 병설(0.5763), 중등(0.5721), 유치원교사(0.5719), 어린이집(0.5574), 임용고시(0.5485), 선생님(0.5374), 원장(0.5356), 어린이집교사(0.5356), 교원자격증(0.5254), 교육대학(0.5109), 유치원선생님(0.5057), 개원(0.4952), 교직이수(0.4932), 특수학교(0.4828), 어린이집선생님(0.4553), 초등(0.4455)
15	대학원 /Noun	석사과정(0.6740), 교육대학(0.6638), 석사(0.6495), 박사(0.6302), 상담심리(0.6109), 대학(0.5673), 교직이수(0.5389), 심리학(0.5303), 심리학과(0.5274), 박사학위(0.5274), 유학(0.5259), 사이버대학(0.5138), 방통대(0.5099), 대학교(0.4998), 사이버대(0.4987), 사이버대학교(0.4894), 상담심리사(0.4741), 코스(0.4739), 진학(0.4673), 상담심리학(0.4598)

순위	단어	단어유사도(word2vec)
16	개론 /Noun	전필(0.7984), 가족복지론(0.7819), 전선(0.7778), 조사론(0.7680), 실천론(0.7625), 자원봉사론(0.7604), 사회복지실천기술론(0.7382), 아동발달(0.7366), 지역사회복지론(0.7296), 정신건강론(0.7265), 인간관계(0.7225), 원론(0.7130), 교필(0.7016), 교육심리학(0.7009), 교육사회학(0.6993), 정론(0.6966), 일선(0.6953), 아동발달론(0.6951), 리더십(0.6898), 마케팅(0.6782)
17	수업 /Noun	강의(0.8010), 인강(0.7109), 오프라인(0.6574), 온라인(0.6480), 시간제(0.6212), 수강신청(0.6200), 강좌(0.6198), 과목(0.5925), 학점은행제(0.5779), 수강(0.5773), 교수(0.5771), 인터넷강의(0.5587), 사이버강의(0.5360), 필수과목(0.5354), 학점은행(0.5271), 개강(0.5233), 온라인강의(0.5165), 사이버(0.5162), 실습(0.5061), 평생교육(0.5018)
18	보육교사 /Noun	사회복지사(0.6527), 건강가정사(0.5862), 보육자격증(0.5833), 취득(0.5509), 급(0.5414), 학점은행제(0.5289), 보육(0.5269), 어린이집(0.5024), 아동학(0.4894), 유정(0.4877), 방과후지도사(0.4784), 복지사(0.4776), 자격증(0.4753), 유치원(0.4753), 학위(0.4635), 과목(0.4632), 학사(0.4630), 학점은행(0.4608), 평생교육사(0.4571), 필수과목(0.4535)
19	학사 /Noun	학위(0.6870), 학점은행제(0.6331), 학점(0.5983), 학점은행(0.5919), 아동학(0.5725), 학점이수(0.5440), 석사(0.5134), 과목(0.4917), 수여(0.4738), 자격증(0.4732), 전문대졸(0.4695), 단기간(0.4655), 졸업장(0.4642), 보육교사(0.4630), 문학사(0.4542), 급(0.4511), 박사학위(0.4486), 평생교육(0.4477), 취득(0.4443), 일반편입(0.4390)
20	편입 /Noun	일반편입(0.7431), 방통대(0.6623), 편입학(0.6570), 입학(0.6124), 사이버대(0.5819), 전과(0.5583), 복수전공(0.5540), 방송통신대(0.5423), 학점은행제(0.5407), 복학(0.5200), 방송대(0.5166), 자퇴(0.5122), 휴학(0.4862), 년제(0.4850), 사이버대학(0.4841), 사이버대학교(0.4817), 신입학(0.4779), 졸업(0.4755), 진학(0.4748), 학위(0.4741)

‘아동’, ‘청소년’의 키워드와 비슷한 ‘어린이’를 키워드로 설정하고 유사도를 분석한 결과, ‘불우’, ‘장애우’, ‘구호’, ‘난민’ 등의 형태소가 도출되었다.

<표 4-11> ‘어린이’ 관련 다빈도 형태소별 유사도

단어	단어유사도(word2vec)
어린이 /Noun	불우(0.66626), 장애우(0.61182), 구호(0.60911), 난민(0.60880), 고아(0.60145), 돕기(0.59752), 개발도상국(0.58338), 불우이웃(0.58106), 도시락(0.56926), 시각장애인(0.56874), 사역(0.56844), 교실(0.56772), 노약(0.56269), 중고생(0.55565), 초등생(0.55531), 영양실조(0.55461), 아프리카(0.55206), 성교육(0.54846), 학습장애(0.54552), 예방(0.54327)

라. ‘청년’ 키워드 분석결과

‘청년’ 관련된 키워드 분석결과는 <표 4-12>와 같다. ‘중소기업’, ‘창업’, ‘입사’ 등과 같은 취·창업과 관련된 진로의 키워드가 다빈도로 나타났으며, ‘군대’와 관련된 키워드도 적지 않게 나타났다.

<표 4-12> ‘청년’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	중소기업/Noun	443	0.6231
2	소득세/Noun	298	0.6205
3	창업/Noun	297	0.6205
4	군인/Noun	471	0.4836
5	교회/Noun	396	0.4826
6	폐지/Noun	281	0.4810
7	예산/Noun	256	0.4807
8	존재/Noun	376	0.4052
9	감면/Noun	298	0.4043
10	누나/Noun	296	0.4042
11	현실/Noun	287	0.4041
12	대한민국/Noun	252	0.4037
13	입사/Noun	240	0.4036
14	인생/Noun	235	0.4035
15	군/Noun	292	0.3524
16	산재/Noun	278	0.3141
17	북한/Noun	263	0.3140
18	다치다/Verb	254	0.3139
19	편입/Noun	245	0.3138

‘청년’ 형태소 유사도 분석결과는 앞서 이야기한 것처럼 취·창업과 관련된 형태소, ‘연말정산’ 등 직장생활과 관련된 형태소가 나타났다. 또한, ‘군대’와 관련되어서는 단순 입대뿐만 아니라 ‘군가산점 폐지’ 같은 사회적 이슈가 나타났다.

<표 4-13> ‘청년’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	중소기업 /Noun	인턴제(0.8401), 패키지(0.8267), 취업자(0.8246), 인턴(0.8208), 감면 신청서(0.8173), 소득세(0.7928), 입사(0.7813), 중소기업(0.7808), 세액 (0.7802), 감면(0.7709), 하청(0.7700), 진흥(0.7631), 주민세(0.7550), 연말정산(0.7525), 재직(0.7521), 정규직(0.7506), 관리직(0.7457), 연 령(0.7423), 하청업체(0.7379), 중견(0.7344)
2	소득세 /Noun	감면(0.9787), 연말정산(0.9228), 환급(0.8800), 감면신청서(0.8735), 주민세(0.8729), 세액(0.8493), 고용보험(0.8453), 소급(0.8420), 실업 급여(0.8176), 산재보험(0.8045), 연말(0.8029), 신청(0.7936), 중소기업 (0.7928), 인턴제(0.7879), 신청서(0.7866), 결정(0.7865), 인턴 (0.7865), 정산(0.7840), 노동부(0.7824), 공제(0.7734)
3	창업 /Noun	대출(0.9179), 청년창업(0.9031), 자금(0.8955), 자금대출(0.8772), 전 세자금(0.8628), 신용(0.8419), 보증(0.8317), 오픈(0.8221), 햇살론 (0.8180), 사업자(0.8156), 프랜차이즈(0.8087), 은행(0.8046), 임대 (0.8029), 자본금(0.7843), 미소금융(0.7767), 금리(0.7740), 전세금 (0.7734), 영세민(0.7682), 음식점(0.7669), 지원(0.7665)
4	군인 /Noun	직업군인(0.8067), 마찬가지로(0.8021), 모순(0.7971), 머릿수(0.7960), 최소한(0.7778), 완수(0.7645), 지키다(0.7639), 임무(0.7581), 주여지 다(0.7490), 보장(0.7397), 만족(0.7254), 부여(0.7172), 순직(0.7086), 예비(0.7081), 자유(0.7042), 적어지다(0.7040), 줄이다(0.6911), 부과 (0.6839), 강간범(0.6691), 선언(0.6676)
5	교회 /Noun	목사(0.8349), 예배(0.7704), 성당(0.7321), 신앙(0.7253), 목회 (0.7165), 선교(0.6976), 선교사(0.6830), 이웃(0.6807), 장로(0.6485), 사역(0.6447), 주변(0.6405), 가깝다(0.6274), 집사(0.6270), 실상 (0.6252), 증축(0.6186), 찬양(0.6182), 연세(0.6153), 신학(0.6152), 모 이다(0.6078), 부흥(0.6044)
6	폐지 /Noun	역차별(0.8636), 양성평등(0.8200), 모병제(0.8086), 군사(0.8068), 노 출(0.8061), 실시(0.7858), 호주제(0.7855), 비리(0.7781), 군가산점 (0.7769), 장교(0.7764), 범죄(0.7747), 성범죄자(0.7723), 없애다 (0.7698), 야기(0.7658), 망언(0.7633), 명예(0.7549), 성범죄(0.7511), 외국인(0.7504), 알아주다(0.7425), 시위(0.7423)
7	예산 /Noun	예산안(0.8424), 다음해(0.8122), 거금(0.8030), 억지(0.7809), 회식비 (0.7803), 금액(0.7693), 출처(0.7579), 수천(0.7554), 홈페이지 (0.7492), 데(0.7453), 인건비(0.7412), 부채(0.7396), 자출(0.7351), 여성가족부(0.7256), 발표(0.7200), 화분(0.7171), 청계천(0.7142), 전 체(0.7089), 절감(0.7055), 예상(0.7018)
8	존재 /Noun	부정(0.8584), 증명(0.8186), 발취(0.8097), 탄생(0.8081), 꼬리 (0.7791), 우월(0.7628), 올바르다(0.7593), 짓(0.7504), 자신(0.7452), 낙약(0.7444), 도래(0.7437), 강조(0.7403), 논증(0.7403), 남녀평등 (0.7359), 그러하다(0.7329), 정부기구(0.7307), 인위(0.7302), 초월 (0.7301), 뿐(0.7242), 모순(0.7237)

순위	단어	단어유사도(word2vec)
9	감면 /Noun	소득세(0.9787), 연말정산(0.9164), 환급(0.8796), 주민세(0.8694), 소급(0.8592), 감면신청서(0.8440), 고용보험(0.8409), 세액(0.8312), 신청(0.8180), 실업급여(0.8047), 결격(0.7991), 산재보험(0.7871), 적용(0.7779), 곤란(0.7751), 중소기업(0.7709), 패키지(0.7707), 사유(0.7648), 연말(0.7645), 생계유지(0.7619), 인턴(0.7604)
10	누나 /Noun	남자친구(0.8870), 가끔(0.8323), 조카(0.7839), 고백(0.7836), 놀다(0.7778), 형(0.7749), 갑자기(0.7716), 여자친구(0.7622), 옆(0.7554), 두렵다(0.7498), 여자애(0.7492), 동생(0.7484), 매일(0.7479), 하루하루(0.7473), 변하다(0.7397), 고집(0.7377), 도저히(0.7336), 언니(0.7316), 눈물(0.7309), 지내다(0.7294)
11	현실 /Noun	제일(0.7788), 만족(0.7498), 솔직하다(0.7433), 자유(0.7384), 적절하다(0.7226), 물론(0.7196), 모순(0.7156), 만큼(0.7138), 표현(0.7082), 만족스럽다(0.6992), 이해(0.6970), 감성(0.6902), 직업군인(0.6795), 논리(0.6794), 의미(0.6775), 나아가다(0.6774), 노력(0.6710), 증명(0.6701), 최선(0.6641), 문젯거리(0.6596)
12	대한민국 /Noun	분단국가(0.7528), 정치(0.7324), 선거(0.7157), 실업률(0.7020), 꿈꾸다(0.7008), 조국(0.6953), 투표(0.6790), 애국심(0.6749), 출마(0.6735), 정치가(0.6713), 대선(0.6652), 아름답다(0.6644), 대통령(0.6611), 평범하다(0.6602), 한국인(0.6575), 정보화(0.6508), 국회의원(0.6503), 추세(0.6488), 노예(0.6462), 특수부대(0.6416)
13	입사 /Noun	사원(0.8689), 재직(0.8439), 인턴(0.8411), 이직(0.8221), 방위산업(0.8121), 계약직(0.7947), 근무(0.7851), 중소기업(0.7813), 협력업체(0.7810), 중순(0.7764), 인턴제(0.7741), 영업(0.7724), 업체(0.7707), 패키지(0.7704), 신입(0.7630), 조리사(0.7624), 퇴사(0.7593), 연말정산(0.7580), 퇴직(0.7533), 조선소(0.7517)
14	인생 /Noun	소중하다(0.8159), 미래(0.8117), 멘토(0.7977), 좋아하다(0.7939), 선택(0.7918), 도와주다(0.7887), 고민(0.7886), 주님(0.7884), 살아오다(0.7870), 살아가다(0.7869), 마음(0.7864), 맘(0.7864), 꿈(0.7834), 어리다(0.7806), 신중하다(0.7716), 부디(0.7710), 향상(0.7681), 진지하다(0.7678), 길(0.7612), 성실하다(0.7547)
15	군 /Noun	군복무(0.8358), 필자(0.8098), 미필(0.8025), 복무(0.7728), 마치고(0.7691), 육군(0.7618), 사단(0.7617), 병장(0.7365), 해병대(0.7308), 부사관(0.7204), 사병(0.7109), 제대(0.7041), 적어지다(0.7034), 장기복무(0.7008), 출국(0.6974), 해군(0.6838), 하사(0.6829), 장교(0.6803), 현역(0.6779), 입대(0.6759)
16	산재 /Noun	산재처리(0.9325), 산재신청(0.9180), 불승인(0.8961), 통보(0.8887), 근로복지공단(0.8673), 공상처리(0.8440), 퇴원(0.8396), 승인(0.8389), 산재보험(0.8386), 허리디스크(0.8382), 산업재해(0.8358), 복직(0.8323), 휴직(0.8185), 종결(0.8128), 요양(0.8104), 합의(0.8100), 입원(0.8065), 서류(0.8032), 노동부(0.7981), 사장(0.7888)
17	북한 /Noun	통일(0.9025), 민족(0.8909), 막다(0.8739), 혁명(0.8615), 현상(0.8478), 세기(0.8436), 사회주의(0.8316), 만능(0.8315), 러시아(0.8291), 중심(0.8279), 평화(0.8264), 바탕(0.8264), 민주주의(0.8247), 추구(0.8244), 김정은(0.8179), 공산주의(0.8177), 역제(0.8167), 기독교(0.8136), 회장(0.8109), 세력(0.8103)

순위	단어	단어유사도(word2vec)
18	다치다 /Verb	허리(0.8953), 아프다(0.8682), 입원(0.8590), 부장(0.8559), 출근(0.8494), 무릎(0.8488), 손가락(0.8418), 나서다(0.8418), 산재신청(0.8395), 월요일(0.8390), 공상(0.8371), 퇴원(0.8254), 무겁다(0.8219), 과장(0.8210), 사장(0.8164), 통화(0.8100), 작업(0.8080), 넘어지다(0.8058), 물건(0.8036), 공상처리(0.8004)
19	편입 /Noun	대학원(0.9122), 방통대(0.9049), 사이버대학(0.8983), 학사(0.8885), 학점은행제(0.8850), 학위(0.8805), 디플로마(0.8760), 진학(0.8637), 스펙(0.8549), 사이버대(0.8528), 전문대(0.8490), 학원(0.8461), 사회 복지학과(0.8443), 수능(0.8398), 년제(0.8396), 대학교(0.8390), 목표(0.8379), 졸업(0.8375), 전문학교(0.8371), 대학(0.8370)

마. ‘중장년’ 키워드 분석결과

‘중장년’ 키워드의 검색결과는 많지 않았다. 앞서 데이터에 대해서 설명하였듯이, 네이버지식IN 검색결과는 2,163 건에 불과하다. 데이터 전처리과정에서 불필요한 결과를 제외하면, 데이터의 양이 더욱 줄어 165 건만 분석 가능하다. 따라서 TF-IDF를 고려한 다빈도 형태소는 추출 가능하지만, 분석 자료의 수가 적어 일반화 과정에 문제가 발생하고 있다. 아래의 표에서도 알 수 있듯이 형태소들의 빈도가 매우 낮게 나타난다. 더욱이 학습모델인 Word2vec에 따른 단어유사도는 계산 자체가 불가능하다.

<표 4-14> ‘중장년’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	모친/Noun	68	0.6313
2	형님/Noun	34	0.6233
3	앓다/Verb	32	0.6228
4	아우/Noun	28	0.6218
5	동영상/Noun	24	0.6209
6	사주/Noun	21	0.6202

순위	단어	빈도(개)	TF-IDF
7	젊다/Adjective	20	0.6199
8	현대/Noun	20	0.6199
9	운/Noun	20	0.6199
10	보통/Noun	20	0.6199
11	폭행/Noun	51	0.4865
12	교회/Noun	23	0.4813
13	출산/Noun	23	0.4044
14	자식/Noun	22	0.4042
15	무시/Noun	22	0.4042
16	느끼다/Verb	21	0.4040
17	삶/Noun	20	0.4039
18	인생/Noun	20	0.4039
19	타다/Verb	36	0.3159
20	달다/Verb	23	0.3144

바. ‘노년’ 키워드 분석결과

‘노년’ 또는 ‘노인’과 관련된 다빈도 형태소의 분석결과는 아래와 같다. ‘요양’, ‘치매’와 같은 건강과 관련된 형태소가 주로 나타났고, ‘연금’과 관련된 형태소도 확인된다. 또한, ‘고령화’, ‘인구’ 등과 같은 사회적 이슈도 도출되었다.

<표 4-15> ‘노년’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	요양원/Noun	1,307	0.6250
2	연금/Noun	1,251	0.6246
3	치매/Noun	940	0.6223
4	요양보호사/Noun	734	0.6207
5	고령화/Noun	1,062	0.4833
6	복지학과/Noun	729	0.4814
7	기초/Noun	663	0.4810
8	국민연금/Noun	1,248	0.4069

순위	단어	빈도(개)	TF-IDF
9	요양/Noun	835	0.4049
10	인구/Noun	804	0.4047
11	사회복지과/Noun	878	0.3532
12	노인복지/Noun	5,117	0.3311
13	할머니/Noun	1,754	0.3183

다빈도 형태소별 유사도를 살펴보면, 위에서 언급한 것 이외에도 기초생활보장, 차상위계층 등 저소득층에 관련된 형태소도 나타났다. 이는 노인 빈곤과 같은 사회적 이슈와 무관하지 않은 것으로 보인다.

<표 4-16> ‘노년’, ‘노인’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	요양원 /Noun	요양병원(0.81618), 요양시설(0.81344), 실버타운(0.68452), 노인복지시설(0.64490), 양로원(0.62790), 공동생활가정(0.61517), 양원(0.60431), 요양(0.60232), 사회복지법인(0.59681), 너싱홈(0.59357), 모시다(0.59114), 입소(0.57730), 개원(0.57217), 시설(0.56722), 병원(0.55879), 간병인(0.55638), 공동생활(0.52948), 치매(0.52880), 방문요양(0.52472), 설립(0.52312)
2	연금 /Noun	국민연금(0.81060), 노령연금(0.72678), 기초연금(0.67585), 개인연금(0.67033), 불로(0.64912), 수령(0.63207), 의료보험(0.63145), 노후(0.62820), 보험료(0.62813), 퇴직금(0.62716), 공무원연금(0.62173), 연금보험(0.61576), 지급(0.61085), 기초노령연금(0.60834), 유족연금(0.60501), 장애연금(0.60483), 저축(0.58741), 환급(0.57397), 해지(0.57077), 원금(0.56702)
3	치매 /Noun	증상(0.75236), 중풍(0.73095), 초기(0.72716), 치매증(0.72505), 노인성(0.71216), 우울증(0.68774), 거동(0.68773), 질환(0.67540), 환자(0.66988), 풍(0.66489), 요양병원(0.65656), 노인성질환(0.64833), 알츠하이머(0.64203), 뇌출혈(0.63893), 뇌경색(0.63422), 간병인(0.63197), 건망증(0.62160), 당뇨(0.61059), 불편하다(0.61015), 앓다(0.60073)
4	요양 보호사 /Noun	간호조무사(0.71967), 요양(0.65989), 복지사(0.64965), 재가(0.62367), 사회복지사(0.60045), 방문요양(0.56993), 양사(0.56567), 근무(0.56013), 도우미(0.53856), 교육기관(0.53563), 운전면허(0.53198), 사무실(0.52868), 센터(0.51666), 요양원(0.51616), 소지(0.51571), 주간(0.51548), 원장(0.51339), 재가복지(0.51034), 조무(0.50806), 자격증(0.50439)
5	고령화 /Noun	저출산(0.80618), 고령화사회(0.79637), 노령화(0.74117), 인구문제(0.67716), 시대(0.66392), 문제점(0.65010), 고령(0.64922), 출산율(0.63915), 인구(0.60961), 이슈(0.60299), 진행형(0.60156), 급속도(0.59968), 속제(0.59857), 대책(0.59399), 진입(0.59344), 가속(0.59136), 수행평가(0.58829), 일꾼(0.57374), 평균수명(0.57200), 증가(0.57056)

순위	단어	단어유사도(word2vec)
6	복지학과 /Noun	보건복지(0.86697), 복지과(0.85273), 체육복(0.76126), 용인대(0.74284), 아동복지학과(0.71293), 한체대(0.70897), 사회복지학과(0.70456), 남서울대(0.68492), 충주대(0.68390), 한서대(0.66255), 사회복지과(0.66250), 충주(0.64727), 대학교(0.64603), 심리상담(0.64517), 물리치료학과(0.64429), 간호학(0.64167), 대구대(0.64127), 학과(0.63771), 실업계(0.63451), 중앙대(0.62467)
7	기초 /Noun	생활수급자(0.89843), 수급대상자(0.74707), 수급(0.71604), 수급자격(0.70048), 한부모(0.69100), 기초생활(0.68210), 생활보호대상자(0.65761), 선정기준(0.65711), 수급자(0.62702), 국민기초생활(0.62486), 대상자(0.61731), 수급권자(0.61548), 한부모가정(0.61520), 주거급여(0.60815), 본인부담경감(0.59969), 의료급여(0.59927), 여부(0.59455), 저소득(0.58775), 차상위계층(0.58073), 재심사(0.57207)
8	국민연금 /Noun	연금(0.81060), 개인연금(0.71771), 공무원연금(0.70345), 연금보험(0.67789), 납부(0.66799), 보험료(0.63680), 퇴직금(0.63341), 환급(0.63204), 기초연금(0.63151), 산재보험(0.62482), 가입(0.62264), 저축(0.61069), 노령연금(0.60958), 고용보험(0.60628), 의료보험(0.60459), 납입(0.58803), 변액(0.58735), 기초노령연금(0.58520), 수령(0.57839), 건강보험(0.57644)
9	요양 /Noun	방문요양(0.69338), 장기요양(0.69170), 요양보호사(0.65989), 노인장기요양(0.64950), 간병(0.64729), 요양시설(0.62311), 요양병원(0.60551), 공동생활가정(0.60523), 요양원(0.60232), 입소(0.59999), 케어(0.59311), 장기(0.59273), 요양급여(0.59228), 간병인(0.58297), 수발(0.58287), 재활치료(0.58126), 장기요양보험(0.57637), 재가급여(0.57177), 재가(0.55754), 공동생활(0.54705)
10	인구 /Noun	증가(0.80124), 출산율(0.79144), 노령화(0.78789), 감소(0.76644), 비율(0.75549), 고령화사회(0.73604), 평균수명(0.73013), 수명(0.72121), 줄어듦(0.71541), 진입(0.70837), 분포(0.69709), 출생률(0.69405), 도시화(0.69282), 점차(0.69086), 늘어나다(0.68878), 가속(0.67787), 두드러지다(0.67680), 피라미드(0.67222), 출산률(0.67001), 추세(0.65263)
11	사회 복지과 /Noun	사회복지학과(0.87168), 유아교육(0.81280), 전문대(0.80658), 아동복지학과(0.78022), 대학교(0.77679), 년제(0.77160), 복지과(0.75026), 심리학과(0.71351), 전문대학(0.70730), 대학(0.69382), 사복(0.68939), 전문학교(0.68740), 대학원(0.67197), 계열(0.66390), 복지학과(0.66250), 사회복지학(0.65976), 시립대(0.64659), 학과(0.63773), 아동복지(0.63690), 간호학(0.63307)
12	노인복지 /Noun	복지(0.60306), 사회복지(0.50203), 사회복지학과(0.49827), 아동복지(0.49805), 아동복지학과(0.46864), 재가복지(0.46487), 사회복지사(0.45643), 재무회계(0.42139), 여러가지(0.42100), 사회복지과(0.41675), 유아교육(0.40371), 실버산업(0.39149), 청소년복지(0.38252), 청소년(0.37686), 꼭(0.36707), 비전(0.36514), 전망(0.36512), 취직(0.36396), 취업(0.35984), 장애인복지(0.35659)
13	할머니 /Noun	할아버지(0.87292), 외할머니(0.69189), 할머니(0.68767), 맥(0.67980), 손녀(0.66416), 이모(0.66202), 손자(0.64957), 아버님(0.60527), 어머님(0.60187), 삼촌(0.58746), 외할아버지(0.58342), 엄마(0.58026), 외삼촌(0.57296), 부모님(0.57008), 고모(0.55876), 연로(0.55297), 손주(0.55031), 외가(0.54617), 아버지(0.53898)

사. ‘장애인’ 키워드 분석결과

‘장애인’과 관련된 키워드는 ‘복지카드’, ‘자동차’ 등과 같은 이용 가능한 복지서비스와 연관된 형태소가 다수 확인되었으며, ‘장애등급’, ‘장애판정’과 같은 형태소도 많이 나타났다.

<표 4-17> ‘장애인’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	복지카드/Noun	4,508	0.6350
2	장애등급/Noun	2,750	0.6273
3	구매/Noun	2,039	0.6242
4	공동명의/Noun	1,817	0.6232
5	중고차/Noun	1,413	0.6214
6	구입/Noun	5,270	0.4950
7	할인/Noun	4,105	0.4911
8	판정/Noun	2,605	0.4860
9	장애인복지/Noun	1,802	0.4832
10	면제/Noun	1,696	0.4829
11	요금/Noun	1,580	0.4825
12	차량/Noun	7,006	0.3669
13	동사무소/Noun	1,482	0.3532
14	입원/Noun	1,379	0.3529
15	발급/Noun	2,000	0.3161
16	할머니/Noun	1,508	0.3150

도출된 다빈도 형태소를 보다 세부적으로 살펴보면, 앞서 언급한 것처럼 관련 복지서비스에 대한 형태소들이 많이 나타났고, 장애등급 및 장애판정과 관련된 형태소들이 많이 나타났다.

<표 4-18> ‘장애인’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	복지카드 /Noun	장애인증(0.75743), 카드(0.67463), 등록증(0.64347), 교통카드(0.63742), 장애인등록증(0.57862), 무임(0.56894), 주민등록증(0.55847), 하이패스(0.55293), 우대권(0.55164), 신분증(0.55094), 신용카드(0.53259), 민증(0.49837), 티켓(0.49411), 체크카드(0.48418), 재발(0.48263), 표지(0.47886), 보호자(0.47291), 표지판(0.47202), 급수(0.47120), 증명서(0.45948), 고속도로(0.45847), 지하철(0.45758), 무임승차(0.45628)
2	장애등급 /Noun	장애판정(0.82003), 장애등급(0.70654), 판정(0.70534), 장애인단(0.67787), 등급(0.66448), 장애등록(0.60482), 장애(0.59863), 재심사(0.59569), 장애연금(0.59193), 재판정(0.57812), 급수(0.57661), 시력장애(0.56464), 장해(0.55468), 산재(0.53019), 정신장애(0.52533), 뇌병변(0.52110), 재검(0.51427), 혹(0.50935), 시각장애(0.50200), 재검사(0.49503), 장애인연금(0.49429), 뇌병변장애(0.49258), 진단(0.48901)
3	구매 /Noun	구입(0.82250), 사려(0.68164), 공동명의(0.64578), 명의변경(0.63652), 운행(0.62542), 중고차(0.60020), 중고(0.59391), 신차(0.58615), 대차(0.57179), 판매(0.56455), 출고(0.56223), 폐차(0.55878), 소유(0.55873), 싼타페(0.55810), 할부(0.55213), 엘피(0.55115), 가솔린(0.55034), 리스(0.54939), 휘발유(0.54333), 차량(0.54276), 인수(0.54212), 보유(0.53915), 오펜타(0.53562)
4	공동명의 /Noun	명의(0.79024), 지분(0.68950), 구입(0.68012), 차량(0.67965), 주소지(0.66794), 장애인차량(0.66075), 등본(0.65319), 면허증(0.64709), 구매(0.64578), 주소이전(0.64511), 단독(0.62342), 명의인(0.61361), 제차(0.60888), 주소(0.60456), 분가(0.60359), 외할아버지(0.60250), 주민등록(0.59887), 명의변경(0.59778), 신차(0.59604), 거주지(0.59458), 세대원(0.59037), 새차(0.59014), 세대주(0.58585)
5	중고차 /Noun	중고(0.83939), 신차(0.77562), 새차(0.77541), 차량(0.69011), 신형(0.69007), 소나타(0.68677), 그랜저(0.67357), 아반떼(0.67127), 중형차(0.67057), 자동차(0.65858), 차종(0.65757), 소형차(0.65535), 차(0.65405), 로체(0.65376), 쏘나타(0.65139), 토스카(0.64703), 중형(0.62850), 사려(0.62399), 리갈(0.61649), 마티즈(0.61236), 스포티지(0.60678), 액티언(0.60614), 승용차(0.60189)
6	구입 /Noun	구매(0.82250), 공동명의(0.68012), 명의변경(0.62202), 운행(0.58195), 중고(0.56043), 가스(0.55346), 엘피(0.54850), 소유(0.54568), 신차(0.54216), 중고차(0.54068), 폐차(0.53825), 휘발유(0.53367), 보유(0.53290), 면세(0.53254), 사면(0.52860), 주유(0.52693), 차량(0.52243), 할부(0.52238), 장만(0.51420), 명의(0.50556), 싼타페(0.49680), 소나타(0.48903), 차사(0.48861)
7	할인 /Noun	할인혜택(0.81521), 혜택(0.61042), 요금(0.60100), 예매(0.59377), 기변(0.58781), 감면(0.57477), 혜택(0.56449), 개통(0.55540), 신규가입(0.54410), 휴대폰(0.53135), 통행료(0.52056), 가입비(0.51915), 명의변경(0.51510), 충전(0.51273), 승차(0.51263), 티켓(0.51005), 명의자(0.50816), 회선(0.50782), 폰(0.50499), 예약(0.50294), 핸드폰(0.49897), 하이패스(0.49883), 가스(0.49863)

순위	단어	단어유사도(word2vec)
8	판정 /Noun	장애판정(0.78529), 장애등급(0.70534), 판명(0.65685), 재검(0.64451), 재판정(0.64375), 재심사(0.64289), 재검사(0.64054), 신검(0.63686), 장애진단(0.62424), 장애등급(0.60924), 뇌병변(0.60421), 신체검사(0.58982), 등급(0.58228), 질병검사(0.57477), 뇌병변장애(0.57219), 진단(0.57169), 장애등록(0.56639), 시력장애(0.56236), 뇌수술(0.56012), 뇌출혈(0.54892), 안면장애(0.54838), 급(0.53831), 급수(0.53580)
9	장애인 복지 /Noun	노인복지(0.50779), 복지(0.49129), 아동복지(0.45781), 사회복지(0.43645), 장애인(0.36389), 심장장애(0.34647), 소지(0.34032), 청 소년복지(0.33830), 다문화가족지원(0.32258), 롯데(0.31429), 가족복지(0.31300), 시각장애(0.31067), 발급(0.30832), 바뀌다(0.30634), 졸업생(0.30312), 장애인등록증(0.30167), 대관(0.29930), 문화누리(0.29887), 산업재해 보상보험(0.29796), 기초생활보장(0.29784), 사회복지과(0.29570), 실버(0.28929), 아동학(0.28494)
10	면제 /Noun	감면(0.68793), 군면제(0.63793), 군대면제(0.63061), 등록세(0.62595), 취득등록세(0.59381), 취득세(0.58310), 신검(0.58128), 재검(0.57865), 자동차세(0.57723), 질병검사(0.56686), 신체검사(0.54575), 예비군(0.53804), 민방위(0.51633), 현역(0.51465), 병무청(0.50887), 공채(0.50234), 병역면제(0.49434), 면세(0.48599), 생계유지(0.48055), 혜택(0.47817), 판정(0.47487), 생계곤란(0.47341), 가입비(0.46561)
11	요금 /Noun	요금제(0.68657), 기변(0.68496), 핸드폰(0.67014), 할부금(0.66813), 가입비(0.65958), 폰(0.65327), 스마트폰(0.64771), 휴대폰(0.64528), 통신사(0.64426), 신규가입(0.63523), 데이터(0.62892), 기기(0.62876), 무제한(0.62761), 갤럭시(0.62644), 모뎀(0.61653), 텔레콤(0.61582), 위약금(0.61220), 옵티머스(0.60697), 올레(0.60449), 할인(0.60100), 올인원(0.59692), 약정(0.59084), 파워콤(0.58528)
12	차량 /Noun	장애인차량(0.81140), 승용차(0.76020), 신차(0.71635), 중고차(0.69011), 공동명의(0.67965), 차(0.66799), 자동차(0.66597), 새차(0.65938), 휘발유(0.65381), 중고(0.64870), 차갑다(0.62269), 중형차(0.61392), 그랜저(0.61114), 소형차(0.60985), 소나타(0.60877), 신형(0.60479), 디젤(0.59206), 쏘렌토(0.58688), 중형(0.58534), 스포티지(0.58008), 레조(0.56764), 싼타페(0.56634), 스타렉스(0.56371)
13	동사무소 /Noun	주민센터(0.81722), 면사무소(0.78725), 구청(0.66988), 읍사무소(0.66760), 시청(0.55703), 신청(0.53871), 담당자(0.53172), 진단서(0.51640), 병무청(0.51614), 우편(0.50589), 장애인증(0.50028), 병원(0.49228), 발부(0.49136), 공단(0.48781), 장애진단(0.48675), 복지과(0.48661), 군청(0.48122), 양육수당(0.47611), 발급(0.47057), 통보(0.46718), 물어 보다(0.45672), 장애등록(0.45012)
14	입원 /Noun	통원치료(0.78860), 퇴원(0.76693), 입원치료(0.73139), 중환자실(0.67977), 통원(0.65964), 항암치료(0.65826), 입원비(0.62553), 진료(0.62264), 간병(0.62168), 재활치료(0.62116), 정신병원(0.60724), 발병(0.59027), 쓰러지다(0.58123), 치료(0.57896), 물리치료(0.57637), 간병인(0.56933), 투석(0.56498), 요양병원(0.55060), 종합병원(0.55001), 보험처리(0.54953), 수술비(0.54937), 치료비(0.54863), 재활병원(0.54678)

순위	단어	단어유사도(word2vec)
15	발급 /Noun	재발(0.72346), 장애인등록증(0.64826), 장애인증(0.62117), 무임승차(0.59156), 부착(0.57882), 발부(0.57205), 장애등록(0.56488), 등록증(0.56291), 교통카드(0.55392), 무임(0.54900), 반납(0.54366), 장애인단(0.54148), 표지(0.52701), 재판정(0.51920), 주민등록증(0.51847), 사본(0.50874), 표지판(0.50782), 신분증(0.49906), 소지(0.49281), 판정(0.48266), 분실(0.48179), 스티커(0.48163), 장애판정(0.47505)
16	할머니 /Noun	할아버지(0.84877), 외할머니(0.72623), 외할아버지(0.69831), 큰아버지(0.69402), 외삼촌(0.68592), 삼촌(0.66576), 할머니(0.66518), 손자(0.66229), 고모(0.64898), 어머니(0.63010), 외숙모(0.60011), 아버지(0.59248), 장모님(0.58509), 손녀(0.58115), 이모(0.58015), 아버님(0.57460), 어머니께(0.56456), 시아버지(0.55589), 모친(0.54954), 작은아버지(0.54930), 어머니(0.53087), 아부지(0.51746), 누님(0.51295)

아. ‘한부모’ 키워드 분석결과

‘한부모’와 관련된 형태소들은 주로 복지서비스와 관련된 것이 주로 분석되었다. 특히 ‘복지급여’, ‘양육비’, ‘임대아파트’, ‘국가장학금’과 같은 형태소들이 도출되었으며, 이를 신청하기 위한 ‘등본’, ‘증명서’ 등과 같은 형태소도 나타났다.

<표 4-19> ‘한부모’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	한부모가족/Noun	507	0.6348
2	등본/Noun	207	0.6232
3	복지급여/Noun	198	0.6229
4	오빠/Noun	161	0.6214
5	한부모가정/Noun	2,175	0.5423
6	양육비/Noun	467	0.4911
7	양육/Noun	259	0.4849
8	미혼모/Noun	196	0.4830
9	친정/Noun	195	0.4830
10	지원/Noun	194	0.4829
11	임대아파트/Noun	175	0.4824
12	년생/Noun	174	0.4823

순위	단어	빈도(개)	TF-IDF
13	키우다/Verb	608	0.4161
14	국가장학금/Noun	453	0.4122
15	증명서/Noun	444	0.4120
16	장학금/Noun	335	0.4093
17	대상자/Noun	322	0.4089
18	가구/Noun	294	0.4082
19	학비/Noun	180	0.4054
20	대학생/Noun	177	0.4053

보다 구체적으로 살펴보면 정확한 복지서비스명도 언급되고 있다. 예를 들어, ‘기초생활보장’, ‘우선돌봄차상위’, ‘법정’, ‘임대주택’ 등이 유사도 분석결과 나타났다. 또한, 자격기준 및 서비스 신청에 필요한 서류와 관련된 형태소들도 언급되었다. 한편, ‘국가장학금’이 다빈도 형태소로 나타났는데, 지원기준에 있어 ‘한부모가정’의 대학생이 국가 장학금 지원대상자에 포함되기 때문으로 판단된다.

<표 4-20> ‘한부모’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	한부모 가족 /Noun	보호대상자(0.76918), 대상자(0.75390), 조손(0.74904), 지원대상자(0.73725), 해당(0.71673), 지원대상(0.71160), 포함(0.71043), 지원신청(0.69846), 법정(0.69433), 여야(0.69275), 우선돌봄차상위(0.69209), 선정(0.69002), 수급(0.68702), 저소득층(0.68444), 법(0.67864), 저소득(0.66932), 대상(0.66686), 부모(0.66582), 한부모가정(0.66155), 차상위계층(0.66134)
2	등본 /Noun	동거인(0.88132), 주민등록(0.87882), 세대주(0.87522), 주소지(0.86826), 자식(0.82934), 미성년자(0.81932), 성인(0.80364), 분다(0.79105), 형제(0.78951), 분리(0.78817), 저(0.78129), 부모님(0.77187), 할머니(0.75612), 어머니(0.75509), 할아버지(0.74643), 돌아가다(0.74010), 호적(0.73971), 동생(0.73440), 등록(0.73351), 주소(0.73110)
3	복지급여 /Noun	수급(0.95302), 대상자(0.91821), 우선돌봄(0.91050), 우선돌봄차상위(0.89377), 지원대상자(0.87286), 분류(0.87187), 차상위계층(0.86167), 확인서(0.86105), 입(0.85114), 의료급여(0.85044), 속(0.84828), 결정통지서(0.84307), 자활(0.84164), 경감(0.83109), 지원사업(0.82919), 본인부담경감(0.81909), 장애수당(0.81782), 분위(0.81257), 부담금(0.80782), 수혜(0.80130)

순위	단어	단어유사도(word2vec)
4	오빠/ Noun	언니(0.97833), 할머니(0.95269), 동생(0.93553), 누나(0.91577), 할아버지(0.90804), 돌아가다(0.90147), 형(0.89391), 남동생(0.87956), 둘이서(0.87081), 여동생(0.86818), 계시다(0.85595), 남매(0.84708), 엄마(0.83527), 어머니(0.82018), 분다(0.81341), 댁(0.80610), 어머니(0.79602), 성인(0.79546), 외할머니(0.78951), 아버지(0.78371)
5	한부모 가정 /Noun	기초생활수급자(0.89938), 기초수급(0.89057), 기초생활수급(0.81148), 다르다(0.77477), 한부모(0.73518), 조건(0.73462), 궁금하다(0.73004), 저소득층(0.72275), 혜택(0.70918), 차상위계층(0.69246), 기초생활(0.68764), 어떨다(0.68327), 수급자(0.67717), 지원이(0.67441), 자격(0.67032), 성인(0.66980), 한부모가족(0.66155), 선정(0.66052), 법정(0.65992), 지원금(0.65954)
6	양육비 /Noun	위자료(0.85405), 일체(0.78052), 합의(0.74111), 명당(0.74099), 양육수당(0.72812), 재산분할(0.71985), 남편(0.70521), 전 남편(0.69990), 주일(0.69031), 아내(0.68522), 아이(0.68491), 양육(0.68237), 매달(0.67930), 개월(0.66502), 친권(0.66431), 개인회생(0.66207), 출생신고(0.66192), 넘겨주다(0.65817), 풀다(0.65186), 동안(0.64641)
7	양육 /Noun	친권(0.95966), 합의(0.84544), 아내(0.83196), 넘겨주다(0.82825), 숙려기간(0.81830), 친권자(0.81709), 재판(0.80767), 협의(0.80231), 재혼(0.79771), 이혼(0.77221), 동거(0.76564), 부부(0.76373), 출생신고(0.75588), 유기(0.74855), 혼인신고(0.74649), 아이(0.74213), 미성년(0.73593), 재산분할(0.71988), 위자료(0.71021), 임신(0.71010)
8	미혼모 /Noun	싱글맘(0.85469), 맞춤(0.79593), 모자가정(0.79224), 장려금(0.79023), 모자(0.76861), 부자가정(0.74251), 기초생활수급(0.73086), 관하다(0.72141), 전세임대(0.72012), 모이다(0.69801), 초등생(0.69601), 미혼(0.69110), 다문화가정(0.68586), 입소(0.68244), 보육원(0.68154), 긴급(0.67978), 청소년(0.67823), 보육료(0.67525), 시스템(0.67279), 출산(0.66997)
9	친정 /Noun	친정집(0.94502), 친정엄마(0.93487), 었다(0.92558), 지내다(0.90071), 주소(0.88384), 전남편(0.88243), 신용불량자(0.87348), 신랑(0.85452), 살다(0.85247), 원룸(0.84069), 친오빠(0.83192), 집(0.82547), 가지다(0.81650), 댁(0.81607), 혼자(0.81367), 작다(0.80337), 전업주부(0.79789), 어머니(0.79277), 시부모(0.78830), 전혀(0.78693)
10	지원 /Noun	궁금하다(0.85200), 언제(0.82484), 꼬다(0.79372), 가능하다(0.78916), 어떠하다(0.77609), 얼마나(0.77267), 어떨다(0.76958), 건지다(0.76754), 무엇(0.76709), 지원금(0.74467), 기초수급(0.74325), 조건(0.73886), 가능(0.72647), 저소득층(0.70587), 수혜(0.70439), 찾아보다(0.70248), 지원금액(0.70054), 자격(0.69895), 선정(0.69793)
11	임대 아파트 /Noun	임대주택(0.91982), 임대(0.82576), 모자가정(0.81951), 전세임대(0.80801), 영구임대(0.78572), 가능하다(0.75051), 기초생활수급(0.73953), 지원금(0.73339), 전세대출(0.72812), 들어가다(0.72546), 조건(0.71462), 흑시(0.71316), 궁금하다(0.71299), 기초수급(0.70456), 싱글맘(0.70152), 영세민(0.70127), 기초생활수급자(0.69259), 국민임대(0.69052), 양육수당(0.68064), 매입(0.67698)

순위	단어	단어유사도(word2vec)
12	년 /Noun	대학생(0.93558), 고등학생(0.91378), 여동생(0.88435), 남동생(0.86244), 누나(0.84086), 세입(0.83564), 살이(0.82789), 명(0.82257), 형(0.81654), 단돌(0.81577), 둘이서(0.81142), 나이(0.80779), 중학생(0.80589), 후반(0.80240), 스무살(0.78268), 생일(0.77980), 구성(0.77939), 직장인(0.77812), 돌아가다(0.77498), 식구(0.76876)
13	키우다 /Verb	아이(0.94525), 아기(0.89800), 혼자(0.88794), 낳다(0.87392), 딸아이(0.85418), 애(0.81945), 데리다(0.81329), 딸(0.81005), 둘째(0.80622), 혼인신고(0.80049), 출생신고(0.79012), 아가(0.78565), 뱃속(0.76848), 아내(0.76738), 전남편(0.76336), 친권(0.75884), 모이다(0.75031), 아들(0.75017), 친정집(0.74928), 두다(0.74750)
14	국가장학금 /Noun	장학금(0.87008), 한국장학재단(0.86087), 학기(0.81630), 소득분위(0.79619), 신입생(0.79440), 학자금대출(0.79165), 유형(0.79010), 분위(0.78907), 든든학자금(0.74729), 학번(0.73215), 장학(0.72628), 드림(0.70974), 등록금(0.69995), 장학재단(0.69979), 드림장학금(0.69903), 신청(0.69339), 재수(0.69074), 차상위계층(0.69032), 대학(0.68344), 낳다(0.67734)
15	증명서 /Noun	떼다(0.94601), 가족관계증명서(0.92245), 발급(0.91176), 팩스(0.86856), 확인서(0.85261), 체크(0.83982), 제출(0.83495), 서류(0.79578), 증빙(0.78995), 수급(0.78326), 우선돌봄차상위(0.77452), 증명(0.76951), 결정통지서(0.76869), 서약(0.76701), 동의(0.76385), 우선돌봄(0.76208), 출하(0.75952), 복지급여(0.75883), 문화누리(0.75389), 본인부담경감(0.75376)
16	장학금 /Noun	학자금대출(0.88021), 장학(0.87711), 국가장학금(0.87008), 장학재단(0.85989), 한국장학재단(0.84265), 등록금(0.80783), 드림(0.80143), 분위(0.78101), 유형(0.76485), 낳다(0.75589), 수혜(0.75491), 교내(0.75131), 국가(0.74281), 면제(0.72463), 대학(0.72120), 소득분위(0.71867), 신입생(0.71632), 원서(0.71537), 학기(0.71515), 승인(0.70909)
17	대상자 /Noun	수급(0.92277), 복지급여(0.91821), 우선돌봄차상위(0.90373), 우선돌봄(0.90305), 의료급여(0.89786), 확인서(0.89739), 지원대상자(0.89726), 지원사업(0.89196), 자활(0.88977), 경감(0.88084), 결정통지서(0.87726), 대상(0.87708), 장애수당(0.87603), 부담금(0.86838), 본인부담경감(0.86389), 수급권자(0.85657), 부가급여(0.83556), 보호대상자(0.83159), 장애인연금(0.82771), 분류(0.82698)
18	가구 /Noun	최저생계비(0.93342), 이하(0.91997), 기준(0.89953), 연간(0.89490), 소득인정액(0.88566), 액(0.85610), 포함(0.83541), 초과(0.83442), 연소득(0.82772), 제외(0.82400), 건강보험료(0.82292), 급여(0.80840), 미만(0.79971), 농어촌(0.79417), 형성(0.79356), 아반떼(0.78947), 소득평가액(0.78600), 다세대(0.78253), 소득환산액(0.78058), 여야(0.77662)
19	학비 /Noun	학교(0.90573), 급식비(0.89242), 등록금(0.83777), 사립(0.74894), 급식(0.72795), 대학(0.70974), 지원(0.70500), 끓기다(0.69474), 면제(0.69401), 고등학교(0.69264), 교육비(0.65986), 들어가다(0.65529), 대학가(0.65190), 내년(0.63639), 방과후(0.63233), 다니다(0.63196), 운영비(0.62488), 입학금(0.61783), 진학(0.61226), 대학교(0.61144)

순위	단어	단어유사도(word2vec)
20	대학생/Noun	년생(0.93558), 고등학생(0.91516), 올해(0.89545), 살이(0.84680), 학생(0.84313), 살인(0.83457), 학년(0.82453), 곱다(0.82022), 동생(0.81710), 누나(0.81492), 단둘(0.80364), 월생(0.80183), 올라가다(0.80141), 남동생(0.79978), 여동생(0.79535), 후반(0.78387), 재학(0.76304), 스무살(0.76266), 중학생(0.76236), 직장인(0.76095)

자. ‘다문화(새터민)’ 키워드 분석결과

다빈도 형태소의 결과로 ‘중국인’, ‘조선족’, ‘필리핀’, ‘베트남’ 등 외국인 관련 형태소가 다빈도로 나타났고, 이와 함께 ‘기득권’, ‘차별’ 등과 같은 형태소들도 확인되었다. 결국, 다문화는 동아시아 국적의 사람들을 의미하는 형태소가 높은 빈도로 표현되었고, 동시에 그들에 대한 차별이 사회적 이슈로서 대두됨을 알 수 있다.

<표 4-21> ‘다문화(새터민)’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	다문화가정/Noun	311	0.6454
2	중국인/Noun	100	0.6249
3	한국어/Noun	83	0.6233
4	한국인/Noun	82	0.6232
5	기득권/Noun	70	0.6220
6	조선족/Noun	62	0.6212
7	국적/Noun	60	0.6211
8	필리핀/Noun	60	0.6211
9	외세/Noun	58	0.6209
10	베트남/Noun	56	0.6207
11	동아리/Noun	56	0.6207
12	외국인/Noun	517	0.5161
13	외국/Noun	139	0.4876
14	민족/Noun	107	0.4852
15	차별/Noun	104	0.4850
16	통일/Noun	85	0.4835
17	상담사/Noun	70	0.4824

순위	단어	빈도(개)	TF-IDF
18	전선/Noun	65	0.4820
19	중국/Noun	223	0.4149
20	국제/Noun	64	0.4049

‘다문화’, ‘새터민’ 형태소의 유사도 분석결과에서 ‘다문화가정’에 대한 정책적 지원에 대한 형태소들이 나타났고, 외국인에 대한 ‘차별’, ‘불법체류자’, ‘범죄자’, 외국인으로 인한 ‘역차별’ 등 부정적인 의미를 전달하는 형태소들이 다수 나타났다.

<표 4-22> ‘다문화’, ‘새터민’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	다문화 가정 /Noun	센터(0.98829), 상담(0.97445), 다문화(0.96433), 가정(0.96187), 다문화 가족지원(0.95491), 관하다(0.95484), 프로그램(0.94336), 리더십(0.94260), 노인복지(0.94183), 청소년(0.94166), 활동(0.94156), 학과(0.94016), 관련(0.93888), 다문화가족(0.93690), 상담사(0.93681), 분야(0.93509), 가족복지론(0.93495), 사회학(0.93476), 아동(0.93322), 아동센터(0.93097)
2	중국인 /Noun	침탈(0.98658), 외국(0.98543), 기득권(0.97940), 들이다(0.97642), 민족(0.97612), 잠식(0.97493), 외세(0.97400), 역사(0.97325), 한국 인(0.96943), 조선족(0.96562), 끝다(0.96346), 중국(0.96314), 한국 (0.96292), 단일(0.96222), 일본(0.95858), 매국(0.95804), 자국 (0.95433), 역차별(0.95374), 우리(0.95259), 몰려들다(0.95248)
3	한국어 /Noun	교원자격증(0.99751), 취직(0.99560), 진학(0.99384), 실습(0.99260), 상담심리(0.99137), 대학교(0.99090), 보육교사(0.99008), 대학(0.99000), 사회복지학과(0.98999), 응시(0.98934), 학사(0.98926), 소지(0.98800), 학점은행제(0.98716), 사이버대학(0.98583), 대학원(0.98495), 년제(0.98394), 자격(0.98394), 수업(0.98316), 학위(0.98296), 사회복지과(0.98252)
4	한국인 /Noun	기득권(0.99544), 일본(0.99033), 외세(0.99008), 민족(0.98886), 자국 (0.98861), 우리(0.98718), 먹다(0.98708), 침탈(0.98542), 역사 (0.98499), 매국(0.98485), 강대(0.97850), 시비(0.97811), 들이다 (0.97807), 몰려들다(0.96956), 중국인(0.96943), 식민지(0.96751), 단 일(0.96683), 조선족(0.96654), 힘쓰다(0.96634), 범죄자(0.96563)
5	기득권 /Noun	외세(0.99616), 한국인(0.99544), 역사(0.99416), 민족(0.99401), 침 탈(0.99279), 일본(0.99217), 우리(0.99044), 매국(0.98879), 자국 (0.98833), 들이다(0.98825), 단일(0.98301), 강대(0.98247), 먹다 (0.98208), 중국인(0.97940), 시비(0.97808), 몰려들다(0.97534), 잠 식(0.97401), 식민지(0.97277), 조선족(0.97173), 끝다(0.96920)

순위	단어	단어유사도(word2vec)
6	조선족 /Noun	꼴다(0.99741), 잡식(0.99521), 중국화(0.99445), 제주도(0.99399), 무비(0.99384), 투자(0.99341), 포석(0.99286), 전략(0.99269), 천지(0.99199), 추종(0.99143), 민제(0.99140), 식민지(0.99134), 범죄자(0.99127), 짱개(0.99121), 들이다(0.99118), 자본(0.99009), 뜯다(0.98856), 매국(0.98834), 영향력(0.98754), 지자체(0.98574)
7	국적 /Noun	태어나다(0.99011), 경우(0.98936), 배우자(0.98721), 아내(0.98601), 싹다(0.98574), 들어오다(0.98546), 아빠(0.98526), 혼인신고(0.98433), 만원(0.98296), 돌아가다(0.98238), 고구려(0.98224), 버리다(0.98207), 부끄럽다(0.98174), 러시아(0.98083), 등본(0.98069), 베트남(0.97986), 지랄(0.97929), 합의(0.97913), 땅(0.97912), 자기(0.97890)
8	필리핀 /Noun	충성(0.99408), 처벌(0.99388), 힘쓰다(0.99309), 약소민족(0.99156), 대한민국(0.99145), 일조(0.99138), 식민지(0.99113), 인격(0.99096), 상류층(0.99092), 살기(0.99068), 귀화(0.99048), 저열(0.99045), 의심(0.99041), 뇌물(0.99030), 깃(0.99000), 취급(0.98995), 매국노(0.98974), 추종(0.98969), 천지(0.98932), 미국(0.98868)
9	외세 /Noun	역사(0.99854), 매국(0.99711), 기득권(0.99616), 들이다(0.99577), 자국(0.99507), 강대(0.99339), 일본(0.99094), 한국인(0.99008), 식민지(0.98788), 몰려들다(0.98708), 잡식(0.98610), 단일(0.98565), 조선족(0.98550), 꼴다(0.98465), 우리(0.98447), 범죄자(0.98445), 짱개(0.98394), 침탈(0.98355), 중국화(0.98340), 힘쓰다(0.98235)
10	베트남 /Noun	아버지(0.99662), 아빠(0.99662), 아내(0.99620), 어머니(0.99570), 혼인신고(0.99499), 아들(0.99468), 보내다(0.99449), 엄마(0.99440), 별거(0.99342), 오빠(0.99234), 오기(0.99086), 가솔신고(0.99082), 벌다(0.99028), 나가다(0.98972), 생활비(0.98936), 합의(0.98914), 버리다(0.98883), 외박(0.98721), 애(0.98713), 누나(0.98676)
11	동아리 /Noun	양로원(0.99122), 자원봉사(0.98974), 새터민(0.98798), 인천(0.98624), 과제(0.98585), 독서(0.98582), 저소득층(0.98565), 내용(0.98520), 돕다(0.98460), 고아원(0.98345), 글짓기(0.98337), 세부(0.98298), 복지학과(0.98166), 다큐멘터리(0.98116), 도서관(0.98102), 치료(0.98091), 토론(0.98058), 장소(0.98058), 돕기(0.98017), 전형(0.97998)
12	외국인 /Noun	국민(0.99151), 시키다(0.92700), 외국(0.92425), 만들다(0.92023), 역차별(0.92015), 차별(0.91628), 특혜(0.89855), 국가(0.89725), 정부(0.89449), 양산(0.89091), 정치(0.87505), 중국인(0.87481), 우리나라(0.87307), 잡식(0.85972), 경제(0.85366), 침탈(0.85231), 불법체류자(0.85170), 꼴다(0.85098), 나라(0.84890), 한국(0.84582)
13	외국 /Noun	침탈(0.98645), 중국인(0.98543), 나라(0.96854), 민족(0.96819), 기득권(0.96728), 한국(0.96725), 들이다(0.96459), 만들다(0.96245), 역차별(0.96198), 잡식(0.95906), 외세(0.95905), 한국인(0.95894), 중국(0.95834), 단일(0.95559), 국민(0.95427), 우리나라(0.95422), 시키다(0.95409), 역사(0.95268), 꼴다(0.94839), 조선족(0.94788)
14	민족 /Noun	침탈(0.99546), 기득권(0.99401), 우리(0.99019), 한국인(0.98886), 일본(0.98282), 먹다(0.98234), 외세(0.98169), 역사(0.97994), 단일(0.97658), 중국인(0.97612), 들이다(0.96973), 외국(0.96819), 자국(0.96771), 시비(0.96754), 매국(0.96751), 나라(0.96632), 중국(0.96609), 한국(0.96518), 강대(0.95836), 잡식(0.94975)

순위	단어	단어유사도(word2vec)
15	차별 /Noun	시키다(0.99520), 양산(0.99361), 역차별(0.99319), 만들다(0.99080), 정부(0.98875), 불법체류자(0.98732), 국가(0.98509), 남발(0.98474), 침해(0.98333), 실정(0.98160), 인권(0.97785), 법안(0.97768), 영향력(0.97733), 골다(0.97567), 자본(0.97478), 정치(0.97454), 뜯다(0.97417), 특혜(0.97337), 이자스민(0.97324), 유입(0.97221)
16	통일 /Noun	남한(0.99692), 단군(0.98596), 북한(0.98238), 약하다(0.98171), 동북공정(0.98156), 보이다(0.98046), 싫어하다(0.97872), 틀리다(0.97858), 옛날(0.97759), 무섭다(0.97756), 내수(0.97719), 심하다(0.97692), 잘못(0.97623), 반대(0.97617), 좁다(0.97557), 스럽다(0.97553), 지키다(0.97550), 하필(0.97541), 영국(0.97523), 감정(0.97517)
17	상담사 /Noun	분야(0.99324), 건강가정사(0.99013), 학과(0.98930), 가정상담사(0.98767), 자격(0.98664), 관심(0.98593), 기관(0.98363), 학생(0.97761), 소지(0.97749), 계열(0.97724), 학교(0.97512), 사이버대학(0.97389), 사회복지학과(0.97258), 공인(0.97249), 복수전공(0.96938), 관련(0.96847), 지도사(0.96779), 취직(0.96691), 미술회화(0.96546), 상담(0.96358)
18	전선 /Noun	문화(0.96354), 전필(0.91761), 문제(0.91455), 사회(0.90599), 경제(0.89566), 탐구(0.88871), 체육관광부(0.87702), 보호(0.87693), 현대(0.87597), 국무총리(0.87061), 보건 의료(0.87002), 예술(0.86227), 피해자(0.86058), 등(0.86010), 위(0.85955), 여성(0.85678), 빈곤(0.85562), 방지(0.85517), 고려인(0.85274), 예우(0.85217)
19	중국 /Noun	한국(0.99775), 나라(0.98284), 민족(0.96609), 침탈(0.96535), 중국인(0.96314), 외국(0.95834), 기독교(0.94169), 한국인(0.93783), 우리(0.92710), 먹다(0.92254), 일본(0.91586), 외세(0.91354), 역사(0.91088), 단일(0.90868), 들이다(0.90240), 우리나라(0.89025), 자국(0.88562), 잠식(0.88520), 매국(0.88385), 시비(0.88110)
20	국제 /Noun	노사(0.99908), 기반(0.99894), 성과(0.99893), 체계(0.99892), 관리(0.99876), 마련(0.99874), 음주운전(0.99867), 임무(0.99852), 환경부(0.99852), 폐지(0.99851), 소수자(0.99848), 격차(0.99846), 민주(0.99842), 외교통상부(0.99840), 증가(0.99817), 언급(0.99816), 질서(0.99814), 선진(0.99807), 유지(0.99807), 전반(0.99803)

차. ‘저소득’ 키워드 분석결과

저소득층과 관련된 다빈도 형태소는 아래와 같다. 우선, ‘전세자금’, ‘전세’, ‘월세’, ‘주택’ 등 주거와 관련된 형태소들이 분석되었고, ‘한부모 가정’, ‘기초생활수급자’ 등 경제적으로 어려운 계층에 대한 형태소가 다빈도로 나타났으며, ‘국가장학금’과 같은 제도 관련 형태소도 있다.

<표 4-23> '저소득' 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	전세자금/Noun	785	0.6303
2	한부모가정/Noun	441	0.4837
3	은행/Noun	374	0.4827
4	유형/Noun	348	0.4823
5	이자/Noun	296	0.4815
6	장학금/Noun	1,015	0.4135
7	국가장학금/Noun	616	0.4085
8	등록금/Noun	413	0.4060
9	한부모/Noun	288	0.4044
10	이사/Noun	282	0.4043
11	전형/Noun	1,117	0.3617
12	전세/Noun	496	0.3549
13	집다/Verb	397	0.3538
14	직/Noun	320	0.3530
15	주택/Noun	298	0.3527
16	컴퓨터/Noun	295	0.3527
17	기초생활수급자/Noun	577	0.3172
18	월세/Noun	495	0.3164
19	이하/Noun	297	0.3145

다빈도 형태소별 유사도를 통하여 보다 세부적으로 살펴보면, '전세 자금'과 관련해서는 '전세대출', '햇살론', '새희망홀씨' 등의 대출과 관련된 형태소와 '주택기금', '임대아파트' 등과 같은 주거환경에 대한 형태소들이 나타났다. 주택과 관련해서는 '자금대출', '전월세', '국민 임대', '공공임대', '영구임대', '다세대' 등이 있다. 그리고 저소득층 가정 내의 학생과 관련된 '전형', '국가장학금', '등록금' 같은 형태소 등도 나타났다. 그리고 기초생활보장 수급자 선정기준인 '최저생계비', '소득인정액', '건보료' 등과 같은 형태소들도 다수 나타났다.

<표 4-24> '저소득' 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	전세자금 /Noun	전세대출(0.87450), 햇살론(0.81565), 대출(0.80431), 새희망홀씨(0.79184), 신혼부부(0.78348), 전세집(0.78081), 버팀목(0.76925), 영세민(0.76170), 자금대출(0.73879), 자금(0.73831), 주택기금(0.73783), 전세임대(0.73149), 고금리(0.72985), 전세금(0.72525), 생계지원(0.72263), 우리카redit(0.72000), 대금(0.71687), 회생(0.71611), 금융(0.71468), 임대아파트(0.71453)
2	한부모 가정 /Noun	기초수급(0.89354), 기초생활수급자(0.89028), 한부모(0.87762), 차상위계층(0.82999), 한부모가족(0.81397), 기초생활수급(0.80313), 기초생활(0.79169), 기초수급대상자(0.78314), 수급자(0.76021), 자격(0.72582), 수급(0.70331), 혜택(0.70273), 미성년자(0.70194), 성인(0.69543), 모자가정(0.69532), 박탈(0.68948), 자녀(0.68668), 대상자(0.67414), 교육비지원(0.67368), 지원(0.67321)
3	은행 /Noun	은행권(0.87897), 농협(0.87378), 금융(0.87344), 햇살론(0.86659), 이율(0.84380), 신용대출(0.83986), 우리은행(0.83983), 승인(0.82771), 신한은행(0.81921), 은행가(0.81808), 조회(0.81345), 지자체장(0.80419), 국민은행(0.80399), 신용(0.80324), 사금융(0.80291), 새마을금고(0.80113), 보증(0.79997), 담보대출(0.79973), 저축은행(0.79506), 전세금(0.79414)
4	유형 /Noun	수혜(0.89082), 소득분위(0.84336), 근로장학생(0.82990), 이종수혜(0.82378), 드림장학금(0.82171), 장학생(0.81376), 공지(0.81049), 지원금액(0.80816), 국장(0.80348), 우수(0.80132), 한국장학재단(0.79454), 사랑드림(0.78808), 국가장학금(0.77535), 장학(0.77143), 선발(0.77018), 심사(0.76469), 중복(0.76170), 근로장학금(0.75886), 평점(0.73873), 페이지(0.73577)
5	이자 /Noun	대출금(0.91660), 원금(0.90748), 프로(0.87328), 갚다(0.87003), 금리(0.86777), 신한은행(0.85921), 신용대출(0.85538), 상환(0.85109), 캐피탈(0.84589), 저축은행(0.83908), 보증(0.83535), 농협(0.82914), 환하다(0.82882), 사금융(0.82223), 이율(0.81769), 우리은행(0.81765), 적금(0.81669), 전세금(0.81493), 한도(0.81188), 은행권(0.80721)
6	장학금 /Noun	국가장학금(0.87300), 장학재단(0.85535), 드림장학금(0.81536), 장학(0.80153), 우수(0.76894), 한국장학재단(0.76756), 사랑드림(0.75430), 소득분위(0.75369), 이종수혜(0.75265), 등록금(0.74720), 신입생(0.73317), 드림(0.73065), 수혜(0.72787), 근로장학금(0.72730), 학기(0.72385), 유형(0.71956), 성적(0.68490), 분위(0.67738), 학자금대출(0.67310), 국장(0.66720)
7	국가 장학금 /Noun	드림장학금(0.87744), 장학금(0.87300), 소득분위(0.85510), 한국장학재단(0.84749), 사랑드림(0.80808), 분위(0.79489), 신입생(0.79118), 근로장학금(0.78118), 학기(0.77853), 유형(0.77535), 이종수혜(0.77037), 장학재단(0.76911), 우수(0.76588), 수혜(0.76473), 드림(0.75360), 탈락(0.75011), 심사(0.74900), 국장(0.73993), 장학(0.69407), 원클릭(0.68746)

순위	단어	단어유사도(word2vec)
8	등록금 /Noun	기숙사비(0.80531), 학비(0.80496), 수업료(0.79072), 전액(0.75271), 급식비(0.74790), 장학금(0.74720), 학자금대출(0.74397), 입학금(0.74071), 고지서(0.73877), 기숙사(0.73128), 장학(0.72812), 학자금(0.72130), 생활비(0.71638), 장학재단(0.71428), 재단(0.71246), 국가장학금(0.68432), 총당(0.67465), 이중수혜(0.65996), 사립대(0.65749), 사랑드림(0.63417)
9	한부모 /Noun	기초수급(0.88955), 차상위계층(0.88805), 한부모가정(0.87762), 기초수급대상자(0.87472), 기초생활수급(0.84844), 한부모가족(0.83723), 기초생활수급자(0.83309), 기초생활(0.81403), 대상자(0.80799), 가정(0.79097), 자녀(0.75171), 긴급복지지원(0.74172), 해당(0.71158), 모자가정(0.70642), 수급자(0.69534), 보호대상자(0.69183), 여야(0.68448), 저소득(0.68093), 분류(0.67882), 면제(0.67600)
10	이사 /Noun	투룸(0.82502), 전세(0.82073), 계약(0.79201), 전세집(0.79122), 보증금(0.76598), 아파트(0.76440), 전세계약(0.75814), 계약금(0.75474), 계약기간(0.75357), 전세금(0.74844), 당첨(0.73983), 주택공사(0.73617), 세입자(0.72517), 집주인(0.72385), 임대(0.72085), 입주(0.71152), 비우다(0.70854), 우리은행(0.70638), 원룸(0.70061), 재계약(0.69349)
11	전형 /Noun	전형도(0.87826), 특별전형(0.84783), 사회배려(0.81595), 수시(0.81122), 원서(0.80761), 기회균등(0.80720), 일반전형(0.80499), 합격(0.79456), 정시(0.78981), 경쟁률(0.78706), 농어촌(0.77676), 뽑다(0.76012), 모집(0.74977), 가망(0.72621), 고르다(0.72303), 회전(0.71985), 커트라인(0.71723), 특목고(0.71144), 예비(0.70298), 불다(0.69736)
12	전세 /Noun	전세집(0.95007), 보증금(0.93196), 투룸(0.93159), 아파트(0.89926), 빌라(0.88003), 월세(0.87195), 원룸(0.86692), 평수(0.86124), 임대아파트(0.85148), 매매(0.84818), 평(0.83961), 전세금(0.83288), 주택공사(0.82458), 분양(0.82205), 임대(0.82123), 이사(0.82073), 임대주택(0.80735), 자금대출(0.80519), 주공아파트(0.79614), 세입자(0.79529)
13	집다 /Verb	집(0.82826), 월세(0.80783), 식구(0.79945), 원룸(0.75835), 할머니(0.75125), 시댁(0.74477), 무주택(0.74426), 친정(0.73603), 평수(0.72903), 신랑(0.72296), 아버님(0.71594), 외할머니(0.71272), 평(0.70860), 이모(0.70777), 었다(0.70734), 투룸(0.70449), 정이(0.70113), 할아버지(0.69694), 차갑다(0.69671), 요양원(0.69354)
14	직 /Noun	지방직(0.91789), 행정(0.86070), 행정직(0.85793), 직렬(0.85566), 출입국관리(0.85506), 특채(0.84436), 급(0.83612), 충청남도(0.83400), 계리직(0.83182), 세무(0.82374), 강원도(0.81933), 익산(0.81901), 충북(0.81270), 사회복지공무원(0.80913), 기술직(0.80178), 행정고시(0.79890), 소방(0.79483), 임용(0.79421), 채용(0.78754), 컴활(0.78597)

순위	단어	단어유사도(word2vec)
15	주택 /Noun	자금대출(0.90426), 전월세(0.89530), 분양(0.88482), 보유(0.88261), 국민임대(0.87678), 공공임대(0.87571), 임대(0.87455), 자금(0.87010), 빌라(0.86976), 구입(0.85000), 용자(0.84882), 임대주택(0.84308), 영구임대(0.83316), 소유(0.83308), 금융공사(0.83223), 자가(0.83052), 담보대출(0.83038), 임대료(0.82903), 다세대(0.82076), 시세(0.81802)
16	컴퓨터 /Noun	통신비(0.69177), 식권(0.65305), 원클릭(0.64166), 수거(0.61958), 무료(0.60973), 종이(0.60718), 방과후(0.60582), 담임(0.60524), 학교(0.59739), 교육비지원(0.59269), 공예(0.58713), 처가(0.58206), 인터넷(0.57620), 교육비(0.57559), 공문(0.57286), 우유(0.57008), 지원(0.56969), 프로그래밍(0.56953), 프린트(0.56896), 학비(0.55795)
17	기초생활 수급자 /Noun	기초수급(0.90463), 한부모가정(0.89028), 기초생활(0.86440), 기초생활수급(0.85486), 차상위계층(0.84899), 수급자(0.84447), 한부모(0.83309), 한부모가족(0.82774), 기초수급대상자(0.78005), 수급(0.76719), 자격(0.74520), 순위(0.73528), 대상자(0.73384), 탈락(0.72759), 우선돌봄(0.71490), 해당(0.69417), 분류(0.69383), 긴급복지지원(0.68463), 박탈(0.68420), 세대분리(0.68312)
18	월세 /Noun	보증금(0.91978), 원룸(0.89971), 전세(0.87195), 투룸(0.84225), 전세집(0.83030), 평수(0.81089), 집다(0.80783), 집도(0.78445), 짜다(0.78202), 반지하(0.75878), 평(0.75587), 강북(0.74429), 차도(0.73833), 아파트(0.73491), 빌라(0.72761), 천만원(0.72669), 부채(0.72446), 임대아파트(0.72189), 신랑(0.72152), 방(0.70983)
19	이하 /Noun	최저생계비(0.87926), 연소득(0.86564), 건보료(0.84748), 건강보험료(0.83533), 가구(0.81996), 미만(0.81962), 지역가입(0.79391), 액(0.79339), 소득인정액(0.79231), 이내(0.78127), 산정(0.77787), 월소득(0.77537), 공제(0.75914), 소득환산액(0.74680), 보험료(0.74569), 기준(0.74090), 의료보험료(0.74079), 합산(0.73985), 포함(0.73830), 연간(0.73263)

카. ‘교육’ 키워드 분석결과

교육 관련된 형태소 분석결과는 ‘영유아’, ‘아동청소년’과 비슷하게 관련 학과나 진학 및 진로에 관련된 형태소가 다빈도로 나타났다.

<표 4-25> ‘교육’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	평생/Noun	3,215	0.6234
2	특수교육/Noun	2,297	0.6211

순위	단어	빈도(개)	TF-IDF
3	사이버/Noun	1,910	0.4809
4	유아교육/Noun	16,107	0.4276
5	평생교육/Noun	7,354	0.4130
6	유아/Noun	2,955	0.4057
7	전문학교/Noun	2,309	0.4047
8	보육/Noun	1,910	0.4040
9	교사/Noun	4,171	0.3555
10	학위/Noun	4,039	0.3553
11	수업/Noun	3,849	0.3551
12	유치원/Noun	2,784	0.3535
13	대학원/Noun	2,334	0.3529
14	사회복지과/Noun	1,846	0.3522
15	보육교사/Noun	9,435	0.3238
16	학사/Noun	5,594	0.3189
17	편입/Noun	4,487	0.3174
18	진학/Noun	2,283	0.3146
19	면접/Noun	1,884	0.3141

다빈도 형태소별 유사도 분석결과도 ‘학교’, ‘학과’, ‘진학’, ‘진로’ 등의 형태소가 대다수를 차지하였다.

<표 4-26> ‘교육’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	평생 /Noun	정보과학(0.49138), 광운대(0.40221), 지식산업(0.39515), 유아(0.39061), 사회복지사(0.38933), 사이버(0.38731), 평생교육(0.37553), 온누리(0.36977), 패스(0.35544), 학점은행제(0.34780), 가족생활(0.34769), 청소년지도사(0.33626), 원격(0.33399), 미술치료(0.33348), 복지사(0.32833), , 평생교육사(0.32063), 중퇴(0.31632)
2	특수교육 /Noun	특수교사(0.59416), 특수(0.57814), 나사렛대(0.51258), 특교(0.50207), 언어치료(0.49877), 재활학과(0.49495), 유아교육학(0.49247), 초등(0.49173), 유아교육(0.48304), 사범대(0.47767), 사회복지학과(0.46752), 대구대(0.46633), 국어교육(0.44339), 교육학(0.43025), 중등(0.42657), 교대(0.42604), 특수학교(0.41248), 역사교육(0.40984), 나사렛대학교(0.40657), 심리학과(0.39833)

순위	단어	단어유사도(word2vec)
3	사이버 /Noun	사이버교육(0.72688), 원격교육(0.60020), 원격(0.59718), 평생교육(0.59700), 온라인(0.59369), 사이버대학(0.59202), 온라인수업(0.59097), 사이버강의(0.58940), 학점은행(0.58237), 오프라인(0.57424), 인터넷강의(0.55682), 학점은행제(0.55221), 온라인교육(0.51684), 사이버대학교(0.51628), 온라인강의(0.51251), 에듀(0.50371), 부설(0.50215), 시간제(0.50174), 에듀월(0.49863), 수강(0.49759)
4	유아교육 /Noun	유아교육학(0.78326), 유교(0.71564), 사회복지과(0.66561), 아동복지학과(0.62539), 보육학과(0.59747), 전문대(0.58444), 유치원교사(0.57100), 유치원선생님(0.57067), 사회복지학과(0.55879), 한양여대(0.53840), 아동복지(0.52547), 실업계(0.51170), 수원여대(0.50846), 내신(0.50714), 유치원(0.50638), 경인여대(0.50529), 내신등급(0.50146), 전문대학교(0.49685), 장안대(0.48601), 동남보건대(0.48447)
5	평생교육 /Noun	원격교육(0.64850), 사이버교육(0.63142), 학점은행제(0.60761), 사이버(0.59700), 학점은행(0.57495), 사이버대학교(0.55449), 교육기관(0.52737), 사이버대학(0.52606), 시간제(0.51531), 사이버대(0.49451), 전산(0.49282), 디지털대학교(0.48780), 온라인수업(0.48422), 인터넷강의(0.47462), 오프라인(0.46864), 사이트(0.46808), 에듀월(0.46621), 사이버강의(0.46573), 온라인(0.45967), 온라인강의(0.45055)
6	유아 /Noun	유아(0.57483), 영유아(0.57390), 특수(0.52919), 아동건강(0.52350), 영아(0.51920), 아동(0.50038), 아동복지학(0.49749), 교육공학(0.47567), 장애아(0.47496), 아동복지(0.45269), 보육학과(0.45117), 교육철학(0.44479), 초등(0.44406), 유아교육(0.44231), 아동문학(0.43576), 백석예술대학교(0.43575), 몬테소리(0.43184), 보육(0.42764), 생활지도(0.42209), 아동발달(0.41716)
7	전문학교 /Noun	전문대학교(0.68675), 한양(0.68253), 한서(0.66289), 전문대(0.58679), 직업전문학교(0.52590), 고려전문학교(0.49840), 서강(0.48592), 유아교육학(0.48305), 전문대학(0.48065), 아동복지학과(0.47039), 사이버대학교(0.46164), 배화여대(0.44763), 유아교육(0.44163), 중앙대(0.43160), 대학교(0.43003), 디지털대학교(0.41462), 대학(0.40968), 사회복지학과(0.40872), 취업률(0.39850), 광운대(0.38744)
8	보육 /Noun	아동복지(0.62822), 보육학과(0.59296), 보육학(0.58025), 영아(0.58008), 영유아(0.52233), 생활지도(0.52206), 보육교사(0.51528), 아동(0.48601), 사회복지(0.48002), 어린이집(0.46600), 복지과(0.46513), 보육자격증(0.46061), 유아교육(0.45940), 보육시설(0.45666), 사회복지과(0.45360), 사복(0.43611), 결식(0.43286), 유치원(0.42974), 유아(0.42764), 아동복지론(0.42419)
9	교사 /Noun	교사자격증(0.68905), 특수교사(0.55537), 중등(0.55373), 선생님(0.54233), 병설(0.48103), 유치원교사(0.47834), 교원자격증(0.47403), 어린이집교사(0.46186), 보육교사(0.44851), 교육대학(0.43762), 초등(0.43484), 임용고시(0.43297), 원장(0.42275), 유치원선생님(0.42211), 병설유치원(0.41659), 교직이수(0.41503), 유아교육법(0.40402), 유치원(0.40363), 유치(0.39608), 영어선생님(0.38887)

순위	단어	단어유사도(word2vec)
10	학위 /Noun	졸업장(0.65586), 학점(0.57011), 학점은행제(0.56926), 학사(0.54134), 사회복지학(0.53737), 학점이수(0.52197), 작위(0.51753), 위증(0.51504), 자격증(0.51377), 학점은행(0.50439), 전문(0.50276), 편집(0.49295), 졸업(0.48826), 아동학(0.48683), 급(0.48172), 대졸(0.46505), 최종학력(0.46045), 교원자격증(0.44315), 이수(0.44136), 박사학위(0.44032)
11	수업 /Noun	강의(0.76740), 온라인강의(0.65449), 인터넷강의(0.61097), 사이버강의(0.58762), 인강(0.58720), 온라인수업(0.57613), 오프라인(0.54613), 학점이수(0.52880), 온라인(0.51789), 과목(0.51194), 강좌(0.50054), 교양과목(0.48383), 실습(0.46995), 수강(0.45339), 학점(0.44994), 학점은행제(0.43568), 수강신청(0.43357), 필수과목(0.42351), 커리큘럼(0.41631), 사이버교육(0.41338)
12	유치원 /Noun	어린이집(0.78378), 병설유치원(0.72759), 국공립(0.64620), 병설(0.63417), 유치원선생님(0.60854), 어린이집선생님(0.60760), 어린이집교사(0.60259), 유치원교사(0.59859), 원장(0.55883), 사립(0.54700), 공립(0.52260), 국공립어린이집(0.50749), 유아교육(0.50638), 유치(0.49134), 보육원(0.49072), 유아교육학(0.48567), 임용고시(0.47743), 아동복지학과(0.47315), 종일반(0.46222), 임용고사(0.46158)
13	대학원 /Noun	교육대학(0.78089), 석사(0.61974), 석사과정(0.60999), 상담심리(0.57184), 상담심리학(0.55230), 교육학(0.53008), 심리학과(0.52159), 심리학(0.51690), 교원자격증(0.50592), 대학(0.50380), 사이버대학(0.50148), 교적이수(0.49681), 사회복지학(0.48824), 사이버대학교(0.48448), 대학교(0.46867), 사이버대(0.44318), 방통대(0.44290), 편입학(0.42843), 년제(0.42066), 학사(0.41802)
14	사회 복지과 /Noun	사회복지학과(0.68358), 유아교육(0.66561), 전문대(0.60377), 복지과(0.59437), 아동복지학과(0.56726), 아동복지(0.56291), 유교(0.54429), 사복(0.52570), 보육학과(0.51485), 유아교육학(0.49747), 복지학과(0.49600), 전문대학교(0.49229), 동남보건대(0.48854), 대학(0.48471), 사회복지(0.48453), 아동복지학(0.46792), 경인여대(0.46558), 신흥대(0.46165), 복수전공(0.45499), 영진전문대(0.45494)
15	보육교사 /Noun	사회복지사(0.72080), 평생교육사(0.62708), 어린이집(0.55194), 취득(0.52210), 자격증(0.51824), 보육(0.51528), 급(0.51441), 건강가정사(0.48834), 온라인(0.48748), 학점은행제(0.48709), 사회복지(0.46469), 어린이집교사(0.45880), 교사(0.44851), 보육자격증(0.44542), 유치원(0.44528), 방과후지도사(0.44386), 요양보호사(0.43329), 주부(0.42644), 사이버(0.42598), 유정(0.42045)
16	학사 /Noun	학점은행제(0.57709), 학점(0.56146), 학위(0.54134), 공학사(0.50461), 학점은행(0.49526), 석사(0.49270), 학점이수(0.46237), 사회복지학(0.46156), 자격증(0.45790), 방송대(0.43723), 급(0.43630), 전공(0.43597), 년제(0.43331), 졸업장(0.43018), 독학사(0.42976), 사이버대학(0.42158), 대학원(0.41802), 과목(0.41420), 이수(0.40534)

순위	단어	단어유사도(word2vec)
17	편입 /Noun	일반편입(0.70009), 편입학(0.66709), 방통대(0.57028), 입학(0.53175), 학점은행제(0.50006), 학위(0.49295), 전과(0.48876), 사이버대(0.47922), 방송대(0.47328), 진학(0.46066), 조기졸업(0.46039), 교직이수(0.45583), 년제(0.44973), 방송통신대학(0.44417), 사이버대학(0.44406), 복수전공(0.43860), 사이버대학교(0.43677), 졸업(0.42825), 방송통신대(0.41840), 자퇴(0.41643)
18	진학 /Noun	입학(0.58899), 졸업(0.53653), 진학문(0.48555), 편입(0.46066), 중퇴(0.42269), 졸업생(0.41591), 교직이수(0.41422), 사범(0.41352), 가다(0.41346), 포기(0.39146), 자퇴(0.38786), 취업(0.38708), 편입학(0.38030), 상담심리(0.37972), 갈거다(0.37861), 갈다(0.37516), 공부(0.37199), 다니다(0.37006)
19	면접 /Noun	논술(0.49069), 수능(0.47563), 고사(0.46416), 적성검사(0.44601), 수시(0.42948), 학생부(0.41819), 시험(0.41802), 입사(0.40993), 수능점수(0.40741), 임용고시(0.39922), 뽑다(0.39446), 합격(0.39343), 실기(0.39113), 필기(0.38819), 검정고시(0.38770), 전형(0.38657), 필기시험(0.38474), 구술(0.37505), 정시(0.37200), 경쟁률(0.36744)

타. ‘고용’ 키워드 분석결과

고용관련 다빈도 형태소는 ‘일용직’, ‘공사’ 등이 높은 빈도로 나타났고, ‘임금’과 ‘퇴직금’, ‘실업급여’ 등도 적지 않은 빈도로 나타났다. 또한, ‘산재보험’, ‘고용보험’, ‘국민연금’과 같은 고용관련 복지제도와 관련된 형태소들도 다빈도로 나타났다.

<표 4-27> ‘고용’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	일용직/Noun	1,102	0.6244
2	공사/Noun	1,073	0.6242
3	업체/Noun	1,038	0.6239
4	임금/Noun	1,031	0.6238
5	사업장/Noun	1,021	0.6237
6	퇴직금/Noun	995	0.6235
7	현장/Noun	796	0.6219
8	작성/Noun	736	0.6214

순위	단어	빈도(개)	TF-IDF
9	산재보험/Noun	2,142	0.4910
10	실업급여/Noun	1,902	0.4894
11	퇴사/Noun	1,314	0.4856
12	사장/Noun	1,298	0.4855
13	고용보험/Noun	4,036	0.4228
14	근로자/Noun	1,399	0.4084
15	근로복지공단/Noun	1,096	0.4068
16	보험료/Noun	941	0.4059
17	국민연금/Noun	938	0.4059
18	납부/Noun	882	0.4056
19	근로/Noun	852	0.4054
20	공단/Noun	812	0.4052

다빈도 형태소별 유사도를 살펴보면 ‘일용직’의 경우, ‘아르바이트’, ‘건설회사’, ‘노무자’, ‘인부’, ‘노임’ 등의 건설 현장에서의 일용근로와 관련된 형태소들이 관련 토픽으로 나타났다.

<표 4-28> ‘고용’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	일용직 /Noun	일용(0.76440), 일용직근로자(0.75588), 일용근로자(0.72449), 일당(0.71460), 아르바이트생(0.69842), 신다(0.67650), 현장(0.67499), 건설현장(0.66732), 산재가입(0.66513), 일일(0.65838), 건설회사(0.64699), 노무자(0.63377), 상용(0.61571), 근로자(0.61518), 공사현장(0.61270), 장인어른(0.60689), 노임(0.58536), 인부(0.58441), 공사(0.58201), 건설업(0.57983)
2	공사 /Noun	하도급(0.80681), 건설(0.78468), 창호(0.76601), 노임(0.75392), 도급(0.75319), 외주(0.74771), 발주(0.73283), 건축(0.73102), 건설업(0.72992), 시공(0.72265), 공사비(0.71839), 수주(0.71333), 건물(0.69899), 일팔(0.69605), 공종(0.69298), 준공(0.69045), 공사현장(0.69011), 면허(0.68948), 하청(0.67728), 업체(0.67181)
3	업체 /Noun	하청(0.81493), 하청업체(0.78540), 하도급(0.75527), 원청(0.75403), 도급(0.74006), 건설회사(0.68679), 아웃소싱(0.68549), 건설(0.67884), 공사(0.67181), 발주(0.66013), 면허(0.65267), 용역(0.65044), 건설업(0.64669), 창호(0.63922), 급사(0.63400), 계약(0.62331), 용역업체(0.62323), 소속(0.61823), 현장(0.61746), 인테리어(0.61152)

순위	단어	단어유사도(word2vec)
4	임금 /Noun	보수(0.62849), 인건비(0.59687), 급여(0.59285), 노임(0.58959), 최저임금(0.58634), 퇴직금(0.58500), 일당(0.58437), 체불(0.57336), 산정(0.57138), 상여금(0.56979), 급료(0.54539), 매월(0.53832), 근로(0.53387), 합(0.53243), 총액(0.53040), 미지(0.52422), 일일(0.52268), 인원(0.52204), 최저(0.52106), 월급(0.51924)
5	사업장 /Noun	사업자(0.72241), 상시근로자(0.71952), 사업자등록(0.67802), 상시(0.67709), 일용직근로자(0.63843), 미만(0.63769), 산재가입(0.63657), 개인사업자(0.63029), 사업주(0.61821), 관리번호(0.61749), 근로자(0.61027), 폐업(0.60972), 일용근로자(0.60401), 대표자(0.59965), 사업(0.57912), 번호(0.57194), 사업자등록증(0.57126), 대표이사(0.57090), 가입(0.55700), 피고용자(0.55405)
6	퇴직금 /Noun	월급(0.68771), 퇴직(0.66592), 정산(0.66042), 급여(0.65722), 연차(0.62860), 그동안(0.61293), 실업급여(0.60649), 임금(0.58500), 미지급(0.57952), 상여금(0.56631), 퇴사(0.56620), 통보(0.55782), 육아휴직급여(0.55238), 만일(0.55044), 육아휴직(0.54984), 연봉(0.54091), 원장(0.53662), 보너스(0.53476), 출산휴가(0.53390), 부당(0.51806)
7	현장 /Noun	공사현장(0.80241), 하청업체(0.79342), 본사(0.78661), 건설현장(0.78239), 일용직근로자(0.77850), 노무자(0.75742), 파트(0.72838), 일용(0.71537), 소장(0.69078), 사무실(0.68880), 하청(0.68609), 건설회사(0.68603), 인부(0.68020), 작업(0.67524), 일용직(0.67499), 공사(0.66643), 일용근로자(0.64300), 도급(0.63623), 개시(0.63443), 미장(0.63413)
8	작성 /Noun	제출(0.82837), 근로계약서(0.77115), 계약서(0.75035), 공지(0.69888), 복사(0.69829), 싸인(0.67816), 고용계약서(0.67590), 신고서(0.66499), 접수(0.66112), 확인서(0.65081), 신청서(0.63979), 서면(0.63948), 자필(0.63830), 인수인계(0.63060), 서명(0.62150), 수정(0.61742), 요청(0.61654), 서류(0.59780), 확인(0.59129), 급여통장(0.58554)
9	산재보험 /Noun	산재(0.73693), 산재보험료(0.65499), 산재가입(0.64483), 산재처리(0.58474), 근재보험(0.56574), 일용직(0.55965), 소급(0.55102), 일용직근로자(0.55098), 산업재해(0.54569), 일용근로자(0.53873), 가입(0.52979), 사업주(0.52842), 하수(0.52104), 보험(0.50171), 국민연금(0.49867), 사업장(0.49066), 일괄(0.48235), 보험처리(0.48218), 장인어른(0.47806), 납부(0.47678)
10	실업급여 /Noun	조기(0.66953), 재취업(0.66648), 취업수당(0.65876), 퇴사(0.65374), 권고사직(0.65181), 퇴직(0.64094), 환급(0.64024), 그만두다(0.62942), 수급(0.62923), 실직(0.62663), 고용보험(0.62219), 퇴직금(0.60649), 이직(0.59610), 육아휴직급여(0.58698), 직장(0.56832), 만약(0.56436), 구직활동(0.56312), 기초수급대상자(0.55994), 출산휴가(0.55714), 육아휴직(0.55577)
11	퇴사 /Noun	퇴직(0.87041), 사직(0.82341), 권고사직(0.82274), 권고(0.74906), 그만두다(0.72015), 자진(0.71073), 개인사정(0.70147), 통보(0.69522), 전직(0.69388), 이직(0.68542), 복직(0.68341), 폐업(0.67039), 실업급여(0.65374), 계약해지(0.64918), 상실(0.64170), 사정(0.63817), 입사(0.63581), 인수인계(0.63106), 해고(0.63048), 복귀(0.60870)

순위	단어	단어유사도(word2vec)
12	사장 /Noun	달라(0.74939), 해달라다(0.68323), 미루다(0.66797), 가게(0.64774), 업주(0.64263), 직원(0.63462), 약속(0.62661), 다하다(0.62011), 인수인계(0.61794), 몰다(0.60255), 그냥(0.60206), 이야기(0.60126), 소장(0.60115), 형님(0.60094), 낱다(0.59996), 그동안(0.59823), 싸인(0.59479), 고용주(0.59003), 자기(0.58813), 사무실(0.58489)
13	고용보험 /Noun	고용보험료(0.66076), 소급(0.66015), 국민연금(0.64210), 실업급여(0.62219), 지역가입(0.59571), 의료보험(0.57839), 환급(0.57218), 가입(0.57062), 납입(0.57044), 이력(0.56726), 건강보험(0.55835), 연말정산(0.55093), 자동(0.55031), 보험료(0.53438), 보험(0.53172), 누락(0.52949), 재직(0.52946), 정직(0.51882), 납부(0.51175), 조회(0.50662)
14	근로자 /Noun	상시근로자(0.66250), 사업주(0.65627), 상시(0.63848), 상용(0.62459), 근로(0.62452), 일용직(0.61518), 근로계약(0.61264), 사업장(0.61027), 미만(0.60881), 외국인(0.60608), 자의(0.58939), 고용주(0.58406), 사용자(0.57916), 일용근로자(0.57766), 의무(0.56781), 정규직(0.56581), 비정규직(0.54192), 일용(0.53832), 일일(0.52283), 고용기간(0.51789)
15	근로 복지공단 /Noun	세무서(0.75873), 공단(0.75257), 근로복지(0.74881), 접수(0.72554), 담당자(0.71918), 공문(0.70071), 안내(0.70029), 건강보험공단(0.69919), 홈페이지(0.69779), 팩스(0.69653), 안내문(0.69614), 청구서(0.69222), 우편(0.69197), 노동청(0.68418), 노동부(0.67633), 신청접수(0.67119), 고지서(0.66511), 승인(0.66427), 관리번호(0.66305), 날아오다(0.66013)
16	보험료 /Noun	고용보험료(0.84228), 납부(0.79134), 산재보험료(0.72842), 건강보험료(0.72807), 연말정산(0.69946), 징수(0.68314), 소급(0.67699), 고지서(0.67159), 원천징수(0.66790), 연금보험료(0.66721), 미납(0.66673), 세금(0.65589), 영수증(0.64799), 보험금(0.64650), 금액(0.64110), 소득신고(0.63359), 지역가입(0.62467), 국세청(0.62365), 차액(0.62154), 원천(0.61997)
17	국민연금 /Noun	건강보험(0.88732), 의료보험(0.82615), 지역가입(0.78537), 건강보험료(0.74846), 연금(0.72832), 납입(0.71615), 연금보험(0.71361), 가입자(0.68287), 소득세(0.67649), 주민세(0.67502), 건강(0.67403), 국민건강보험(0.66700), 직장가입자(0.65751), 고용보험료(0.64212), 고용보험(0.64210), 원천징수(0.63678), 미납(0.63055), 체납(0.62944), 노인장기요양보험(0.62044), 장기요양(0.61834)
18	납부 /Noun	고용보험료(0.81417), 미납(0.81128), 보험료(0.79134), 누락(0.75324), 징수(0.75208), 건강보험료(0.73740), 납입(0.72373), 체납(0.72331), 고지서(0.71417), 연말정산(0.71347), 산재보험료(0.70554), 원천징수(0.70209), 정산(0.69546), 개산(0.69093), 반환(0.68422), 소급(0.68192), 신고(0.67800), 부과(0.66901), 공제(0.66541), 납(0.65114)
19	근로 /Noun	일용(0.63744), 근로자(0.62452), 근로계약(0.61425), 계약서(0.60974), 취업규칙(0.59348), 일수(0.57118), 자부담분(0.56768), 신고서(0.56628), 미만(0.56300), 상시(0.56208), 확인서(0.56067), 근무(0.55841), 계약직(0.55469), 작성(0.55438), 고용계약서(0.54728), 근로계약서(0.54699), 규약(0.53955), 일용근로자(0.53873), 정규직(0.53546), 체크(0.53486)

순위	단어	단어유사도(word2vec)
20	공단 /Noun	근로복지(0.85408), 근로복지공단(0.75257), 관리공단(0.67268), 안내(0.66232), 안내문(0.65464), 공문(0.64498), 홈페이지(0.62699), 청구서(0.62469), 우편(0.61757), 국민연금공단(0.61688), 담당자(0.61395), 세무서(0.60362), 건강보험공단(0.60020), 고지서(0.59892), 우편물(0.59003), 입하(0.58477), 고용부담금(0.57751), 구청(0.57377), 센터(0.57356), 내역(0.57263)

파. ‘주거’ 키워드 분석결과

주거관련 다빈도 형태소 분석결과로 ‘주거급여’, ‘생계급여’, ‘기초수급’ 등 기초생활보장제도와 관련된 형태소들이 높은 빈도를 가지는 것으로 나타났다. 뿐만 아니라 관련 규정들도 다빈도로 나타났으며, ‘보증금’, ‘월세’ 등과 같은 주거비용과 관련된 형태소들도 있다.

<표 4-29> ‘주거’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	주거급여/Noun	248	0.6249
2	지구/Noun	148	0.6210
3	생계급여/Noun	148	0.6210
4	건물/Noun	141	0.6207
5	시행/Noun	136	0.6205
6	건축/Noun	128	0.6202
7	개정/Noun	126	0.6201
8	권리/Noun	157	0.4819
9	규정/Noun	149	0.4816
10	중심/Noun	143	0.4815
11	자연/Noun	133	0.4812
12	도시/Noun	258	0.4074
13	가구/Noun	254	0.4073
14	건설/Noun	146	0.4045
15	주택/Noun	348	0.3572
16	거주/Noun	202	0.3540
17	보증금/Noun	131	0.3524

순위	단어	빈도(개)	TF-IDF
18	기초수급/Noun	125	0.3523
19	기초생활수급자/Noun	241	0.3164
20	월세/Noun	200	0.3156

각 형태소별로 유사도를 분석하면 다음과 같다. 먼저 ‘주거급여’와 관련해서는 기초생활보장제도 및 세부 급여서비스인 ‘생계급여’, ‘교육급여’ 등의 형태소들이 나타났고, ‘취업성공패키지’, ‘장애수당’ 같은 복지서비스 등도 있고, 이와 함께 수급자격과 관련된 형태소들도 나타났다. 기초생활수급자와 관련된 형태소로는 ‘한부모가정’, ‘생계비’, ‘조건부수급자’, ‘임대아파트’ 등이 나타났다. 한편, 보증금과 관련된 형태소는 ‘전세’, ‘월세’ 등 계약형태와 ‘대출’, ‘목돈’ 등과 같은 비용, ‘임대아파트’ 등과 같은 주거형태 등이 나타났다.

<표 4-30> ‘주거’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	주거급여 /Noun	생계급여(0.99139), 주거비(0.96388), 수급(0.95493), 생계비(0.95109), 기초수급(0.93744), 취업성공패키지(0.93022), 기초생활(0.92648), 장애수당(0.92177), 매달(0.92061), 수급비(0.91631), 수급자(0.91558), 입금(0.91018), 교육급여(0.90651), 차감(0.90539), 신청(0.90529), 동사무소(0.90431), 자격(0.90411), 생계(0.89882), 기초생활수급자(0.89573), 치료비(0.89239)
2	지구 /Noun	건설(0.91195), 현상(0.90768), 운동(0.90635), 도심(0.90375), 도시(0.90326), 공장(0.90212), 세계(0.89829), 성장(0.89666), 문화(0.89388), 첨단(0.89191), 공원(0.89051), 댐(0.88670), 산업(0.88511), 타운(0.88455), 영화관(0.88194), 소비(0.88057), 경제(0.88016), 문화유산(0.87868), 국토(0.87465), 택지(0.87333)
3	생계급여 /Noun	주거급여(0.99139), 수급(0.96493), 주거비(0.95034), 생계비(0.94473), 장애수당(0.94431), 취업성공패키지(0.94190), 차감(0.94093), 수급자(0.93703), 매달(0.93451), 기초생활(0.92912), 기초수급(0.92631), 입금(0.91789), 교육급여(0.91695), 자격(0.90733), 수급비(0.90545), 계산(0.90446), 치료비(0.90108), 생계(0.89723), 신청(0.89655), 조건부(0.89600)

순위	단어	단어유사도(word2vec)
4	건물 /Noun	층(0.97252), 아파트(0.89857), 상가(0.86901), 개조(0.86535), 공공임대(0.86394), 분양(0.85984), 제곱미터(0.85523), 다세대(0.85351), 면적(0.85007), 주택(0.85001), 유닛(0.82796), 평정(0.82143), 일반(0.81492), 단독주택(0.81156), 주차장(0.80781), 음식점(0.80378), 영업(0.79917), 부지(0.79324), 연립(0.79229), 평(0.79186)
5	시행 /Noun	조정(0.93023), 실시(0.92993), 보건복지부장관(0.92641), 의거(0.91904), 호(0.91863), 규정(0.91694), 개정(0.91576), 군수(0.91540), 신설(0.90397), 구청장(0.89856), 범위(0.89509), 공무원(0.89218), 담당(0.88981), 일부(0.88718), 발부(0.88672), 거치다(0.88543), 위원회(0.88539), 규칙(0.88396), 평가(0.87182), 의하다(0.87136)
6	건축 /Noun	건축물(0.95449), 녹지(0.93821), 주거지역(0.92805), 지역(0.92796), 중일반(0.91138), 연계(0.90205), 단지(0.87873), 게임(0.86189), 지상(0.86060), 상업(0.85915), 지목(0.85729), 인쇄(0.85218), 다양하다(0.85161), 특성(0.84905), 택지(0.84797), 평생교육(0.84400), 주유소(0.84322), 자연(0.84235), 관광(0.84222), 스마트(0.84179)
7	개정 /Noun	규정(0.98609), 의하다(0.98369), 법률(0.96716), 군수(0.96281), 규칙(0.96052), 항의(0.95750), 법령(0.94784), 범위(0.94368), 신설(0.94214), 구청장(0.94212), 결정(0.94174), 법(0.94058), 보건복지부장관(0.93908), 보장기관(0.93831), 대통령령(0.93625), 특례법(0.93362), 헌법(0.93236), 개정안(0.93232)
8	권리 /Noun	가치다(0.97733), 국민(0.96339), 자유(0.91816), 타인(0.91053), 대리인(0.89254), 필요(0.88849), 법조인(0.88396), 제한(0.87994), 존엄성(0.86769), 정당(0.86523), 주권(0.86310), 지위(0.86070), 정부(0.85986), 동의(0.85918), 동등(0.85897), 공공복리(0.85826), 필요하다(0.85374), 어떠하다(0.84805), 행위(0.84573), 법정(0.84297)
9	규정 /Noun	개정(0.98609), 의하다(0.98174), 군수(0.96831), 재판관(0.96090), 규칙(0.96024), 보건복지부장관(0.95754), 신설(0.95609), 법률(0.95402), 범위(0.95330), 헌법(0.95074), 구청장(0.94998), 의무(0.94104), 의결(0.94068), 배상(0.93999), 구금(0.93896), 개정안(0.93653), 따르다(0.93576), 특례법(0.93406), 대통령령(0.93379), 단서(0.93361)
10	중심 /Noun	기술(0.97220), 도시(0.96809), 발전(0.96622), 개발(0.96450), 자원(0.96217), 발달(0.95344), 유비쿼터스(0.94743), 국토(0.94636), 문화(0.94631), 성장(0.94349), 제품(0.94056), 영향(0.94049), 변화(0.93974), 새롭다(0.93607), 생태(0.93053), 경제(0.93022), 다양하다(0.92976), 특성(0.92854), 인프라(0.92786), 배경(0.92760)
11	자연 /Noun	과학(0.96169), 관광(0.96092), 소비자(0.95334), 인재(0.94742), 개론(0.94327), 디자인(0.94068), 전망(0.93978), 미래(0.93925), 경영(0.93468), 대구(0.93425), 식생활(0.93407), 녹지(0.93074), 도심(0.92927), 재료(0.92756), 시스템(0.92606), 산업(0.92559), 특성(0.92514), 컴퓨터(0.92508), 설계(0.92443), 발달(0.92419)
12	도시 /Noun	중심(0.96809), 문화(0.95940), 기술(0.95935), 유비쿼터스(0.95782), 발전(0.95433), 제품(0.95198), 국토(0.95105), 성장(0.95067), 개발(0.94932), 발달(0.94658), 에너지(0.94498), 국제(0.94386), 회계기준(0.94102), 기반(0.94100), 자원(0.93773), 행정(0.93710), 녹색(0.93561), 산업(0.93388), 조성(0.93329), 추진(0.93303)

순위	단어	단어유사도(word2vec)
13	가구 /Noun	포함(0.94317), 세대(0.92366), 장애(0.92230), 특례(0.92094), 수당(0.91935), 차량(0.91836), 소득인정액(0.91257), 전기요금(0.90548), 등록(0.90532), 감면(0.90478), 기준(0.89717), 미만(0.89645), 최저생계비(0.89611), 소득(0.89021), 연금(0.88579), 대상(0.88431), 대상자(0.88333), 별도(0.87859), 자녀(0.87356), 의료비(0.87111)
14	건설 /Noun	공장(0.96462), 교통(0.95609), 추진(0.94905), 에너지(0.94830), 운동(0.94821), 첨단(0.94811), 시민(0.94566), 조성(0.94163), 도로(0.93742), 육성(0.93722), 수요(0.93665), 댐(0.93665), 공원(0.93605), 청주(0.93594), 정비(0.93413), 기업(0.93350), 국제(0.93144), 행정(0.93128), 체육(0.92936), 환경보호(0.92859)
15	주택 /Noun	용도(0.89131), 종(0.88823), 개조(0.88014), 건물(0.85001), 매입(0.84970), 일반(0.83862), 숙소(0.83845), 분양(0.82939), 다세대(0.80640), 근린생활(0.80001), 제곱미터(0.79745), 연면적(0.79045), 공공임대(0.78700), 임대주택(0.77910), 저소득층(0.77835), 상가(0.77420), 임대(0.77259), 허가(0.77006), 연립(0.76974), 단독주택(0.76937)
16	거주 /Noun	임대(0.90170), 미혼(0.90036), 세대(0.89709), 개인회생(0.89260), 입원(0.88989), 신혼부부(0.88986), 주민등록(0.88124), 자녀(0.88040), 단돌(0.87959), 무상(0.87909), 평정(0.87128), 명의(0.86517), 남동생(0.86448), 평(0.86425), 급(0.86055), 예정(0.85790), 전세(0.85470), 맞벌이(0.84798), 빌라(0.84531), 지적장애(0.84296)
17	보증금 /Noun	전세(0.97989), 월세(0.96959), 대출(0.96262), 임대아파트(0.95021), 월급(0.94744), 원금(0.93607), 목돈(0.93328), 어머니(0.93265), 병원비(0.92959), 계약(0.92906), 수입(0.92211), 할부(0.92084), 전세금(0.92020), 아드님(0.91951), 주택청약(0.91946), 카드(0.91717), 통장(0.91589), 밀리다(0.91362), 신랑(0.91327), 수급비(0.91040)
18	기초수급 /Noun	기초생활수급자(0.98502), 기초생활(0.96457), 수급(0.96335), 생계비(0.96110), 생계(0.95500), 한부모가정(0.94701), 인가(0.94582), 주거비(0.94058), 무주택(0.94005), 주거급여(0.93744), 임대아파트(0.93165), 생계급여(0.92631), 조건부수급자(0.92630), 세대분리(0.91900), 조건부(0.91835), 지원비(0.91766), 신청(0.91662), 탈락(0.91581), 자격(0.91526), 동사무소(0.91461)
19	기초생활 수급자 /Noun	기초수급(0.98502), 한부모가정(0.97354), 기초생활(0.97233), 생계(0.95454), 수급(0.94406), 생계비(0.93443), 자격(0.93259), 무주택(0.93052), 조건부수급자(0.92613), 기초생활수급(0.91845), 인가(0.91004), 조건부(0.90683), 생계지원(0.90366), 조건(0.90046), 지원비(0.89754), 주거급여(0.89573), 신청(0.89560), 수급자(0.89461), 세대분리(0.89169), 임대아파트(0.88920)
20	월세 /Noun	전세(0.97791), 보증금(0.96959), 임대아파트(0.95661), 대출(0.93266), 생활비(0.92205), 원룸(0.92186), 막내(0.92077), 목돈(0.92013), 어머니(0.91078), 병원비(0.90862), 전세자금(0.90200), 이사(0.90173), 밀리다(0.89927), 신랑(0.89674), 명의(0.89341), 직장(0.89195), 동생(0.89129), 아드님(0.88921), 월급(0.88898), 계약(0.88607)

하. ‘건강’ 키워드 분석결과

건강과 관련된 형태소들은 복지서비스와 관련하여서는 ‘건강보험’, ‘국민연금’, ‘산재’ 등이 다빈도로 분석되었다. 그 이외에 ‘건강검진’, ‘질병’, ‘진단’, ‘입원’, ‘검사’ 등이 다빈도 형태소로 나타났다.

<표 4-31> ‘건강’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	건강하다/Adjective	1,671	0.6251
2	건강검진/Noun	888	0.6205
3	건강보험/Noun	1,785	0.4853
4	질병/Noun	1,214	0.4827
5	암/Noun	1,186	0.4825
6	진단/Noun	1,245	0.4056
7	보험료/Noun	1,164	0.4053
8	건강가정사/Noun	1,046	0.4048
9	국민연금/Noun	935	0.4044
10	입원/Noun	1,580	0.3548
11	검사/Noun	1,096	0.3532
12	산재처리/Noun	907	0.3525
13	산재/Noun	1,864	0.3172
14	처리/Noun	1,157	0.3151
15	보육교사/Noun	1,014	0.3147

건강과 관련된 복지서비스를 중심으로 살펴보면, ‘건강보험’과 관련하여서는 ‘보험료’, ‘의료보험’, ‘국민연금’, ‘고용보험’, ‘산재보험’ 등의 형태소가 유사도가 높은 것으로 나타났고, ‘국민연금’도 비슷하였다. 한편, ‘산재’와 관련해서는 ‘산재승인’, ‘보상’, ‘보험처리’, ‘산업재해’ 등과 같은 형태소들이 다수 나타났고, ‘권고사직’, ‘해고’, ‘퇴사’ 등과 같은 부정적인 형태소도 있다.

<표 4-32> '건강' 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	건강하다 /Adjective	행복하다(0.57354), 몸(0.55239), 약하다(0.55227), 건강(0.53581), 연세(0.53102), 신체(0.50533), 평소(0.50496), 허약(0.50302), 체력(0.49451), 앓다(0.48319), 몸무게(0.47990), 젊다(0.47793), 밝다(0.47370), 년생(0.46542), 아주(0.46197), 기력(0.45941), 소홀하다(0.45880), 새해(0.45680), 늙다(0.45403), 평상시(0.45356)
2	건강검진 /Noun	검진(0.88069), 검사(0.74730), 재검(0.66790), 암검진(0.58310), 내시경(0.57364), 판정(0.56864), 정기(0.55825), 유방암(0.55048), 난청(0.54974), 예방접종(0.54602), 조직검사(0.54255), 정밀검사(0.54076), 진단(0.52762), 영유아건강검진(0.51912), 고혈압(0.51593), 진료(0.51328), 확진(0.51004), 결핵(0.50771), 결절(0.50613), 진찰(0.49060)
3	건강보험 /Noun	의료보험(0.73613), 가입자(0.71117), 지역가입(0.66964), 국민연금(0.66054), 보험(0.62891), 실손보험(0.61619), 건강보험공단(0.61406), 납부(0.59956), 직장가입자(0.59350), 실비보험(0.59142), 연금보험(0.58504), 고용보험(0.57898), 보험료(0.57760), 산재보험(0.57473), 의료보험료(0.57182), 진료비(0.56223), 산재보험료(0.55941), 꼬박(0.55498), 국민건강보험(0.55458), 건보료(0.54988)
4	질병 /Noun	질환(0.66839), 병(0.64594), 뇌혈관(0.63363), 입원치료(0.60821), 노인성(0.60799), 만성(0.58812), 심장병(0.58276), 상해(0.58233), 악관(0.58113), 동맥경화(0.58084), 발병(0.58031), 입원비(0.57230), 간염(0.56936), 재해(0.56810), 대질(0.56549), 성인병(0.56525), 통원(0.56179), 감염(0.55738), 실신(0.55649), 간병(0.55380)
5	암 /Noun	갑상선암(0.75814), 위암(0.70231), 간암(0.68873), 임파선(0.68728), 췌장암(0.67495), 말기(0.66246), 상피(0.66002), 후두(0.65195), 피부암(0.64849), 갑상선(0.64557), 유방암(0.63604), 폐암(0.63495), 전이(0.63328), 심장병(0.63048), 진단(0.63048), 대장암(0.62782), 항암치료(0.62737), 뇌출중(0.60558), 자궁경부암(0.59894), 투병(0.58131)
6	진단 /Noun	판명(0.71238), 성형수술(0.69553), 갑상선암(0.68454), 유방암(0.67159), 급성(0.66645), 말기(0.66288), 재검(0.65698), 종양(0.65557), 심근경색(0.65311), 이식(0.65302), 판정(0.64814), 확진(0.64671), 전이(0.64645), 뇌출중(0.64500), 입원치료(0.63730), 갑상선(0.63238), 진폐(0.63165), 암(0.63048), 신부전증(0.62584), 시술(0.62442)
7	보험료 /Noun	건보료(0.74681), 납부(0.71809), 의료보험료(0.70889), 건강보험료(0.70508), 납입(0.69969), 연금(0.69762), 가입자(0.67313), 지역가입(0.66770), 국민연금(0.66314), 보험(0.65021), 산재보험료(0.64702), 환급(0.64581), 환급금(0.63599), 금액(0.62880), 연금보험(0.60795), 직장가입자(0.60512), 보험금(0.60112), 고용보험료(0.59592), 산정(0.59498), 보험(0.58792)
8	건강 가정사 /Noun	자격증(0.79790), 보육교사(0.75711), 사회복지사(0.70810), 사이버대학(0.65830), 학점은행제(0.65316), 취득(0.64027), 평생교육사(0.62151), 강의(0.61914), 과목(0.61748), 선택과목(0.60318), 청소년상담사(0.59178), 이수(0.58162), 학점은행(0.56496), 수강신청(0.55878), 청소년지도사(0.55723), 수강(0.55601), 교육기관(0.55416), 필수과목(0.54472), 학위(0.54260), 자격(0.54240)

9	국민연금 /Noun	고용보험(0.82719), 산재보험료(0.76393), 지역가입(0.76212), 주민세(0.74239), 납부(0.73072), 가입자(0.71787), 체납(0.70872), 건보료(0.70493), 고용보험료(0.70252), 소득세(0.68170), 직장가입자(0.67893), 갑근세(0.67636), 연금보험(0.67432), 보험료(0.66314), 건강보험(0.66054), 연금(0.65338), 역서(0.63229), 장애연금(0.62942), 건강보험료(0.62576), 고지서(0.62292)
10	입원 /Noun	퇴원(0.79788), 입원치료(0.77817), 통원치료(0.75790), 재활병원(0.72114), 통원(0.69992), 중환자실(0.68607), 입원비(0.65297), 재활치료(0.63780), 수술(0.63474), 정신병원(0.62401), 수술비(0.61022), 치료(0.58145), 이식(0.57727), 일초(0.57404), 실밥(0.56837), 간병인(0.56241), 치료비(0.56043), 병원비(0.55523), 간병(0.55029), 병상(0.54055)
11	검사 /Noun	조직검사(0.78615), 정밀검사(0.78612), 내시경(0.75517), 건강검진(0.74730), 재검(0.72304), 촬영(0.71970), 검진(0.71022), 진찰(0.68585), 정밀(0.68215), 엑스레이(0.67419), 대학병원(0.64362), 찍다(0.63176), 확진(0.63014), 소견서(0.62903), 초음파(0.61863), 골수(0.61672), 의사 선생님(0.61138), 진료(0.61069), 예방접종(0.60788), 재검사(0.60638)
12	산재처리 /Noun	산재신청(0.85741), 산재(0.83375), 공상처리(0.78132), 산재승인(0.75308), 처리(0.72415), 공상(0.69384), 합의(0.67602), 복직(0.66238), 보상(0.65580), 노무사(0.64591), 휴업급여(0.63670), 승인(0.63390), 치료비(0.62676), 보험처리(0.62432), 다치다(0.61957), 산재보험(0.61867), 퇴사(0.61753), 권고사직(0.61503), 산업재해(0.61180), 회사(0.61058)
13	산재 /Noun	산재처리(0.83375), 산재신청(0.78889), 산재승인(0.77971), 승인(0.74468), 휴업급여(0.69827), 산재보험(0.69585), 산업재해(0.69050), 종결(0.68818), 공상(0.66788), 근로복지공단(0.66208), 보험처리(0.64826), 공상처리(0.64142), 처리(0.63132), 불승인(0.61445), 다치다(0.61214), 복직(0.60832), 노무사(0.59627), 실업급여(0.59056), 치료비(0.59051), 권고사직(0.58462)
14	처리 /Noun	산재처리(0.72415), 산재신청(0.70914), 공상처리(0.68652), 권고사직(0.66814), 복직(0.65443), 공상(0.65156), 퇴사(0.64947), 해고(0.63930), 사업주(0.63432), 산재(0.63132), 날인(0.62938), 산재가입(0.62608), 청구(0.61674), 휴직(0.61210), 보험처리(0.61188), 산재보험(0.61146), 근로복지공단(0.60720), 고용주(0.60313), 병가(0.59280), 승인(0.58652)
15	보육교사 /Noun	사회복지사(0.79676), 건강가정사(0.75711), 자격증(0.75356), 이수(0.72218), 학점은행제(0.72008), 취득(0.70354), 평생교육사(0.66297), 수강신청(0.66023), 과목(0.65698), 학점은행(0.65622), 수강(0.65502), 학위(0.64842), 필수과목(0.64057), 학사(0.63972), 자격시험(0.63674), 강의(0.62515), 고졸(0.61595), 가정사(0.61189), 아동학(0.60338), 실습(0.60316)

거. ‘서민금융’ 분석결과

앞서 데이터의 특징에서 설명하였듯이, ‘서민금융 + 복지’ 키워드는 포털사이트에서 검색결과가 거의 없다. 따라서 복지로의 ‘서민금융’ 하위 카테고리인 ‘생계자금’, ‘주거자금’, ‘고용지원’, ‘학자금’ 관련된 다빈도 형태소들을 분석하였고, 그 결과는 <표 4-33>와 같다. 우선, ‘학자금대출’, ‘한국장학재단’, ‘국가장학금’ 등 대학생 학비와 관련된 형태소가 다빈도로 나타났고, ‘연체’, ‘신용’, ‘이자’, ‘은행’, ‘빚’ 등 대출과 관련된 형태소들도 있다.

<표 4-33> ‘서민금융’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	학자금대출/Noun	1,032	0.6498
2	연체/Noun	315	0.6258
3	자금/Noun	223	0.6227
4	한국장학재단/Noun	215	0.6224
5	신용/Noun	185	0.6214
6	상환/Noun	596	0.4926
7	이자/Noun	385	0.4871
8	은행/Noun	229	0.4831
9	고용/Noun	218	0.4828
10	유형/Noun	214	0.4827
11	장학금/Noun	956	0.4217
12	등록금/Noun	757	0.4173
13	국가장학금/Noun	586	0.4136
14	금융/Noun	273	0.4068
15	대학생/Noun	177	0.4047
16	학비/Noun	172	0.4046
17	갚다/Verb	515	0.3593
18	빚/Noun	343	0.3560
19	연봉/Noun	330	0.3558
20	통장/Noun	265	0.3161

다빈도 형태소들을 기준으로 유사도를 살펴보면, ‘학자금대출-한국장학재단-장학금-등록금-국가장학금-대학생-학비’ 등이 모두 유사한 것으로 나타났다. 이를 제외하고 살펴보면, ‘연체’와 관련해서는 ‘신용카드’, ‘카드’, ‘할부’ 등의 형태소와 함께 ‘저축은행’, ‘캐피탈’, ‘현금서비스’, ‘부채’ 등의 형태소도 있다. ‘자금-신용-상환-이자-은행’과 관련해서도 대출과 관련된 형태소들이 다수 나타났다.

<표 4-34> ‘생계자금’, ‘주거자금’, ‘고용지원’, ‘학자금’ 관련 유사도

순위	단어	단어유사도(word2vec)
1	학자금 대출 /Noun	사정(0.80762), 대출(0.80659), 등록금(0.80228), 대출금(0.78894), 학비(0.76824), 학자금(0.76575), 든든학자금(0.76038), 이자(0.76022), 상환(0.75858), 학기(0.75376), 장학재단(0.75079), 생활비(0.73795), 용자(0.73463), 외환은행(0.73049), 모자라다(0.72960), 장학금(0.72519), 면제(0.72507), 국가장학금(0.71533), 받다(0.70844), 갚다(0.70388)
2	연체 /Noun	신용카드(0.94640), 카드(0.94264), 체크카드(0.92854), 농협(0.89005), 사금융(0.88561), 주거래(0.87820), 국민은행(0.87795), 은행(0.87046), 미납(0.86759), 연체기록(0.86076), 할부(0.85742), 대략(0.85599), 저축은행(0.85539), 적금(0.85455), 캐피탈(0.85391), 현금서비스(0.85120), 결제(0.84032), 부채(0.83955), 꾸준하다(0.83900), 솔로몬(0.83436)
3	자금 /Noun	주택기금(0.96548), 생계(0.92892), 신용대출(0.92326), 창업(0.90973), 새희망홀씨(0.90684), 자금대출(0.90626), 한도액(0.90041), 햇살론(0.89157), 전세자금(0.89089), 현대(0.88626), 캐피탈(0.88222), 가능(0.87732), 보증(0.87088), 신용회복(0.86708), 신혼부부(0.86087), 기업은행(0.86078), 신한은행(0.85871), 직장인(0.85843), 무직자(0.85743), 현대캐피탈(0.85226)
4	한국 장학재단 /Noun	수혜(0.93463), 농어촌(0.93345), 든든학자금(0.93234), 홈페이지(0.90132), 드림장학금(0.89899), 국장(0.89246), 감면(0.88150), 신입생(0.88099), 이중(0.88078), 공지(0.87513), 국가장학금(0.87023), 장학재단(0.86350), 중복(0.86029), 장학생(0.85017), 장학(0.84609), 소득분위(0.84129), 순위(0.83634), 완료(0.82555), 이중수혜(0.82453), 직전(0.82437)
5	신용 /Noun	신용등급(0.96034), 조회(0.94158), 금융(0.93081), 은행(0.92045), 신용카드(0.91591), 햇살론(0.88430), 금리(0.88270), 신한은행(0.88266), 은행권(0.88262), 이율(0.87670), 캐피탈(0.87227), 사금융(0.87038), 리드코프(0.86587), 신용대출(0.86469), 바꾸다(0.86290), 한도(0.86143), 저축은행(0.85443), 원리금(0.85013), 부결(0.83863), 연체기록(0.83857)

순위	단어	단어유사도(word2vec)
6	상환 /Noun	환하다(0.86246), 대출금(0.86058), 원금(0.85242), 장학재단(0.84970), 이자(0.84061), 납부(0.82528), 입금(0.82466), 계좌(0.82238), 분할 납부(0.82182), 든든학자금(0.82174), 중도(0.81415), 자동(0.80673), 실행(0.80264), 반환(0.79544), 이중수혜(0.78446), 중지(0.78270), 원리금(0.77536), 거치(0.77520), 승인(0.77405)
7	이자 /Noun	대출금(0.95598), 원금(0.93646), 갚다(0.89725), 남다(0.86880), 은행(0.85876), 저축은행(0.85136), 상환(0.84061), 원리금(0.83606), 러쉬앤캐쉬(0.82525), 갚(0.82496), 나머지(0.82139), 리드코프(0.81382), 균등(0.81123), 솔로몬(0.80834), 농협(0.80519), 부담스럽다(0.80400), 사금융(0.79783), 빌리다(0.79580), 백만원(0.78813), 내다(0.78730)
8	은행 /Noun	한도(0.94524), 저축은행(0.94023), 사금융(0.92616), 캐피탈(0.92580), 농협(0.92259), 신용(0.92045), 원리금(0.91896), 신한은행(0.91421), 은행권(0.91338), 솔로몬(0.90972), 금리(0.90630), 신용카드(0.90376), 카드(0.90354), 금융(0.90080), 리드코프(0.89960), 부결(0.89829), 국민은행(0.89800), 보증(0.89779), 론(0.89763), 주거래(0.89380)
9	고용 /Noun	센터(0.98075), 장려금(0.89649), 노동부(0.84129), 워크넷(0.83943), 사업주(0.83708), 확인서(0.82280), 사무소(0.81214), 기관(0.80346), 재취업(0.78395), 국민건강보험공단(0.77400), 체계(0.76629), 제도(0.76096), 변경(0.75495), 업무(0.75256), 산재보험(0.75048), 채용(0.75029), 고용센터(0.74992), 산재(0.74406), 작성(0.74077), 촉진(0.73965)
10	유형 /Noun	교내(0.95733), 국가장학금(0.94573), 국장(0.93970), 입학금(0.91471), 이중수혜(0.91341), 소득분위(0.91174), 중지(0.90117), 장학금(0.90085), 뜨다(0.88970), 반환(0.88576), 입금(0.88229), 분위(0.88046), 장학(0.88005), 장학생(0.87796), 드림장학금(0.87535), 든든학자금(0.87202), 면제(0.87077), 장학재단(0.87052), 고지서(0.86994), 수업료(0.86849)
11	장학금 /Noun	교내(0.90426), 유형(0.90085), 이중수혜(0.89322), 국가장학금(0.89299), 장학재단(0.86208), 반환(0.84686), 든든학자금(0.82968), 장학(0.81534), 분위(0.81474), 중지(0.81281), 국장(0.80972), 전액(0.80972), 소득분위(0.80773), 학기(0.79996), 매(0.79249), 입금(0.79115), 한정(0.79073), 학교(0.78681), 가요(0.78642), 등록금(0.78546)
12	등록금 /Noun	매(0.88133), 학기(0.86819), 계절(0.86518), 장학재단(0.84964), 학비(0.82746), 기숙사비(0.80353), 학자금대출(0.80228), 장학금(0.78546), 날다(0.78529), 반수(0.78454), 국가장학금(0.78099), 반환(0.77996), 전액(0.77979), 이중수혜(0.77971), 야하다(0.77948), 학교(0.77501), 든든학자금(0.77155), 입학금(0.76982), 원서(0.76826), 성적(0.76665)

순위	단어	단어유사도(word2vec)
13	국가 장학금 /Noun	유형(0.94573), 국장(0.92536), 이중수혜(0.91801), 든든학자금(0.91359), 소득분위(0.89858), 장학금(0.89299), 면제(0.88949), 학기(0.88812), 뜨다(0.88517), 농어촌(0.87829), 드림장학금(0.87198), 한국장학재단(0.87023), 반환(0.86785), 교내(0.86731), 중지(0.86659), 장학재단(0.86132), 수혜(0.86049), 탈락(0.84050), 신입생(0.83524), 완료(0.83473)
14	금융 /Noun	이율(0.96588), 금리(0.96141), 햇살론(0.94358), 은행권(0.93865), 신용(0.93081), 상품(0.93002), 전세자금(0.91276), 새희망홀씨(0.91140), 신용대출(0.91038), 캐피탈(0.90228), 은행(0.90080), 신용회복위원회(0.89442), 리드코프(0.89222), 저축은행(0.88754), 론(0.88725), 바꾸다(0.88489), 한도(0.88239), 한도액(0.87933), 낮다(0.87658), 조회(0.87534)
15	대학생 /Noun	재학(0.89525), 대학교(0.83614), 학생(0.83030), 내년(0.82728), 학년(0.81993), 사립대(0.79207), 올해(0.78356), 학번(0.77453), 재학생(0.74943), 예비(0.74047), 고등학생(0.72253), 마치고(0.71871), 학비(0.71521), 대학입학(0.71473), 신입생(0.71149), 동생(0.70975), 남학생(0.70957), 예정(0.70818), 거들다(0.70723), 다르다(0.70530)
16	학비 /Noun	대학(0.83348), 등록금(0.82746), 대학교(0.80907), 아르바이트(0.77900), 생활비(0.77898), 기숙사비(0.77878), 학자금(0.77547), 전액(0.77340), 학자금대출(0.76824), 학생(0.76452), 학업(0.76240), 비싸다(0.75480), 학교(0.74264), 다니다(0.73827), 들어가다(0.73280), 대학생(0.71521), 학기(0.71443), 사정(0.71164), 형편(0.71078), 마련(0.70417)
17	갚다 /Verb	이자(0.89725), 대출금(0.88088), 빌리다(0.85242), 남다(0.84884), 갚(0.84533), 빚(0.84353), 돈(0.81688), 내다(0.80013), 모으다(0.79784), 원금(0.79705), 당장(0.79584), 내기(0.79019), 빠듯(0.78229), 한번(0.77462), 꾸준하다(0.77443), 얼마(0.77277), 밀리다(0.76801), 나중(0.76783), 마련(0.76239), 여유(0.75921)
18	빚 /Noun	갚다(0.84353), 월세(0.84013), 빌리다(0.83284), 나가다(0.82256), 전세금(0.81146), 병원비(0.80741), 빚(0.80661), 보증금(0.80391), 양육비(0.79987), 집(0.79879), 두다(0.79518), 용돈(0.79442), 갚(0.79429), 천만원(0.79239), 반지하(0.78927), 빠듯(0.78577), 모으다(0.78508), 신용불량(0.78360), 신랑(0.78114), 해결(0.77615)
19	연봉 /Noun	계약직(0.84732), 재직(0.82112), 직장(0.79755), 실수령액(0.78725), 최소(0.78575), 보험(0.77866), 정직(0.77542), 정규직(0.77254), 삼성(0.77215), 근무(0.76459), 세후(0.76257), 연체기록(0.75196), 정도(0.75056), 입사(0.74961), 월급(0.74949), 진급(0.74672), 대다(0.74480), 넘다(0.73662), 잔업(0.73489), 현대(0.71736)
20	통장 /Noun	금액(0.91114), 이체(0.89690), 보험료(0.88376), 내면(0.86983), 나머지(0.86979), 계좌(0.86775), 가압류(0.85821), 건강보험료(0.85564), 만원(0.85406), 현금(0.85210), 수령(0.85141), 자동(0.85035), 생계비(0.84681), 가량(0.83632), 입금(0.83096), 제외(0.82850), 내역(0.82832), 결제(0.82665), 이름(0.82285), 마이너스(0.82280)

너. ‘문화’ 키워드 분석결과

문화관련 다빈도 형태소는 다양하게 나타났다. 문화라는 키워드가 하나의 영역이라기보다는 다양한 분야에서 사용되는 용어이기 때문에 복지와 관련된 사항을 도출하기는 어려웠다.

<표 4-35> ‘문화’ 관련 TF-IDF를 고려한 다빈도 형태소

순위	단어	빈도(개)	TF-IDF
1	유럽/Noun	757	0.6215
2	외국어/Noun	708	0.6211
3	예술/Noun	629	0.6205
4	관광/Noun	587	0.6201
5	통일/Noun	893	0.4829
6	역사/Noun	852	0.4826
7	시민/Noun	757	0.4820
8	수리/Noun	666	0.4814
9	민족/Noun	646	0.4813
10	경영/Noun	643	0.4813
11	주의/Noun	612	0.4811
12	참여/Noun	607	0.4810
13	생산/Noun	595	0.4810
14	발달/Noun	587	0.4809
15	예산/Noun	579	0.4809
16	행정/Noun	894	0.4057
17	중국/Noun	853	0.4055
18	국제/Noun	786	0.4051
19	도시/Noun	777	0.4050
20	인구/Noun	706	0.4047

<표 4-36> ‘문화’ 관련 다빈도 형태소별 유사도

순위	단어	단어유사도(word2vec)
1	유럽 /Noun	북부(0.75210), 아시아(0.75153), 남부(0.74437), 프랑스(0.73917), 스웨덴(0.72603), 덴마크(0.70964), 이탈리아(0.69808), 서부(0.69537), 그리스(0.69282), 아메리카(0.69128), 동남아시아(0.68335), 동유럽(0.68315), 영국(0.68305), 남아메리카(0.67914), 스페인(0.67646), 아프리카(0.66936), 중동(0.66771), 독일(0.66326), 동부(0.66308), 유럽연합(0.65776)

순위	단어	단어유사도(word2vec)
2	외국어 /Noun	한국지리(0.84768), 수리(0.84131), 근현대사(0.83391), 언어(0.79737), 근현(0.78639), 사문(0.76376), 표준점수(0.75842), 분위(0.75681), 등급(0.74240), 한문(0.73742), 세계사(0.73528), 언(0.73141), 국사(0.72529), 메가스터디(0.70886), 수학(0.70551), 표점(0.70266), 모의고사(0.69859), 영역(0.68751), 정석(0.68345), 점수(0.68219)
3	예술 /Noun	대극장(0.67650), 뮤지컬(0.66704), 음악(0.65669), 스포츠(0.65493), 백석(0.64688), 레저(0.64547), 장르(0.64459), 콘텐츠(0.64436), 체육(0.64381), 무용(0.64239), 실버(0.62946), 대중문화(0.62942), 인류학(0.62263), 연극(0.62258), 신학교(0.62250), 민화(0.61814), 공연(0.61290), 디지털(0.60054), 사이버대(0.60037), 전당(0.59304)
4	관광 /Noun	레저(0.76498), 콘텐츠(0.70222), 영문과(0.69405), 러시아어(0.69057), 산업대학(0.68680), 전주대(0.68209), 비즈니스(0.67705), 경상(0.67269), 실버(0.66726), 통역사(0.66573), 청강(0.66556), 외식(0.66432), 콘텐츠(0.66325), 레저(0.66079), 경영학(0.64668), 비즈니스(0.64524), 패션(0.64239), 호텔경영(0.64007), 국문(0.63731), 뷰티(0.63417)
5	통일 /Noun	북한(0.72441), 남북통일(0.70891), 상봉(0.70546), 분단(0.70213), 남한(0.68576), 남북관계(0.67303), 남북한(0.66969), 남북(0.64614), 조국(0.63715), 한반도(0.62365), 평화통일(0.61937), 민족(0.59776), 평화(0.56747), 한민족(0.56278), 도약(0.53842), 이루다(0.53384), 군사(0.53291), 민족국가(0.53241), 국정교과서(0.52546), 염원(0.52356)
6	역사 /Noun	힌두교(0.72378), 인물(0.71215), 풍습(0.71050), 불교(0.69237), 유래(0.69074), 문화유산(0.68358), 이탈리아(0.68303), 전통(0.67336), 스페인(0.65508), 유산(0.65157), 핀란드(0.63526), 소설(0.63297), 국기(0.63023), 전통문화(0.62670), 조상(0.61420), 문화재(0.61078), 노르웨이(0.60622), 영국(0.60413), 히브리어(0.60367), 주제(0.60264)
7	시민 /Noun	연대(0.66916), 민주(0.66574), 의식(0.65940), 자발(0.65420), 참여(0.64122), 나눔(0.63086), 참여연대(0.62810), 실천(0.60654), 경실련(0.60234), 공동체(0.59932), 국가인권위(0.59583), 모임(0.59373), 구현(0.59277), 시민단체(0.59163), 시민사회(0.58881), 유권자(0.58835), 환경운동(0.58430), 단체(0.58337), 행복추구권(0.58123), 질서(0.57536)
8	수리 /Noun	외국어(0.84131), 사문(0.78475), 표준점수(0.76248), 정석(0.74816), 근현(0.74178), 분위(0.73945), 표점(0.73524), 수외(0.72860), 메가스터디(0.72105), 언(0.71055), 등급(0.69897), 한국지리(0.68398), 가군(0.68136), 점수(0.67887), 만점(0.67257), 언어(0.65565), 연수위담(0.65403), 영역(0.65237), 언외(0.65015), 법사(0.64328)
9	민족 /Noun	한민족(0.80836), 민족주의(0.76034), 평화(0.75250), 계승(0.74650), 침략(0.74349), 일제(0.73380), 대항(0.73287), 외세(0.72036), 불교(0.71147), 단결(0.69761), 사상(0.69032), 유교(0.68760), 한반도(0.68581), 식민지(0.68552), 미풍양속(0.68218), 공존(0.67833), 남북한(0.67420), 동질(0.67213), 사회주의(0.66984), 자긍심(0.66918)
10	경영 /Noun	경영학(0.77559), 마케팅(0.72811), 상경(0.71175), 경영정보(0.68619), 마이(0.67502), 경제학(0.66861), 비즈니스(0.64969), 원론(0.64556), 윤리경영(0.64245), 재무(0.63305), 중부대(0.63051), 통계학(0.62954), 소신(0.62872), 외식(0.62591), 행정학(0.62526), 법학과(0.62450), 세무(0.62036), 일본학(0.62018), 어문(0.61963), 고려대(0.61742)

순위	단어	단어유사도(word2vec)
11	주의 /Noun	근본(0.79280), 사상(0.77124), 이데올로기(0.76958), 삼균(0.75957), 파시즘(0.75779), 권위(0.75660), 나치즘(0.74218), 자유주의(0.73000), 이념(0.72882), 주체성(0.72700), 사회주의(0.72538), 만능(0.72285), 문화상(0.71445), 개인주의(0.71407), 자본주의(0.71237), 계급(0.70882), 상실(0.70874), 사대주의(0.70870), 공산주의(0.70314), 이기주의(0.69586)
12	참여 /Noun	시민(0.64122), 수렴(0.63230), 실천(0.63183), 자원봉사(0.62085), 통(0.61313), 직접(0.60780), 감시(0.58090), 연대(0.57797), 봉사(0.57789), 적극(0.57246), 행사(0.56907), 리더(0.56402), 토론회(0.56167), 토대(0.55880), 모임(0.55586), 여가활동(0.54722), 활동(0.54435), 보람(0.54180), 구현(0.53539), 자발(0.53427)
13	생산 /Noun	재화(0.80937), 대량(0.79745), 소비(0.76916), 증대(0.74640), 공급(0.72669), 시장(0.72457), 원료(0.71689), 활발해지다(0.70625), 교환(0.70544), 가능해지다(0.70485), 이윤(0.70456), 식량(0.69839), 수공업(0.69737), 노동(0.69551), 에너지(0.69444), 생산자(0.69435), 축적(0.69281), 상업(0.69221), 수단(0.68230), 수요(0.67898)
14	발달 /Noun	변화(0.72784), 발명(0.70471), 평야(0.68183), 농업(0.67445), 공업(0.67230), 고기압(0.67069), 급격(0.66430), 확산(0.65835), 상업(0.65762), 고도화(0.64882), 공업화(0.64002), 생산량(0.63908), 등장(0.63828), 만들어지다(0.63716), 변동(0.63401), 남부(0.63265), 동식물(0.62924), 지하자원(0.62865), 산지(0.62838), 도시화(0.62519)
15	예산 /Noun	지출(0.67192), 금액(0.67119), 특별법(0.62354), 여성가족부(0.62159), 조원(0.61628), 지원금(0.61169), 낭비(0.61096), 편성(0.60855), 억원(0.60221), 삭감(0.59966), 예산안(0.58848), 부채(0.58281), 청문회(0.57708), 세월호(0.57107), 재정(0.56944), 용도(0.56886), 투입(0.56883), 행정기관(0.55843), 조세(0.55016), 세금(0.54836)
16	행정 /Noun	지방행정(0.67266), 회계(0.61747), 분권(0.60931), 계산(0.60920), 행정학(0.60318), 재경(0.59677), 공무원(0.59012), 실마리(0.58507), 관료제(0.58338), 타운(0.57150), 지방직(0.56947), 개혁(0.56749), 정부조직(0.56486), 자치행정(0.56477), 법무(0.56290), 행정직(0.56226), 교육행정(0.56096), 범죄수사(0.55191), 지방정부(0.55126), 용인시청(0.55056)
17	중국 /Noun	러시아(0.79023), 인도(0.72737), 대국(0.72256), 일본(0.71109), 아시아(0.69200), 강국(0.68396), 브라질(0.68028), 페루(0.66185), 예멘(0.64960), 대등(0.64675), 북경(0.64649), 군사력(0.64227), 동북공정(0.64189), 스페인(0.64135), 강대국(0.64135), 싱가포르(0.63691), 인도네시아(0.63583), 이탈리아(0.63561), 동남아(0.63559), 홍콩(0.63265)
18	국제 /Noun	협력(0.71412), 통상(0.70413), 연맹(0.64879), 기구(0.64657), 교섭(0.64480), 조약(0.64353), 회계기준(0.64351), 수상(0.64315), 무역(0.62942), 산학(0.62555), 한중(0.62303), 진출(0.61498), 교류(0.60887), 수교(0.60398), 세미나(0.59622), 샌프란시스코(0.59556), 국제기구(0.59243), 본부(0.58699), 정세(0.58684), 외교(0.57702)

순위	단어	단어유사도(word2vec)
19	도시 /Noun	도심(0.74542), 대도시(0.73196), 중심지(0.70653), 농촌(0.68720), 평야(0.67794), 해안(0.67325), 복합(0.66587), 늪(0.66225), 경관(0.65733), 자연환경(0.65576), 울창(0.65273), 위치(0.65091), 농경지(0.64882), 사계절(0.64832), 하천(0.64732), 넓다(0.64589), 산지(0.63930), 일대(0.63781), 관광객(0.63772), 편리하다(0.63139)
20	인구 /Noun	생산량(0.75568), 본포(0.73238), 면적(0.72415), 강수량(0.70889), 기후(0.68347), 감소(0.66584), 고령화(0.65696), 밀집(0.65458), 지형(0.65452), 노령화(0.63426), 출산율(0.63373), 증가(0.62296), 산지(0.62188), 기온(0.61979), 피라미드(0.61785), 활엽수(0.61575), 부족(0.61495), 식량(0.61111), 연평균(0.60478), 촌락(0.60173)

2. 분석의 한계

앞에서 설명한 결과는 불특정 다수를 대상으로 한 인터넷 문서를 기반으로 한 분석이기 때문에 데이터의 오탈자가 다수 존재해 데이터 전처리 작업에만 상당한 절차와 시간이 요구되었다. 그리고 연구에 많은 시간을 할애해 텍스트를 수정하였음에도 인터넷의 특성상 이용되는 문자인 은어, 단축단어 등의 다양한 상황을 모두 해결하기에는 부족해 모델링에 약점이 있다. 예를 들어, 포털사이트는 연관검색어와 관계가 없는 문서가 수집되기도 하는데, 이는 복지영역별로 추출한 문서임에도 불구하고 전혀 관계가 없는 문서들이 데이터로 포함됨을 의미한다. 또한 기계학습의 특성상 대용량의 자료가 수집되지 않으면 일반화된 결과 도출에 어려움이 발생한다. 특히, 텍스트의 구조적인 문제로 다른 기계학습보다 모델링을 위해서 더 많은 양의 데이터가 요구된다. 또한, 설정한 키워드가 일반 국민이 얼마나 자주 사용하는 키워드인가에 대한 검토와 함께 분석을 위한 특정 웹사이트를 자주 사용하는 계층에 대한 검토도 필요하다. 앞서, ‘영유아’, ‘아동청소년’ 등과 같은 키워드 분석결과에서도 볼 수 있듯이, 다수의 분석 문장이

‘진학’, ‘취업’ 등과 관련된 형태소들로 본 연구의 내용에는 부합하지 않았다. 또한 인터넷을 활용하는 연령과 소득을 중심으로 살펴본다면, 10~40대 사이의 연령층의 인터넷 사용률은 98% 이상으로 매우 높은 반면 50대 이후의 인터넷 이용률은 상대적으로 낮으며, 저소득층(가구 소득 100만원 미만)의 경우 인터넷 이용률은 37.6%로 월평균 가구의 소득 400만원 이상 가구(95.1%)보다 57.5%p 낮게 나타났다.²¹⁾ 결국 보다 정확한 분석을 위해서 연령과 가구소득 등을 고려해야 함에도 불구하고, 포털데이터 사용에 있어서 공개범위의 제한 등으로 이를 반영할 수 없는 한계가 있다.

21) 한국인터넷진흥원(2015), 2015년 인터넷이용실태조사

제3절 초기상담 기능강화 방안

상담에서 대상자와 상담자의 유대관계 형성은 상호 존중과 신뢰가 바탕이 된다. 상호 존중과 신뢰가 바탕이 된 관계를 라포(rapport)라 하며, 관계의 형성을 위해 관련된 지식과 기술의 습득도 필요하지만, 마음이 통할 수 있는 상태를 유지하는 것이 중요하다. 이를 위해서 생각 및 언어를 동일시하는 것이 필요하다. 즉, 동일한 단어를 써도 다른 표현으로 이해하면 라포의 형성은 어려울 것이다.

복지영역별 토픽 분석결과는 상담을 통해 대상자가 고민하는 것을 특정 단어를 중심으로 접근하여 유대감을 형성하고, 이를 기반으로 상담자는 대상자와 라포를 형성해 대상자의 욕구를 정확히 파악하여 적절한 서비스 연계를 수행한다. 그리고 복지영역별 사회적 이슈를 시간적 흐름에 따라 분석하여 적절하게 대응하고, 현재 복지사업의 발전방안과 신규 서비스의 개발에도 활용한다.

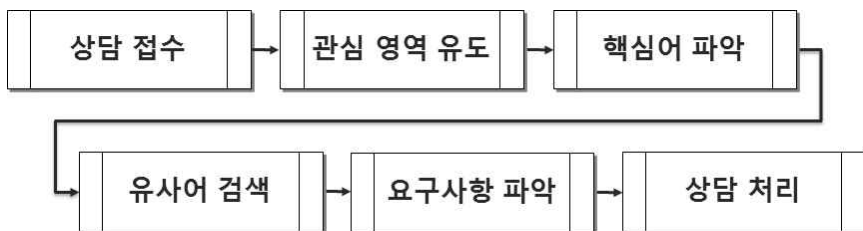
다음으로 복지로 복지영역별 정보안내를 위한 메뉴인 ‘전문가상담’, ‘관련정보’를 최근 대상자의 관심영역과 비교 검토하여 발전방안 및 고도화를 추진할 수 있을 것이다. 즉, 도출된 결과의 ‘아동청소년’의 키워드가 ‘심리상담’, ‘상담사’ 등의 요구가 많은데, 이를 기초로 ‘관련 정보’의 메뉴를 다양화하면 이용률 제고와 함께 복지로 검색 기능을 보다 손쉽게 이용할 수 있는 기반이 될 것이다.

앞의 내용은 정책지원과 복지로에 결과를 활용하는 방안이고, 이와 함께 초기상담 기능 강화방안으로 첫째, 초기상담에서 복지영역별로 상담과정에 요구하는 단어의 정확한 의미 파악과 관련된 질문에 이용할 수 있다. 즉, 대상자의 직·간접적인 표현을 명확히 이해하고 이를 파악하기 위한 추가적인 질문을 제안하여 업무의 효율성을 높일 수 있다. 예를 들어 ‘임신출산’ 관련해 상담 접수된 경우, 다빈도 단어인

‘시댁’, ‘와이프’, ‘남자친구’ 등과 관련 얘기를 시작으로 유의성 높은 단어로 확장해 나가는 방식이라고 할 수 있다. 이와 같은 방법으로 상담의 전문성 확보와 상담기능을 강화 할 수 있을 것이다.

한편 앞에 제시한 텍스트마이닝의 토픽 분석결과는 방대한 양으로 한 번에 파악하기는 쉽지 않다. 따라서 이를 시스템으로 구현해 상담 업무의 효율성을 도모할 수 있을 것이다. 즉, [그림 4-4]처럼 상담의 키워드를 입력하여 관련된 유사한 단어를 제공하여 상담이 수행될 수 있다면 일선 담당 공무원의 업무 부담을 줄일 수 있을 것이다.

둘째, 분석된 복지영역에서의 주요 토픽을 이용해 기존의 욕구분류 체계(10대 복지욕구 영역)도 보완할 수 있을 것이다. 비록 분석에서 이용된 텍스트는 인터넷 웹 사이트에서 크롤링하여 수집된 데이터로 초기상담의 대상자의 상담내용과는 차이가 있을 수 있지만 욕구사정 단계가 내담자(사이트에서는 질의자)가 필요로 하는 것(Needs)을 파악 하는 과정이라고 할 때, 위의 토픽 분석의 결과를 욕구사정 단계에 반영하는 것은 의미가 있다. 또한, 사례관리의 욕구조사가 위기가구의 관리를 중심으로 발달되었다면, 기존의 선별적 복지에서 점차 보편적 복지로의 전환됨에 따라 욕구조사도 보다 폭넓은 대상의 욕구를 반영 할 수 있도록 보완되어야 할 것이다. 결론적으로 <표 4-37>와 같이 10대 욕구 영역을 분석을 중심으로 세분화할 수 있을 것이다.



[그림 4-4] 텍스트마이닝 결과를 이용한 상담 절차

〈표 4-37〉 통합사례관리 10대 복지욕구 영역

연번	욕구 영역	세부 욕구 영역
1	안전	가족 내 안전유지
		가족 외부로부터의 안전유지
2	건강	신체적 건강유지
		정신적 건강유지
3	일상생활 유지	의식주 관련 일상생활 유지
		여가생활 활용
4	가족 관계	관계형성
		가족 돌봄
5	사회적 관계	친인척 및 이웃 간 관계형성
		소속된 집단 및 사회생활
6	경제	기초 생활해결
		자산관리
7	교육	기초학습 습득 및 향상
		교육환경 개선
8	고용	취(창)업
		고용유지
9	생활환경	주거 내부환경 개선과 안전
		주거 내·외부환경과 지역사회
10	법률 및 권익보장	권익보장
		법률적 지원

앞서 분석한 주요 토픽을 10대 욕구영역에 맞추어 재분류 한다면, 우선, ‘안전’ 욕구에서는 기존의 가족 내·외로부터의 폭력, 응급상황에 대한 상담과 함께 ‘보호시설’, ‘상담프로그램’ 등과 같은 ‘복지서비스’ 등을 제시할 수 있을 것이다. ‘건강’ 욕구에서는 실제 앓고 있는 질병 이외에도 ‘건강검진’, ‘암검진’, ‘예방접종’, ‘영유아건강검진’ 등 예방 측면에서의 상담도 가능할 것이다. ‘가족관계’ 욕구에서는 영유아의 돌봄을 고려해 양육과 관련된 ‘시설’, ‘양육비’ 등을 보다 구체적으로

제시할 수 있을 것이다. ‘고용’ 욕구에서는 기존의 취·창업(구직) 및 고용유지 뿐만 아니라 ‘휴직’, ‘휴가’, ‘실업’, ‘산업재해’ 등의 사례도 제시할 수 있을 것이다. 또한, ‘법률 및 권익보장’ 욕구에서는 이혼에 따른 ‘소송’과 ‘친권양육’, ‘양육비 청구소송’ 등의 구체적 사례를 제시할 수 있을 것으로 보인다.

또한, 10대 욕구영역 외에도 특정 취약계층에 대한 욕구를 제시할 수 있을 것으로도 분석된다. ‘노인’, ‘장애인’, ‘한부모’, ‘미혼모’ 등이 갖는 욕구와 복지서비스를 보다 구체적으로 제시할 수 있을 것이다. 예를 들어, 노인의 경우에는 기초연금이나 국민연금 등 ‘노후 경제적 생활’에 대한 상담이 이루어질 수 있고, 치매나 요양 등 ‘건강 상태의 상담’도 필요로 하고 있다. 장애인의 경우에는 ‘장애판정’부터 관련 ‘장애연금’, ‘장애수당’, ‘장애인등록차량’, ‘할인’ 등의 복지서비스로 이어지는 상담 등 대상자 특성에 맞게 보다 ‘특성화된 상담’이 가능할 것이다. ‘한부모가족’과 ‘미혼모’의 경우에도 ‘선정기준’부터 ‘지원금’, ‘시설’ 등과 같은 복지서비스 등 보다 구체적인 상담이 이루어질 수 있을 것이다.

결 론

제5장 결론

정부는 ‘읍면동 중심의 복지전달체계 구축’을 위해서 읍면동과 지역 사회 ‘복지허브화’를 추진해 국민이 주민센터만 방문해도 원스톱으로 통합서비스를 제공받을 수 있는 읍면동 중심 전달체계 강화를 지속적으로 추진하고 있다. 이는 전국의 모든 읍면동에서 체계화된 업무와 확대된 복지기능으로 인적안전망과 민간복지 자원을 효과적으로 활용하여 제공하는 것이다.

주민과 인접한 읍면동의 주민센터를 공적 부조의 집행에서 종합적 서비스를 제공하는 기관으로 전환이 2016년 전국 700여개(933개소) 읍면동에서 2018년 전국으로 확대된다. 기존 읍면동을 신청과 접수의 기능에서 다양한 복합욕구를 가진 방문자의 문제를 적극적으로 해결하는 현장 맞춤형 통합서비스 기능으로 재편과 함께 읍면동에 맞춤형 복지전담팀 설치와 방문상담 및 사례관리 등으로 담당자에게 상담의 업무를 부여하고 중요성도 증대하였다.

초기상담이란 사회복지통합업무 절차에서 급여를 제공하기 위하여 신청단계에서 대상자와 상담하는 것으로 상담의 주요 사항은 대상자 및 가구의 특성, 개인 및 가구의 주요 문제와 복지 욕구를 파악하여 지원 가능한 서비스를 안내·신청하도록 하는 과정이다. 복지 체감도

제고를 위해서 초기상담 단계에 지역주민과 행정기관의 대면 과정은 매우 중요하다고 할 수 있는데, 초기상담 단계에서 신뢰가 형성되면 서비스에 대한 만족도 역시 긍정적으로 나타날 것이다.

본 연구는 일반적인 국민이 생각하는 복지영역별(임신출산, 영유아, 아동청소년, 장애인, 한부모가정 등의 16개 영역)로 관련된 텍스트를 수집(웹 크롤링)하여 각 영역별로 관심 사항과 관련된 토픽을 텍스트 마이닝 방법론을 적용하여 정보를 정형화하고자 한다. 이를 통해서 현재의 초기상담과 사례관리의 10대 욕구조사를 세분화하는 방안을 제시하고자 한다.

복지영역별 토픽 분석을 통해서 정형된 정보를 이용하는 방안으로 첫째, 초기상담에서 면담자의 상담어의 의미를 적극적으로 파악하여 추가질문을 구성하는 등 상담업무의 전문성을 높일 수 있을 것이다. 더불어 상담을 위한 전문응답 시스템을 구축하면 방대한 정보를 쉽게 이용할 수 있어 업무의 효율도 향상 될 것이다. 둘째, 현재 복지로의 복지영역별 안내를 위한 ‘전문상담’, ‘관련정보’ 메뉴를 최근 이용자의 관심영역 등으로 새롭게 구성하여 국민의 요구에 신속하게 대응 및 반영해 시스템 활용과 이용자 만족도를 제고한다. 마지막으로 시간적 흐름에 따른 추가적인 요구를 분석하여 현재의 복지사업의 개선 등의 의사결정도 지원할 수 있을 것이다.

텍스트마이닝은 데이터마이닝의 분야로 비정형 텍스트를 데이터로 하여 정형화된 정보를 도출하는 과정이라고 할 수 있다. 특정 영역의 여러 문서로부터 구조화된 정보를 모형화한 것으로, 언어의 특성상 다양한 형태로 표현되기 때문에 정형화된 데이터의 표본의 수보다는 상대적으로 많은 수의 표본이 필요 하게 된다. 이는 구조화되지 않은 문서에서 패턴을 분석하여 근거 기반의 정보를 도출하기 위한 최소의 기초 데이터로 이용되는 것이다.

본 연구에서는 빅데이터 분석 방법론 중에서 텍스트마이닝을 이용하여 복지영역별 다빈도 단어, 토픽 등의 연관어를 도출하였고, 이를 읍면동 초기상담 기능강화를 위하여 적용 방안을 제안하였다. 그러나 데이터의 한계가 있어, 효과적이고 전문적인 상담을 위해서 희망복지지원단의 통합사례관리 욕구상담 내용을 이용한 텍스트마이닝 수행이 필요하다.

사례관리의 욕구상담 자료를 활용한 분석결과는 초기상담에 직접적으로 활용할 수 있는 여러 방안이 제시될 수 있다. 예를 들어, 10대 욕구를 기준으로 다빈도 단어, 토픽을 분석할 수 있는데, 이는 앞에서 기술한 동일한 단어도 연령대에 따라 다른 의미를 포함하고 있음을 전제한 것으로 욕구별 요구사항의 파악에도 활용 가능하다. 다음으로 욕구와 토픽 기반의 복지서비스 연계가 가능하다. 즉, 상담을 통하여 축적된 정보를 활용해 대상자에게 적합한 서비스를 추천하는 것이다. 이는 대상자의 욕구와 제공 가능한 복지급여를 기반으로 복지급여의 정보가 부족한 국민들에게 적절한 서비스를 제안하게 되는 것이다. 더불어 향후에 이와 관련된 연구가 계속되어 읍면동 복지전달 체계의 완성에 기여하길 바란다.



참고문헌

1. 국내 저서

- 김성근, 허용창, 유자영(2015). 「사회복지전담공무원의 업무부담 합리화 방안 연구」. 한국행정연구원.
- 김승권, 이해영, 신윤경, 이형진, 이애희(2013). 「복지전달체계 개편 우수사례 매뉴얼」. 한국보건사회연구원.
- 김은하, 추병주, 김경준, 최현수, 오미애, 김용대, 김경범, 이창수, 신동민, 주명원, 위세아(2015). 「사회보장정보시스템을 활용한 복지 사각지대 발굴방안 연구」. 사회보장정보원.
- 김진철, 정영수, 김근온, 김배현(2014). 「빅데이터 활용을 위한 개인정보 비식별화 사례집」. 한국정보화진흥원.
- 배상태(2014). 「과학기술정책 네트워크 분석 기반의 빅데이터 활용방안 연구」. 한국과학기술평가원.
- 성은미, 민소영, 박경숙, 안혜영, 백민희, 황상미(2015). 「경기도 읍면동 복지 기능 강화중심의 전달체계 개편 모형개발 연구」. 경기복지재단.
- 정근하, 정철우(2010). 「텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구」. 한국과학기술기획평가원.
- 정홍원(2013). 「지역사회 통합적 복지서비스를 위한 지방자치단체 복지기능 강화」. 한국보건사회연구원.
- 강범일, 송민, 조화순(2013). “토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구”. 『한국문헌정보학회지』, 47(4), pp.315-334.
- 강정훈, 임호정, 고원식(2015). “빅데이터 기반 텍스트 마이닝 가시화 기술”.

- 『한국가시화정보학회 2015년도 추계학술대회 발표논문집』, pp.184-186.
- 권혜경, 이희경(2013). “초기상담에서 상담자 경력별 공감정확도와 상담만족도”. 『한국심리학회지』, 25(3), pp.397-424.
- 권혜경, 이희경(2013). “초심상담자와 숙련상담자가 지각하는 내담자 강점과 활용방안에 대한 개념도 연구”. 『한국심리학회지』, 25(3), pp.397-424.
- 김경재, 안현철(2009). “개인화된 추천시스템을 위한 사용자-상품 매트릭스 축약기법”. 『한국정보기술응용학회』, 16(1), pp.97-113.
- 김상윤, 이수원(2015). “Word2vec 모델을 이용한 대체어 후보 자동 추출”. 『한국정보과학회 2015년 동계학술발표회 논문집』, pp.769-771.
- 김윤석, 서영훈(2013). “기계 학습을 이용한 한글 텍스트 감정 분류”. 『한국엔터테인먼트산업학회 2013 추계학술대회 논문집』, pp.206-210.
- 김현중, 조성준, 강필성(2014). “KR-WordRank_WordRank를 개선한 비지도 학습기반 한국어 단어 추출 방법”. 『대한산업공학회』, 40(1), pp.18-33.
- 류애정(2013). “공공영역 사례관리자의 역할수행에 대한 실증적 분석”. 『한국지역사회복지학』, 44, pp.181-201.
- 박선형(일자 없음). “통합적_사례관리”의 강점중심 사회복지실천 특성에 대한 소고”. 『사회과학논총』, 31(2), pp.87-116.
- 박수선(2013). “조손가족 기능강화를 위한 사례관리 실천 과정 기초분석”. 『한국가정관리학회』, 31(2), pp.41-57.
- 박용욱(2012). “형태소 분석기의 어휘적 중의성 해결에 관한 연구”. 『한국전자통신학회』, 7(4), pp.783-787.
- 박은정, 조성준(2014). “KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지”. 『제26회 한글 및 한국어정보처리 학술대회 논문집』, pp.133-136.
- 박종희, 박은정, 조동준(2015). “북한 신년사 (1946-2015) 에 대한 자동화된 텍스트 분석”. 『한국정치학회보』, 49(2), pp.27-61.
- 서대석, 박미은, 서진(2011). “사례관리에서 자원연계 방해요인이 사례관리자의 직무스트레스에 미치는 영향”. 『사회과학연구』, 27(4), pp.179-209.
- 성기원(2014). “통합사례관리 실무자 역량강화를 위한 지원방안”. 『2014년 한

- 국사회복지행정학회 추계 학술대회 및 WORKSHOP』, pp.129-161.
- 성은미, 백민희(2013). “주민생활지원국체계의 한계와 대안적인 종합복지사무 행정센터 모형에 대한 연구”. 『한국보건사회연구원(보건사회연구)』, 33(3), pp.442-475.
- 손지은, 김성범(2014). “의사결정나무 모델에서의 중요 룰 선택기법”. 『대한산업공학회지』, 40(4), pp.375-381.
- 오동석, 강상우, 서정연(2016). “Word2Vec을 이용한 반복적 접근 방식의 그래프 기반 단어 중의성 해소”. 『인지과학』, 27(1), pp.43-60.
- 이오준, 유은순(2015). “추천 시스템의 성능 안정성을 위한 예측적 군집화 기반 협업 필터링 기법”. 『지능정보연구』, 21(1), pp.119-142.
- 이용구(2011). “단어 중의성 해소를 위한 지도학습 방법의 통계적 자질선정에 관한 연구”. 『한국비블리아학회』, 22(2), pp.5-25.
- 이재운(2006). “지적 구조 분석을 위한 새로운 클러스터링 기법에 관한 연구”. 『정보관리학회지』, 23(4), pp.215-231.
- 이주연(2011). “미국의 아동대상 사례관리 서비스”. 『한국보건사회연구원(보건복지포럼)』, 175, pp.106-115.
- 이충희, 임준호, 임수종, 김현기(2016). “기분분석사전과 기계학습 방법을 결합한 음절 단위 한국어 품사 태깅”. 『한국정보과학회』, 43(3), pp.362-369.
- 이현주(2007). “주민생활지원서비스 전달체계 구축 목표와 행정조직구조 개편”. 『한국보건사회연구원(보건복지포럼)』, 130, pp.22-35.
- 임병우, 오현주. (2013). “지자체 희망복지지원단에 근무하는 사례관리자의 역할수행 경험에 관한 연구”. 『한국정책연구』, 13(4), pp.131-148.
- 임좌상, 김진만(2014). “한국어 트위터의 감정분류를 위한 기계학습의 실증적 비교”. 『한국멀티미디어학회』, 17(2), pp.232-239.
- 정연정(2014). “한국사례관리실천의 딜레마”. 『한국사회복지행정학』, 16(1), pp.55-88.
- 정호석, 이종욱, 유재학, 이한성, 박대희(2011). “웹 캐스트와 텍스트마이닝을 이용한 축구 경기의 심층분석”. 『한국콘텐츠학회논문지』, 11(10), pp.59-68.

- 최명민, 정병오(2015). “근대사회 사례관리의 계보학적 분석”. 『한국정신보건 사회복지학회 학술발표논문집』, pp.79-105.
- 최지선, 민소영, 엄태영(2015). “희망복지지원단 통합사례관리사의 사례관리역량 자기평가지표개발 연구”. 『한국사회복지행정학』, 17(3), pp.229-254.
- 함철호(2013). “공공전달체계내 사례관리자의 사례관리 수행과 영향요인”. 『한국지역사회복지학』, 44, pp.1-31.
- 현윤진, 한희준, 최희석, 박준형, 이규하, 곽기영, 김남규(2013). “텍스트 분석을 활용한 국가 현안 대응 R&D 정보 패키징 방법론”. 『한국데이터베이스학회』, pp.231-257.
- 홍선미(2013). “공공사례관리도입에 따른 지역복지기관의 사례관리 기능”. 『한국사회복지행정학회 학술대회 자료집』, pp.95-107.
- 관계부처 합동(2016). 「2016년 국민행복 분야 업무계획」.
- 대전복지재단(2014). 「통합사례관리 시범사업 업무 매뉴얼」.
- 보건복지부(2016). 「2016년 희망복지지원단 업무안내」.
- 보건복지부(2016). 「사회복지통합안내지침」.
- 보건복지부(2016). 「읍면동 맞춤형 복지팀 업무 매뉴얼」.
- 서울특별시(2015). 「찾아가는 동주민센터 업무 매뉴얼」.
- 한국보건복지인력개발원(2013). 「위기가구 사례관리업무매뉴얼」.
- 한국인터넷진흥원(2015). 「2015년 인터넷이용실태조사」.

2. 국외 저서

- Feldman, R. and Dagan, I.(1995). “Knowledge discovery in textual databases”. In: *Proceedings of the First International Conference on Knowledge Discovery and Data Mining(KDD-95)*.
- Choi, J.H., Kim H.S. and Im, N.G.(2011). “Keyword Network Analysis for Technology Forecasting”. *Journal of Intelligence and Information Systems*. 17(4). pp.227-240.

- Kim, H.(2007). "Notational Analysis of Sports using Social Network Analysis". *The Korean Journal of Measurement and Evaluation in Physical Education and Sport Science*. 9(1). pp.99-112.
- Qin He.(1999). "Knowledge Discovery Through Co-Word Analysis". *Library Trends*. 48(1). pp.133-159.
- Yoon, B.G. and Park. Y.T.(2007). "Development of New Technology Forecasting Algorithm : Hybrid Approach for Morphology Analysis and Conjoint Analysis of Patent Information". *IEEE Transactions on Engineering Management*. 54(3). pp.588-599.
- James Manyika and Michael Chui.(2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.
- Kwak, H., C. Lee, H. Park, and S. Moon(2010). "What is Twitter, a social network or a news media?". *Proceedings of the 19th International World Wide Web(WWW) Conference*. pp.591-600.
- Sohn J. S. and I. J. Chung.(2011). "A Dynamic Management Method for FOAF Using RSS and OLAP Cube". *Journal of Intelligence and Information Systems*. 17(2). pp.39-60.
- Otte, E. and Rousseau, R.(2002). "Social network analysis: a powerful strategy, also for the information sciences". *Journal of Information Science*. 28(6). pp.441-453.
- Ellison, N., Steinfield, C., and Lampe, C.(2007). "The benefits of Facebook 'firends:' Social capital and college students' use of online social network sites". *Journal of Computer-Mediated Communication*. 12. pp.1143-1168.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean(2013). "Efficient estimation of word representations in vector space". *arXiv:1301.3781v3*.

3. 웹사이트

복지로. <http://www.bokjiro.go.kr>

보건복지콜센터. <http://www.129.go.kr>

네이버 지식IN. <http://kin.naver.com>

Konlpy. <http://www.konlpy.org/en/v0.4.4/>

연구보고서 발간자료 목록

보고서번호	보고서명	연구책임자
연구 12-01	국가복지사업 운영 합리화 방안 연구Ⅰ: 국가복지사업 총괄 조정 방안[1차]	홍성대
연구 12-02	국가복지사업 운영 합리화 방안 연구Ⅱ: 지방자치단체 복지사업 재정 및 현황 분석 -서울, 대전, 경기를 중심으로-	함영진
연구 12-03	국가복지사업 운영 합리화 방안 연구Ⅲ: 2012년도 복지사업 기준 표준화[1차]	이대영
연구 12-04	기초생활보장 사각지대 프로파일 분석연구	김은하
연구 12-05	보건복지사업 산출통계 분류·분석: 한국보건복지정보개발원 산출통계를 중심으로	장영미
연구 12-06	2011년 사회복지통합관리망 복지급여 통계연보	김은하
연구 12-07	국가복지사업 운영 합리화 방안 연구Ⅰ: 법정부 복지사업 유형분류 정책조정에 관한 연구	함영진
연구 12-08	국가복지사업 운영 합리화 방안 연구Ⅱ: 지방자치단체 복지사업 특징 및 영향요인 분석	함영진
연구 12-09	국가복지사업 운영 합리화 방안 연구Ⅲ: 복지사업기준 표준화 원칙 및 방향 마련 연구	이대영
연구 12-10	국민기초생활보장제도 급여지출 연구: 지역간 차이를 중심으로	김은하
연구 12-12	사회서비스 제공기관 공급실태 분석: 3대 돌봄바우처사업을 중심으로	김윤수
연구 13-01	국내 사회서비스분야 바우처사업 사례 연구	김윤수
연구 13-02	사회서비스 전자바우처 통계연보 관리방안 연구	김윤수
연구 13-03	사회서비스 제공기관 공급실태 분석: 장애인활동지원, 지역사회서비스투자사업, 장애아동 가족지원사업을 중심으로	김윤수
연구 13-04	사회서비스 전자바우처사업 발전을 위한 제도개선 방향 연구: 사회서비스이용 및 이용권 관리에 관한 법률을 중심으로	김윤수
연구 13-07	지방자치단체 복지사업정보 통합관리체계 방안 마련 연구	함영진
연구 13-08	수급이력 기반 국가복지사업 실태 분석: 사회서비스바우처 대상자를 중심으로	박선미
연구 13-09	국민기초생활보장제도 수급자격 변동유형별 소득추이 분석	이대영
연구 13-10	국민기초생활보장제도 부양의무자 특성 분석	김은하
연구 13-11	영유아 보육관련 정책선호 및 영향요인 분석	함영진
연구 13-12	사회보장정보시스템(행복e음) 복지급여 통계연보	김은하
연구 14-05	환경변화에 따른 조직발전 전략마련 연구	함영진
연구 14-06	조직브랜딩을 위한 전략 연구	김윤수

보고서번호	보고서명	연구책임자
연구 14-07	복지 사각지대 발굴 및 지원체계를 위한 DB 설계	김은하
연구 14-08	정보화 발전방향 연구: 한국보건복지정보개발원 운영시스템을 중심으로	박선미
연구 15-01	사회보장정보시스템 정책활용을 위한 기초연구	김현아
연구 15-02	보건의로 빅데이터를 활용한 지역 간 의료이용 격차 해소 근거기반 연구	강소영
연구 15-03	수급자 패널 구축의 유용성과 활용에 관한 연구	최항석
연구 15-04	정보시스템을 활용한 사회보장 통계지표 체계 개발에 관한 연구	김은하
연구 15-05	사회보장제도 공통기준 마련을 위한 기초연구	이대영
연구 15-06	사회보장급여 수급 시뮬레이션을 위한 기초연구	추병주
연구 15-07	아이행복카드 통합에 대한 효과성 평가 연구	함영진

빅데이터 텍스트마이닝을 이용한 초기상담 기능강화 방안

발 행 일 : 2016. 12

저 자 : 최항석, 김성훈

발 행 인 : 임 병 인

발 행 처 : 사회보장정보원

주 소 : 서울 중구 퇴계로 173(충무로3가) 남산스퀘어빌딩 14층(우 04554)

전 화 : 02) 6360-6114

홈페이지 : <http://www.ssis.or.kr>

인 쇄 : EM실천

02) 875-9744

ISBN 979-11-86408-13-1

* 잘못된 책은 바꿔 드립니다.