

2021년 기획연구 보고서

Part. 3

개념 거리 측정을 활용한 비영리 법인의
사업 내용 분석과 수입 구조와의 관련성에 관한
탐색적 연구: 비영리 법인 분석에서
질적, 양적 정보 통합의 한 사례

노법래 세명대학교





책임 연구원

노 법 래

아름다운재단 기부문화연구소, 연구위원
세명대학교 사회복지학과 교수

학력

서울대학교 대학원 사회복지학과 박사
서울대학교 대학원 사회복지학과 석사
서울대학교 사회복지학과

주요경력

서울대학교 사회과학연구원

연구실적

- 노법래(2020). 기부는 우리를 행복하게 만들까? 인과적 방향성 검토를 위한 종단 경로모형과 교차지연 패널모형의 결합을 중심으로. 보건사회연구, 40(1), 178-205.
- 노법래 · 문영민(2020). 장애인 노동시장 참여의 종단적 패턴은 어떻게 변화했는가? : 노동시장 참여 시퀀스의 코호트 비교를 중심으로. 한국사회복지학, 72(2), 349-371.
- 노법래 · 양경은(2019). 한국 사회 저출산 논의 구조와 그 변환에 관한 텍스트 마이닝 분석 - 2000년 이후 언론 기사문에 대한 토픽 모델링과 토픽 구조의 시계열적 변동을 중심으로. 한국사회복지학, 71(4), 154-176.
- 노법래 · 김소영(2019). 한국의 계층인식과 기부행위: 계층이동 인식의 매개효과 검증. 한국사회복지행정학, 21(4), 145-165.

개념 거리 측정을 활용한 비영리 법인의 사업 내용 분석과 수입 구조와의 관련성에 관한 탐색적 연구: 비영리 법인 분석에서 질적, 양적 정보 통합의 한 사례

노법래(세명대학교)

1. 서론

본 연구의 목적은 비영리 단체의 사업 내용을 분석하고, 사업 내용에 따라 주요 자원(기부금, 보조금)의 구성에 차이가 있는지 관찰하는 것이다. 본 연구의 목적은 비영리 단체의 운영에 필요한 전략 도출과 방법론적 측면에서 각각 의의가 있다. 우선 비영리 단체의 사업 성격과 자원 간 관련성에 대한 실증적인 검토를 통해 안정적인 사업 운영에 필요한 재무 구조에 대한 함의를 도출할 수 있을 것으로 기대한다. 한편으로 본 연구는 다양하면서도 중첩되는 지점이 많은 비영리 단체의 사업 내용을 데이터에 기반하여 체계적으로 분석하고 그 결과를 단체의 주요 양적 정보와 결합하는 과정을 담고 있다. 이는 비영리 분야와 관련해 흩어져 있는 질적(텍스트) 정보와 정형화된 양적 정보를 결합해 분석하는 시도로서 의미가 있는 것이다.

비영리 법인 운영에 있어 사업의 내용을 근거로 체계적으로 대별하는 연구가 많지 않았고, 이와 같은 분석을 통해 법인의 재무 구조나 조직 특성과 연결지어 분석한 경우는 거의 없었다고 할 수 있다. 본 연구의 시도는 비영리 단체의 활동을 전반적으로 조망하고 그 흐름을 이해하는 데 중요한 의미를 지닐 수 있다.

본 연구는 크게 두 단계를 거쳐 진행하게 된다. 첫 번째 단계는 비영리 법인의 사업 내용에 대한 텍스트 마이닝을 통해 도출된 결과를 통해 사업의 주요 대상(문제)과 그것을 해결하기 위한 방법이 무엇이며, 그 관련성이 어떠한지 기술적 차원에서 검토하는 것이다. 분석의 두 번째 단계는 이전 단계에서 도출된 사업의 내용에 따라 어떤 차이가 있는지 살펴보는 것이다. 여기서는 기부금과 정부보조금을 중심으로 총량적 측면과 양자의 비율을 동시에 검토하고자 한다.

이상의 분석을 위해 본 연구에서는 첫 번째 단계에서는 워드 벡터(word vectors, 혹은, word embedding)를 활용한 개념 간 거리 측정을 수행한다. 두 번째 단계에서는 비영리 법인에 대한 중단 자료 분석을 위해 회귀모형 가운데 하나인 선형혼합효과모형(linear mixed effects model)을 활용할 것이다.

지금까지의 논의를 종합하여 본 연구의 연구 문제를 요약하면 다음과 같다.

연구문제 1: 비영리 법인의 사업 대상(문제)과 실천 방법 사이에 관련성이 있는가?

연구문제 2: 비영리 법인의 사업 내용(대상과 방법)에 따라 수입(기부금, 보조금) 간에 차이가 있는가?

2. 분석 방법

1) 분석 자료

본 연구에서는 한국가이드스타의 비영리 법인 회계 자료를 확보하여 분석에 활용하였다. 해당 자료에는 법인명, 주요 사업 내용, 설립 근거와 같은 기본 정보와 함께 수입 및 지출에 대한 예산 정보가 포함되어 있다. 다만 본 연구의 목적상 비영리 법인 가운데 사업 내용에서 사회복지(보육 시설, 복지관, 노인 생활시설, 장기요양 시설 등)나 의료 시설, 교육(학교 법인, 산학협력 기관, 장학 단체 등)과 관련해서 직접적인 서비스를 제공하는 경우는 분석 대상에서 제외하였다. 또한 순수한 학술 연구나 정책 개발을 사업 목적으로 하는 연구소(원)와 기타 공공 기관도 분석에서 제외했다.

본 연구의 주된 내용이 사업의 내용을 분석하는 것이기 때문에 지나치게 사업 내용에 대한 설명이 짧은 사례는 분석에서 제외하였다. 사업 내용 설명 길이에 대한 기준점을 달리 적용하여 텍스트 분석 결과의 정확성을 검토하는 과정에서 본 연구는 사업 내용에 대한 설명이 40자 미만인 경우는 분석에서 최종 제외하였다. 법인이 속한 지역의 경우 사전에 제공된 주소 가운데 오류 사례가 다수 있어 법인명을 기준으로 다시 위경도 자료와 정확한 주소지를 추출하는 지오코딩(geocoding) 과정을 수행했다.

분석 자료는 격년으로 세 차례 수집된 자료(2015년, 2017년, 2019년)를 활용하였으며, 법인별로 관측 길이가 다른 언밸런스된(unbalanced) 종단 자료를 구성하여 분석에 활용하였다. 앞에서 언급한 케이스 배제 및 텍스트 정보량에 따른 선별에 의해 최종 분석에 포함된 케이스(법인)는 1,444개였으며, 관측치(observation)는 2,846건이었다.

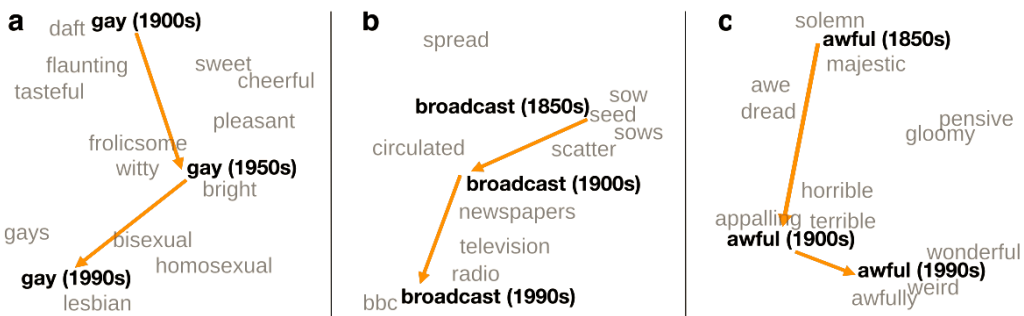
2) 분석 방법

(1) 사업 내용 분석: Word Vector and Conceptual Distance

분석 과정은 크게 두 단계로 나뉜다. 첫 번째 단계는 텍스트 마이닝을 활용한 사업 내용에 대한 분석이다. 본 연구에서는 이를 위해서 단어 벡터(word vectors)를 활용하여 선정된 주제어(key concepts)와의 의미상 거리를 측정하여 사업 내용을 분석하고자 한다. 단어 벡터는 선택된 말뭉치(corpus)에 출현하는 단어의 근접성을 기준으로 단어 간 의미상 유사성과 관련성을 측정하기 위해서 활용된다(Milkolov et al, 2013).

그림 1은 워드 벡터를 활용한 분석 사례를 소개한 것이다. 이 내용은 연구자의 관심 개념(gay, broadcast, awful)이 문헌에서 어떤 단어와 관련성이 높게 포착되고 그 변화가 어떠한지 검토한 결과이다. 각 산점도에서 거리는 워드 벡터로 구현된 의미 공간에서의 관련성의 정도로 이해할 수 있다. 예를 들어, “broadcast”의 경우 1850년대는 주로 농사와 관련된 단어(“씨뿌리기”)로 활용되었으며, 최근 시점인 1990년대에서는 현대적 의미의 방송으로 그 뜻이 변화하고 있음을 살펴볼 수 있다. 본 연구에서는 사업 내용과 관련된 주제어를 선정하고 해당 주제어가 사업 내용에 포함된 단어 간 평균적인 거리를 구하는 방식으로 의미를 파악하는 과정을 거치게 된다.

그림 1. 워드 벡터를 활용한 개념 간 거리 측정 예시(Hamilton et al, 2018)



본 연구에서 활용하는 워드 벡터는 한국어 위키피디아 텍스트에 출현하는 단어의 위치 좌표를 300차원에서 구현한 자료를 활용하기로 한다. 차원이 커질수록 단어 간 거리 측정의 정밀성이 높아질 수 있지만 단어 벡터를 구성하는 시간이 많이 걸리게 된다. 따라서 일반적으로 200~400차원 정도에서 구해진 자료를 활용하는데, 본 연구에서도 일반적으로 많이 활용되는 300차원의 정보를 이용했다.

다음으로 주제어를 선정하는 작업이 필요한데, 본 연구에서는 주요 대상(문제)과 사업 방법이라는 두 영역으로 주제어를 구성하였다. 표 1. 은 법인의 사업 내용을 파악하기 위해 활용된 개념과 구성어를 제시한 것이다. 본 연구에서 대상(문제) 영역은 크게 청소년/청년, 장애, 건강, 노인, 여성/가족, 이민자, 경제적 곤궁, 환경/지역의 8개 영역으로 나누었다. 영역의 세분화는 분석에 활용한 텍스트 정보(사업 내용)의 양에 따라 달라질 수 있는데, 많은 경우 내용이 간략하게 제시되는 경우가 많아 세분화에 한계가 있었다. 특히, 청소년/청년이나 여성/가족의 경우 사업 내용에서 핵심 대상 집단이나 문제 영역에서 일정한 차이가 있는 경우가 있었으나, 자료의 한계로 세분화에 어려움이 있었다. 핵심 문제 영역에 대한 사업 전략을 의미하는 “방법”은 교육, 자활, 연구/정책, 권익/법률, 문화, 상담/치료의 6개 영역으로 나누어 살펴보았다.

주제어를 기준으로 텍스트의 내용과의 거리를 측정할 때 단일한 단어를 활용해도 되지만, 주제어가 담고 있는 의미상의 범위나 문화적 맥락에 따른 핵심적 관계가 있는 단어가 있는 경우 복수의 단어로 구성하는 방법이 타당할 수 있다. 이 경우 개념의 공간에서의 위치는 구성어 전체의 중점(centroid)에 해당된다. 본 연구도 여러 개의 구성어를 활용하여 개념을 포착하고자 하였는데, 표 1.의 마지막 열에는 각 개념별 분석에서 실제 활용한 단어 집합을 제시하였다.

표 1. 사업 내용 분석을 위한 개념 구성

분류	개념	구성어
대상	청소년/청년	청소년, 학생, 아동, 학교, 방과후
	장애	장애, 장애인, 지체장애, 발달장애, 정신장애, 내부장애
	건강	의료, 질병, 중독, 정신건강, 우울, 자살
	노인	노인, 고령자, 노령, 치매, 독거노인
	여성/가족	여성, 가족, 성폭력, 가정폭력, 이혼
	이민자	이민자, 다문화, 외국인
	경제적 곤궁	빈곤, 채무, 신용불량, 저소득, 파산
	환경/지역	환경, 동물, 오염, 재활용, 자연, 보존, 문화재
실천 방법	교육	수업, 교육, 훈련
	자활	자활, 직업, 자격증, 근로, 자립
	연구/정책	연구, 학술, 개발, 정책
	권익/법률	권익, 인권, 법률, 소송
	문화	문화, 예술, 체육, 캠페인
	상담/치료	상담, 치료, 재활

각 법인의 사업 내용은 위에서 제시한 14개 개념과의 거리를 모두 측정하게 된다. 이 경우 문제 영역이나 방법에서 여러 개의 개념과 동시에 가까운 경우가 발생할 수 있다. 예를 들어, 특정 법인이 다문화 가정에 속한 여성에 대한 사업을 주요 사업으로 제시하고 있는 경우 “이민자”와 “여성/가족” 모두에서 높을 가능성이 있는 것이다. 이에 본 연구는 “대상”과 “방법” 각각에서 가장 거리가 가까운 개념을 대표로 선택하였다. 다만, 복수 개념과의 관련성을 분석에서 다루기 위해서 대상과 방법에서 평균 수준보다 높은 인접성을 보이는 개념이 두 개 이상인 경우(복합형)와 하나인 경우(단일형)로 구분하였다.

(2) 회귀분석: Linear Mixed Effects Model

분석의 두 번째 단계에서는 앞에서 진행한 사업 내용에 대한 분류 결과를 활용하여 비영리 법인의 수입 규모에 대한 회귀분석을 수행하고자 한다. 본 연구는 종단 자료를 활용하기 때문에 이를 적절하게 다룰 수 있는 통계 모형을 선택할 필요가 있다. 이에 본 연구에서는 최근 가장 많이 활용되고 있는 종단 모형 가운데 하나인 선형혼합효과모형(linear mixed effects model, LMER)을 택하였다. LMER은 본 연구와 같이 분석 케이스의 관측된 시계열에 차이가 많은 불균형(unbalanced) 종단 자료에서 비교적 강건성(robustness)이 높은 추정 결과를 도출하는 것으로 알려져 있다(Hasselmann, 2018). 본 연구에서는 개별 기관의 관측되지 않은 특수성을 무작위 효과(random effects)로 사업 내용을 비롯한 법인과 지역 특성 변인을 고정 효과(fixed effect)로 모형에서 다루었다.

회귀분석에서는 사업 내용과 관련된 개념과의 관련성 정도를 텍스트 마이닝을 활용하여 측정한 연속 변수 값을 그대로 활용하였다. 이를 통해 사업 내용에서 특정 개념과의 거리가 가까워지거나 멀어짐에 따라 수입 구조에서 어떤 변화가 있는지 살펴볼 수 있을 것으로 기대한다. 회귀모형에는 14개 개념과의 거리(관련성)에 대한 정보와 함께 법인의 설립 기간, 위치하고 있는 지역의 규모와 연도를 더미 변수로 투입하였다. 지역의 규모는 시군구 단위를 기준으로 광역시 이상의 단위는 “대도시”로, 그 외 시는 “일반시”로 분류하였으며, 여기에 속하지 않는 지역은 “농어촌”으로 분류하였다. 회귀분석은 R 환경에서 “lme4” 패키지의 관련 함수를 활용하였다(Bates et al, 2015).

3. 분석 결과

1) 사업 내용 분류 결과 검토

표 2. 는 대상(문제)별로 관련성이 가장 높은 사업내용을 제시한 것이다. 사업 내용은 원자료에 입력된 내용을 별다른 수정 없이 그대로 제시하였다. 다만 지면을 절약하기 위해 내용을 부분적으로 삭제하였다. 표에서 제시된 관련성은 상대적인 거리가 가까울수록 값이 큰 것으로 해석할 수 있다. 표에서 제시된 바와 같이 전반적으로 보았을 때, 텍스트 마이닝을 활용하여 사업의 주요 대상 영역을 검토한 결과에 일정한 타당성이 있는 것으로 보인다. “단일”은 앞에서 언급한 바와 같이 사업 내용에서 특정 집단에 집중도가 높은 경우를 의미하며, “혼합”은 복수의 대상 집단이 포함된 경우를 의미한다. 예를 들어, “건강”의 첫 번째 사업은 전반적인 내용이 건강 증진과 관련된 내용이지만, “저소득층”과 같이 경제적 곤궁을 경험하고 있는 집단을 포함하고 있어 “혼합”으로 분류되었다.

표 2. 대상별 주요 사업 내용

대상	사업 내용	관련성	단일/혼합
청(소)년	<ul style="list-style-type: none"> 1. 재능 발굴 등 청소년 역량강화를 위한 수련 체험활동 사업 2. 청소년 활동 지원사업 	3.507	혼합
	<ul style="list-style-type: none"> 사회적 경제적 취약계층 청년 및 청소년들의 자립을 지원 1. 학교 밖 청소년 대안교육 시설운영 	3.207	혼합
	<ul style="list-style-type: none"> 청소년의 수련활동 및 프로그램 개발 운영 청소년의 문화예술체험 활동진흥 청소년의 자원봉사 	2.865	단일
장애	<ul style="list-style-type: none"> 장애인 활동지원사업 등(심리상담지원사업, 성폭력예방교육, 장애인지원센터운영, 자원봉사자 개발 및 관리 등) 	3.716	단일
	<ul style="list-style-type: none"> 장애인 복지사업과 지역사회 복지사업을 수행하고 발달장애인의 복지증진에 기여함을 목적 	3.416	혼합
	<ul style="list-style-type: none"> 기금의 운용관리 및 배분평가 장애인 인권 신장 및 인식개선 지원사업 장애인 단체 지원사업 	2.843	혼합
	<ul style="list-style-type: none"> 1. 저소득층의료비후원금 2. 의료복지기관후원금 3. 의료복지장학금 4. 이웃 사랑계도실천사업 	4.470	혼합

건강	<ul style="list-style-type: none"> 북한 어린이 건강개선 사업, 조선족 어린이 건강개선 사업, 사랑의약품뱅크 사업(의약품 무료지원 사업), 연구사업 	2.007	혼합
	<ul style="list-style-type: none"> 유기동물 및 길동물 그리고 취약계층의 반려동물 치료 지원사업 / 반려동물 인식개선 교육, 캠페인 사업 	1.747	혼합
노인	<ul style="list-style-type: none"> 지역 주민의 복지향상과 소외된 노인을 위하여 노인복지 증진에 기여함을 목적으로 한다. 	2.559	혼합
	<ul style="list-style-type: none"> 노인복지법에 의한 노인복지사업을 시행하여 지역사회 복지증진에 기여함을 목적으로 한다. 	2.307	혼합
	<ul style="list-style-type: none"> 노인들의 다양한 복지요구에 따른 서비스를 체계적이고 복합적으로 제공하여 지역사회 노인복지 증진에 기여 	1.877	단일
여성가족	<ul style="list-style-type: none"> 1. 가족생활교육 및 상담지원 2. 가족봉사 3. 가족가치확산 캠페인 및 홍보 4. 가족복지 유관단체사업 및 연구지원 5. 취약가족 경제적 지원 	2.703	혼합
	<ul style="list-style-type: none"> 1. 추모, 선양사업 2. 자료수집, 연구, 발간 및 보급 지원사업 3. 어린이 복지, 교육사업 	2.036	단일
	<ul style="list-style-type: none"> 1. 여성인적자원 정책 연구와 프로그램 개발 / 2. 여성인력개발정보센터 설치·운영 / 3. 여성인력개발 	1.517	단일
이민자	<ul style="list-style-type: none"> 외국인 노동자 쉼터, 무료급식소 운영 등을 포함한 이주민 인권 신장 및 복지향상 	2.443	단일
	<ul style="list-style-type: none"> 다문화 이혼여성 및 자녀 자립 쉼터 운영, 다문화가족 사회통합 지원사업, 다문화자녀 및 영유아 보육사업 	1.924	단일
	<ul style="list-style-type: none"> 대한민국과 해외 여러 나라에 있는 한인 변호사 등 간의 네트워크를 수립하고 교류를 촉진하여 세계 한인 	1.873	단일
경제	<ul style="list-style-type: none"> 취약계층 일자리창출 사회서비스지원통한 삶의 질 향상 국민경제의 균형있는 발전 기여 	3.596	단일
	<ul style="list-style-type: none"> 저신용, 저소득 금융소외계층의 경제적 자생기반 마련과 사회적경제 활성화를 지원함으로써 서민생활의 안정 	2.012	단일
	<ul style="list-style-type: none"> 복지제도에서 소외된 저소득 가구의 경제적 자립과 기부문화 확산을 통한 공동체 의식 함양에 이바지함 	1.837	단일
환경	<ul style="list-style-type: none"> 인근 지역의 자연환경과 생활 환경을 적정하게 관리, 보전함으로써 지역사회의 쾌적한 환경 보전에 기여함 	3.098	단일
	<ul style="list-style-type: none"> 자연환경 문화보존 관련 정책연구 및 지원사업, 자연환경 문화보존을 위한 시민, 사회운동사업, 토론회 	2.979	단일
	<ul style="list-style-type: none"> 환경 정책대안(자연생태, 물, 대기, 토지, 에너지, 환경보건 등)개발과 제시 2. 환경관계 조사연구 3. 환경정보센터 운영 	2.121	단일

표 3.은 실천 방법별 주요 사업 내용 예시를 제시한 것이다. 여기서도 분류 결과의 타당성을 검토하기 위해 사업 내용은 수정 없이 제시하였다. 전반적으로 보았을 때, 사업 내용을 토대로한 실천 방법 분류가 타당성이 있는 것으로 판단된다. 다만, 표에서 제시되지는 않았으나, 사업 내용에 대한 설명이 지나치게 짧고 모호하게 기술된 경우나 오타가 이어지는 등의 텍스트 질에 따라서 분류 결과가 정확하지 않은 사례가 부분적으로 있어 추가적인 개선이 필요한 것으로 나타났다.

표 3. 실천 방법별 주요 사업 내용

실천 방법	사업 내용	관련성	단일/혼합
교육/훈련	<ul style="list-style-type: none"> 1.청소년 수련활동 및 수련시설의 설치 운영 2.청소년 대안교육과 청소년상담실설치운영 3.청소년 국제교류 및 문화사업 ... 	4,681	혼합
	<ul style="list-style-type: none"> 컴터 운영(이주노동자, 다문화가정), 방과후 학교운영(이주노동자, 다문화가정) 컴퓨터, 한국어교육, 의료지원, 모자원, ... 	3,164	단일
	<ul style="list-style-type: none"> 북한이탈 청소년, 청년, 다문화청소년 대상으로 전인교육, 공동체교육 및 통일교육 실시 초/중/고등 교육 실시, 민주시민 교육 ... 	2,906	단일
자활/근로	<ul style="list-style-type: none"> 비정규직의 정규직화, 청년 일자리 창출, 저소득 근로자의 고용안정 지원 등 	3,605	단일
	<ul style="list-style-type: none"> 노동상담지원, 방송업계종사자 조직화 지원, 노동환경개선 추모제 진행, 공간 대관, 방송장비대여등의 법인공관관리 운영사업 	3,024	혼합
	<ul style="list-style-type: none"> 사회적경제네트워크사업, 경제활성화사업, 공동화사업, 경제 동체, 사회적경제 판로지원 및 매장운영 	2,508	단일
연구/정책	<ul style="list-style-type: none"> 고유목적사업 현황 1. 융합기술 연구사업 - 판교제로시티 자율주행 실증단지 조성 ... 	2,331	혼합
	<ul style="list-style-type: none"> 1. 야학의 평생교육정책에 관한 연구 및 정보교환 사업 (연대사업) 2. 회원 국내외 연수 사업 (교육훈련사업) 3. 시민평생교육과정 ... 	2,142	단일
	<ul style="list-style-type: none"> 1. 중등, 고등 대안교육 사업 2. 청(소)년 대안진로교육 사업 3. 지역 및 사회적 인식확대를 위한 사업 4. 홍보 및 네트워크 연구 사업 ... 	1,645	혼합
권익/법률	<ul style="list-style-type: none"> 장애인, 고령자, 다문화가정 등의 전자정부 정보접근 이용 보장에 관한PC보급 등 지원사업... 	3,203	혼합
	<ul style="list-style-type: none"> 장애인권익증진 및 인식개선 사업 중증장애인직업재활지원 사업 장애인 특별운송 사업 편의시설 설치 시민 촉진단 사업 ... 	1,761	단일

	<ul style="list-style-type: none"> 1. 장애여성 인권사안에 대한 상담 및 관련사업 2. 장애여성 성폭력상담소 설치운영 및 피해자 지원사업 3. 장애여성 	1,411	혼합
문화	<ul style="list-style-type: none"> 1. 청소년을위한 인권보장활동 2. 청소년을 위한 보호 및 지원 활동 3. 청소년을 위한 상담활동 4. 청소년을위한 문화활동 	5,500	혼합
	<ul style="list-style-type: none"> 1. 제주올레 안내소 지킴이 민간위탁사업 2. 제주올레 그린리더 활동사업 3. 제주올레 글로벌 홍보마케팅 사업 4. 길동무 	2,775	혼합
	<ul style="list-style-type: none"> 우리가 전통생활 및 정신문화 연구와 계승발전을 통해 건전한 미래문화 창달을 목적으로 한다. 	2,633	단일
상담/치료	<ul style="list-style-type: none"> 1. 자살예방 지원사업 2. 저출산 해소 지원사업 3. 고령화극복 지원사업 4. 생명존중 지원사업 	3,380	혼합
	<ul style="list-style-type: none"> 1. 체계화된 다양한 맞춤형 운동프로그램 및 모바일 웹 개발 및 보급사업 2. 운동처방 재활 활동프로그램 운영 및 운동처방 매뉴얼... 	1,532	단일
	<ul style="list-style-type: none"> 저소득암환자지원, 소아청소년 및 저소득 성인 암 환자 심리사회적 지지 프로그램 지원, 국립암센터 국제암대학원대학교 지원, 암 연구... .. 	1,464	단일

표 4.는 법인 사업 내용에서 나타난 주요 영역과 실천 방법의 관련성을 교차표의 형태로 검토한 결과이다. 각 영역별로 가장 큰 비중을 차지하는 실천 방법은 굵게 표시하였다. 표에서 제시된 바와 같이 사업의 주요 대상에 따라서 실천 방법에서 상이한 분포가 관찰됨을 알 수 있다. “건강”과 관련해서는 “상담/치료”가 전체의 24.1%로 가장 큰 비중을 차지했다. 경제적 곤궁과 관련된 “경제”에서는 “자활/근로”가 40.6%를 차지했다. 노인의 경우는 “교육/훈련” (38.1%), “여성/가족”은 “문화” (22.8%), “이민자”와 “장애”는 각각 26.5%와 29.1%로 “권익/법률”이 가장 큰 비중을 차지했다. “청소년/청년”의 경우 “자활/근로”가 21.8%로 가장 큰 비중을 차지하였으나, 근소한 차이로 “교육/훈련”이 21.3%로 뒤를 이었다. 이는 해당 영역이 교육 관련 서비스가 주요한 청소년과 성인기에 접어들어 노동시장 진입과 관련한 욕구가 있는 청년층이 포함되어 있기 때문으로 보인다. 끝으로 환경의 경우 “연구/정책”이 35.1%로 가장 큰 비중을 차지하는 실천 방식인 것으로 나타났다. 교차분석 결과 확률적으로 유의미한 수준에서($p < .001$) 문제 영역과 실천 방법 사이에 관련성이 있는 것으로 나타났다.

표 4. 영역과 실천 방법의 분포

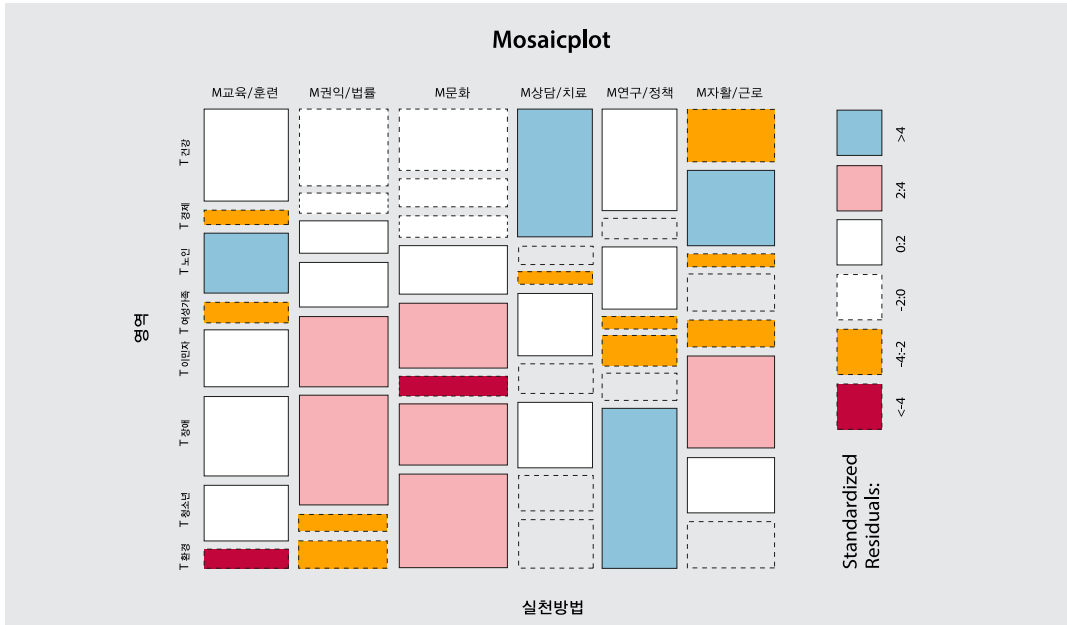
	영역	실천 방법						계
		교육/훈련	권익/법률	문화	상담/치료	연구/정책	자활/근로	
	건강	61(19.9)	52(16.9)	52(16.9)	74(24.1)	33(10.7)	35(11.4)	307(100)
	경제	9(7.0)	13(10.2)	24(18.8)	10(7.8)	20(15.6)	52(40.6)	128(100)
	노인	40(38.1)	22(21.0)	17(16.2)	7(6.7)	11(10.5)	8(7.6)	105(100)
	여성/가족	13(7.1)	32(17.4)	42(22.8)	36(19.6)	36(19.6)	25(13.6)	184(100)
	이민자	38(20.5)	49(26.5)	56(30.3)	17(9.2)	6(3.2)	19(10.3)	185(100)
	장애	54(20.4)	77(29.1)	16(6.0)	38(14.3)	16(6.0)	64(24.2)	265(100)
	청(소)년	37(21.3)	12(6.9)	52(29.9)	20(11.5)	15(8.6)	38(21.8)	174(100)
	환경	12(4.5)	19(7.2)	81(30.6)	28(10.6)	93(35.1)	32(12.1)	265(100)

Chi-square(35) = 432.29, p < .001

그림 2.는 모자이크 그림을 활용해 영역과 실천 방법의 분포를 다시 제시한 것이다. 사각형 형태로 제시된 셀의 크기는 상대적인 비중을 의미하며, 실선으로 그려진 사각형의 색이 진해질수록 예측값에 비해 실제값이 큰 경우를, 반대로 점선으로 그려진 사각형의 색이 진해질수록 예측값이 실제값보다 큰 경우를 의미한다. 따라서, 전자의 경우는 확률적으로 유의미한 수준에서 케이스의 크기가 큰 경우이며, 후자는 그 반대로 이해할 수 있다.

그림에서 제시된 바와 같이 “건강” 영역의 경우 확률적으로 유의미한 수준에서 “상담/치료”의 비중이 크지만, 반대로 “자활/근로”는 유의미하게 낮은 비중을 보였다. “이민자”의 경우 “권익/법률”과 더불어 “문화”도 유의미하게 큰 비중을 차지하는 것을 알 수 있다. 앞서 제시한 교차표에서 “환경”의 경우 “연구/정책”이 가장 큰 비중을 차지했음을 보았는데, “문화” 또한 유의미하게 큰 비중을 차지하고 있음을 알 수 있다. “장애”의 경우 앞서 검토했던 “권익/법률”의 실천 전략에서의 중요성과 더불어 “자활/근로” 또한 중요한 비중을 차지함을 알 수 있다.

그림 2. 영역과 실천 방법 분포에 관한 모자이크 그림



2) 회귀분석 결과

표 5.는 선형혼합효과 모형을 활용한 회귀분석 결과를 제시한 것이다. 세 모형은 각각 기부금 총액, 보조금 총액, 그리고 두 금액의 합에서 기부금이 차지하는 비율을 각각 결과변수로 투입하였다. 분석 결과 제시의 편의를 위해 기부금과 보조금은 모두 만원 단위로 수정하였다. 앞서 진행한 텍스트 마이닝 결과를 활용한 8개 사업 영역과 6개 실천 방법의 개념별 관련성을 모형에 투입하였으며, 기타 변인으로 법인 설립 기간과 연도, 그리고 지역 규모를 함께 다루었다.

분석 결과 기부금과 보조금의 절대 규모에서 사업 성격에 따른 유의미한 차이가 관찰되지 않았다. 이는 기타 변인도 유사했는데, 보조금을 결과변수로 한 모형에서 설립기간은 유의미한 정적 관련성이 있는 것으로 나타났다.

표 5. 회귀분석 결과

구분	변수	DV = 기부금(A)			DV = 보조금(B)			DV = A / (A+B)		
		b	S.E.	t	b	S.E.	t	b	S.E.	t
지역	(Intercept)	83716.5	58667.9	1.43	27841.5	28959.6	0.96	65.86	1.88	34.99
	청소년/청년	12623.9	93302.3	0.14	-37536.8	35166.5	-1.07	12.74	2.26	5.63
	장애	23272.4	97778.3	0.24	15393.5	36827.6	0.42	0.36	2.37	0.15
	건강	-73503.7	97518.2	-0.75	-26596.7	36899.9	-0.72	-4.07	2.38	-1.71
	노인	6640.4	92107.2	0.07	-29192.9	34851.7	-0.84	-2.63	2.24	-1.17
	여성/가족	732.6	78994.9	0.01	-31946.9	29781.8	-1.07	3.26	1.92	1.70
	이민자	-56974.5	78649.7	-0.72	-46620.3	29897.6	-1.56	10.20	1.93	5.30
	경제	47308.0	82082.5	0.58	32386.2	31114.0	1.04	-1.79	2.00	-0.89
환경/지역	12541.9	78081.4	0.16	23050.4	29416.4	0.78	-1.53	1.89	-0.81	
분야	교육	-86521.7	91819.8	-0.94	12911.3	34696.0	0.37	-11.37	2.23	-5.09
	자활	47909.8	83704.6	0.57	50312.2	31801.6	1.58	-3.86	2.05	-1.88
	연구/정책	-35377.7	96454.6	-0.37	31448.6	36395.8	0.86	3.05	2.34	1.30
	권익/법률	37374.0	77742.4	0.48	8189.5	29214.7	0.28	-2.65	1.88	-1.41
	문화	35302.7	72176.3	0.49	-214.8	27097.1	-0.01	3.43	1.74	1.97
	상담/치료	95337.5	91745.1	1.04	1245.2	34589.0	0.04	-2.43	2.23	-1.09
기타	설립기간	1912.5	2837.3	0.67	4963.7	1219.6	4.07	-0.48	0.08	-6.13
	(ref. = 2015)2017	16074.4	18521.2	0.87	-17813.1	20957.1	-0.85	3.46	1.39	2.49
	2019	30790.4	20472.4	1.50	26956.5	20351.3	1.33	-0.27	1.35	-0.20
	(ref. = 대도시) 일반시	-54003.9	61937.0	-0.87	-28857.5	34958.2	-0.83	-18.34	2.26	-8.11
	농어촌	-17140.3	74790.9	-0.23	-47666.1	56030.1	-0.85	-14.54	3.65	-3.99

전반적인 평균 액수에서 사업 내용 간 유의미한 차이는 관찰되지 않았으나, 개별 법인 별로 수입에서 기부금이 차지하는 비중은 사업 내용에 따라 부분적으로 유의미한 차이가 존재하는 것으로 나타났다. “청소년/청년” 부분의 경우 통계적으로 유의미한 수준에서 기부금의 상대적인 비중이 큰 것으로 나타났다. 이는 “이민자”의 경우도 비슷하게 나타났다. 사업 실천 방식에서는 교육의 경우는 유의미한 수준으로 기부금의 상대적인 비중이 낮은 것으로 나타났다. 이는 교육, (직업)훈련 등의 영역에서는 보조금의 비중이 상대적으로 클 수 있음을 의미한다. 한편, “문화” 영역의 경우는 기부금의 상대적인 비중이 유의미하게 큰 것으로 나타났다.

기타 변인에서도 유의미한 관련성이 포착되었다. 비교적 과거에 설립된 법인일수록 보조금의 상대적 비중이 큰 것으로 나타났다. 기간 효과도 부분적으로 관찰되었는데 2015년을 기준으로 했을 때, 2017년 기부금 비중이 상대적으로 높아진 것으로 나타났다. 그러나 2019년에는 유의미한 차이를 관찰하지 못했다. 지역의 효과도 나타났는데, 광역시 이상의 대도시의 경우 기부금의 상대적인 비중이 더 커지는 것을 확인하였다.

4. 소결

본 연구는 텍스트 마이닝 기법을 활용하여 비영리 법인의 사업 내용을 분석하고, 이를 바탕으로 사업 운영에 필요한 수입 규모와 구성에서 어떤 차이가 있는지 살펴보고자 하였다. 이는 두 가치 차원에서 의의가 있는데, 하나는 텍스트 마이닝과 같은 기법을 활용하여 비영리 법인의 성격을 분석하는 방법론적 시도라는 점과 이렇게 추출한 정보와 법인 특성 요인을 결합하는 과정을 통해 질적 데이터와 양적 데이터를 결합하여 다루고자 했다는 점이다. 본 연구의 시도는 비영리 법인의 사업 목적과 실천 방식에서 드러나는 다양성을 실증적으로 다루고, 질적, 양적 데이터가 다양한 형식으로 축적되어 있는 비영리 분야 연구의 범위를 확대함에 기여가 있을 것으로 판단된다.

분석 결과를 통해 도출할 수 있는 탐색적인 시사는 다음과 같다. 먼저, 사업의 목적과 실천 방법 사이에는 일정한 관계가 있음을 확인할 수 있었다. 건강 증진과 관련된 사업은 직접적인 상담이나 치료 프로그램 운영이 많았으며, 환경이나 지역 사회 개선과 관련된 사업은 연구와 정책 개발 등의 실천 내용이 중요한 비중을 차지하고 있음을 확인했다. 이와 같은 결과는 사업 목적과 방법의 관련성을 실증적으로 검토한다는 점에서도 의의가 있지만, 새로운 사업 영역을 도출하거나 새롭게 주목이 필요한 사업 내용을 구상하는 데 도움을 줄 수 있다.

예를 들어, 그림 2.에서 제시된 바와 같이 청소년/청년 분야의 경우 “문화”와 관련된 사업 내용은 유의미한 규모를 차지하지만, “권익/법률”과 관련해서는 유의미하게 구성비중이 낮은 것을 확인하였다. 많은 청년들이 고용의 불안정성, 직장에서의 갑질 문화 등이 존재하는 상황에서 직장내 차별이나 부당대우를 받을 가능성이 크다. 따라서 청년 노동자들의 노동권을 보호하고 필요한 경우 실질적인 법률 서비스를 제공하는 사업의 확대를 생각해 볼 수 있을 것이다.

또 다른 예로서 “이민자”의 경우 “권익/법률”이나 “문화”와 관련해서 유의미한 수준으로 높은 비중을 차지하고 있음을 확인하였다. 이는 많은 경우 외국인 노동자가 경험할 수 있는 부당한 대우에 대한 법률적 보호나, 다문화 인식 개선 등의 사업이 많기 때문으로 해석된다. 그런데 상대적으로 “연

구/정책”과 관련된 영역이 부족함을 확인하였다. 이는 이민자에 대한 체계적인 조사, 학술 활동, 정책 개발과 관련된 사업 영역이 확대될 필요성을 간접적으로 시사하는 결과라고 해석할 수 있다.

다음으로 사업 내용과 수입 구조와의 관련성을 검토한 회귀분석 결과를 통해 다음과 같은 함의를 도출할 수 있다. 우선, 사업 내용에 따라 기부금과 보조금의 규모에서 유의미한 차이는 나타나지 않았다. 그러나 양자의 비율에서는 사업 성격에 따라서 유의미한 차이가 관찰되는 확인했다. 이와 같은 결과는 비영리 법인의 활동 분야 전반에 걸쳐 재정적인 균형이 존재한다는 탐색적인 시사와 함께, 개별적으로 보았을 때는 사업 성격에 따라서 재원의 성격이 달라질 수 있음을 의미하는 것이다.

“청소년/청년” 분야의 경우 기부금의 상대적 비중이 유의미하게 크게 도출되었는데, 이는 해당 분야의 경우 기부금을 통한 자원 마련이 상대적으로 수월한 체계가 수립되어 있음을 뜻한다. 아동, 청소년 분야는 가장 전통적인 기부 분야이며, 어려움을 경험하고 있는 집단에 대한 사회적 공감대가 수월하게 형성되기 때문으로 보인다. 이와 반대로 실천 전략에서 “교육” 영역은 보조금의 비중이 상대적으로 큰 것으로 나타났다. 이는 교육, 훈련 사업 등이 지니는 사회적 투자 성격으로 인하여 기부보다는 정부를 통한 보조금을 통한 운영이 주요를 이루기 때문으로 파악된다. 반대로 “문화”의 경우는 기부의 비중이 상대적으로 높았는데, 인식 개선과 같은 캠페인 활동이나 예술, 스포츠 활동의 경우 기업이나 관심있는 시민들의 자발적 후원이 활발하기 때문으로 보인다.

사업 내용에 따라 수입 구조에서 차이가 생길 수 있다는 본 연구의 결과는 두 가치 측면에서 그 의미를 생각해 볼 수 있다. 첫 번째는 이와 같은 실증 관찰을 바탕으로 비영리 단체의 자원 마련과 사업 성격과의 관련성에 대한 이론 개발에 필요한 시사를 얻을 수 있을 것이다. 비영리 단체의 공익적 활동이 안정적으로 이뤄지기 위해서는 안정적인 자원 확보가 필요하며, 재정적 지속가능성을 높이기 위해 사업의 성격에 따라 어떤 자원 마련 전략이 유용한가를 체계적으로 정리하는 단초가 될 수 있을 것으로 기대한다.

두 번째는 반대로 현재의 수입 구조를 통해 비영리 단체가 지닐 수 있는 취약성에 대한 이해로도

이어질 수 있다는 점이다. 기부금에 대한 의존도가 높은 사업 영역의 경우 기부와 관련된 환경에 따라 재정적 불안정성이 커질 수 있다. 기부와 관련된 환경은 기부 자원 동원에 영향을 미치는 전반적인 경제 상황도 있겠지만, 비영리 단체에 대한 신뢰에 영향을 미치는 사건도 중요한 영향을 미칠 수 있다. 한편, 보조금에 대한 의존도가 상대적으로 큰 영역의 경우는 정부 정책의 핵심 기조와 사업 내용의 비정합성이나 파트너십에서의 위기로 인해 재정적인 어려움에 빠질 수 있다. 아울러 정부에 대한 재정적 의존은 비영리 단체의 독립적인 사업 운영에 방해물로 작용할 수 있을 것이다.

본 연구는 탐색적 수준의 연구로서 여러 측면에서 개선될 필요성이 있다. 우선, 분석 자료와 관련해서 비영리 단체의 사업 내용에 대한 보다 풍부한 텍스트 정보를 활용할 필요성이 있다. 활용한 데이터 내에서 다수의 법인이 사업 내용을 구체적으로 제시하지 않거나 정보 수준이 떨어지는 형식적인 기술에 그치는 경우가 많아 누락되는 경우가 많았다. 또한 분석에 포함된 케이스도 모호한 기술이나 오타 등이 많아서 정확한 개념 측정에서 어려움이 있었다. 따라서 후속 연구에서는 정관에 명시된 사업의 내용이나 홍보 자료 등을 활용하는 노력이 필요할 것으로 보인다.

사업 내용에 대한 개념 측정에서도 개선이 필요할 것으로 판단된다. 본 연구에서는 대상(문제 영역)과 실천 방법으로 사업 내용 분석의 틀을 구성하고 각 차원을 구성하는 핵심 개념과 구성 단어를 연구자의 임의적인 판단에 기대어 설계한 측면이 있다. 따라서 후속 연구에서는 이론적 논의에 기반을 둔 분석틀의 구성이나, 토픽 모델링(topic modeling)이나 키워드 추출(keyword extraction)과 같은 정보 축약 과정을 통해 주요 개념과 핵심어를 구성하여 이를 분석에 활용하는 방법도 생각해 볼 수 있을 것이다.

이상의 개선 과정을 통해 보다 엄밀하게 비영리 법인의 사업 내용을 분석하고 이를 예산 구조와 결합하여 살펴본다면 더욱 정확하고 풍부한 관찰 결과를 도출할 수 있을 것으로 기대한다. 따라서 현재까지 수행한 본 연구의 결과는 탐색적인 수준으로 이해할 필요가 있다. 그리고 더 나아가 예산 활용의 세부적인 사항을 살펴볼 수 있는 결산서 등의 추가적인 데이터가 확보된다면, 사업의 목적과 예산 집행의 정합성을 평가하는 분석도 가능할 수 있을 것으로 생각한다.

참고문헌

- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Hamilton, William L., Jure Leskovec, and Dan Jurafsky. (2016) Diachronic word embeddings reveal statistical laws of semantic change, arXiv preprint arXiv:1605.09096.
- Hesselmann, G. (2018), Applying Linear Mixed Effects Models (LMMs) in within-participant designs with subjective trial-based assessments of awareness—a caveat, *Frontiers in Psychology*, 9, 1–5
- Mikolov, Tomáš, Wen-tau Yih, and Geoffrey Zweig. (2013). Linguistic regularities in continuous space word representations, *Proceedings of the 2013 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies*.

