

토픽 모델링을 활용한 한국의 창업생태계 트렌드 변화 분석

Analysis on Dynamics of Korea Startup Ecosystems Based on Topic Modeling

손희영 (Heeyoung Son) 창업진흥원¹⁾
이명종 (Myungjong Lee) 창업진흥원²⁾
변영조 (Youngjo Byun) 국립한밭대학교³⁾

〈 국문초록 〉

1986년, 한국은 국가발전의 주축인 중소기업 창업지원을 위한 법 제도를 마련하였다. 이를 기반으로 지난 30여년간 창업정책의 수립 및 발전을 거듭하여 매년 100만 개가 넘는 신규 창업기업이 설립되는 역동적인 창업생태계를 구축하였다. 국가의 정책 방향과 사회, 경제, 문화 등의 외부환경 영향, 그리고 창업지원의 역사를 주요 이슈별로 분석하여 도출된 핵심문장 또는 키워드는 시대별 지원의 특징과 국가지원의 중심내용 등을 확인하는 데 매우 유용하다. 본 연구는 한국의 창업생태계 트렌드 변화를 분석하기 위해 1991년부터 2020년 12월까지 30년간의 언론기사에서 ‘창업’, ‘벤처’, ‘스타트업’ 키워드가 포함된 118만여 건을 추출하고 네트워크 분석과 토픽 모델링을 활용하였다. 분석결과, 한국의 창업생태계 트렌드는 기업 및 산업육성, 확산 그리고 규제 완화, 활황 등, 정부 중심으로 스타트업 생태계의 변화와 발전이 이루어졌음을 파악할 수 있었으며, 다빈도 키워드 분석결과, 생태계 구성요인 간의 연계 활동을 통하여 기업가적인 생산성이 창출되었다. 생산성 창출의 주요 요인으로 한국은 대기업의 휴대폰 산업 발전과 이와 관련된 콘텐츠 스타트업의 성장, 인터넷과 쇼핑물 중심의 플랫폼 기업의 발전, 그리고 청년창업과 글로벌 진출, 모바일과 인터넷 인프라 중심의 창업 기업육성 노력 등으로 파악할 수 있었다. 본 연구는 30년간의 언론기사를 텍스트마이닝과 토픽 모델링을 활용하여 트렌드를 도출하였다. 이는 선행연구가 기존 정부와 정책의 변경 시기를 기준으로 트렌드 변화를 분석한 것과 달리, 언론기사의 키워드와 토픽 변화를 기준으로 창업생태계의 트렌드 변화를 분석하였다는 점에서 학술적 의의뿐만 아니라, 30년간의 창업생태계 변화 및 주요이슈를 조명해 봄으로써 향후 창업지원의 방향성을 예측할 수 있는 실무적 시사점을 제공하였다.

주제어: 창업생태계, 창업 트렌드, 스타트업 지원, 텍스트마이닝 분석, 한국창업발전사

1) 제1저자, sonhy@kised.or.kr

2) 제2저자, mjlee@kised.or.kr

3) 교신저자, yjbyun@hanbat.ac.kr

1. 서론

한국은 1986년 「중소기업창업 지원법」 제정으로 창업정책의 제도적 기반이 마련되었으며(김만수, 강재원, 2021), 1991년 「창업기업보육센터의 설립 및 업무운용 준칙」의 고시를 통해 창업보육센터 설립 근거를 마련한 뒤, 1994년 현재의 중소벤처기업진흥공단이 운영한 안산창업보육센터를 시작으로 본격화 되었다(황보윤, 김재형, 2016). 1990년대 말 외환위기로 경제구조조정이 추진되며, 이를 극복하기 위한 수단으로 벤처 활성화를 위한 종합대책을 마련하고, 1997년 「벤처기업육성에 관한 특별조치법」을 제정하였다(김만수, 강재원, 2021). 이를 계기로 활성화된 벤처 붐은 2000년대에 들어서며 소위 ‘벤처버블의 붕괴’에 직면하게 되고, 벤처기업의 질적 내실화를 위한 정책이 추진되었다(하규수 등, 2016). 이후 우리 정부는 하드웨어적 인프라를 확충하고, 자금지원을 확대함과 동시에 기업가정신 확산을 통해 사회 전반에 걸쳐 창업문화를 조성하고, 세분된 창업지원사업을 추진하며 혁신적인 창업생태계를 조성하고자 다양한 정책을 추진하였다(하규수 등, 2016).

그 결과, 현재 한국의 창업생태계는 괄목할만한 성장을 이루었다. 2021년 기준, 부동산업을 제외한 실질 창업기업 수는 전년대비 5.1% 증가한 1,100,589개를 기록했으며, 특히 기술창업은 정보통신업, 전문과학기술업 등이 큰 폭으로 증가하며 역대 최대인 23만 개를 달성하였다(중소벤처기업부, 2022). 이러한 기술창업 증가와 더불어 한국 유니콘 기업은 11개(2021.12.31. 기준, CB Insight)가 창출되며 한국의 창업기업이 글로벌 스타트업으로 성장하는 성공사례를 만들어가고 있다. 그뿐만 아니라, GERA(Global Entrepreneurship Research Association)의 연차별 Global Entrepreneurship Monitor (GEM) Global Report(2019-2021)에 따르면, 한국의 기

업가정신 종합지수는 2019년 5.13으로 참여국 54개국 중 15위에서 2020년 5.49로 44개국 중 9위, 2021년에는 5.7로 50개국 중 6위를 차지하며, 매년 한국의 기업가정신 생태계가 발전하고 있는 것을 확인할 수 있다.

이처럼 지속해서 성장해온 한국의 창업생태계 변화와 관련하여 한국에서는 다양한 접근을 통해 연구가 진행됐지만, 체계적이고 종합적으로 창업정책 변화를 분석한 연구는 제한적이다(김만수, 강재원, 2021). 또한, 대부분의 연구가 정책의 관점에서 특정 기간을 대상으로 단편적 특징을 연구하는 데에 그치고 있다. 이에 본 연구는 1991년부터 2020년까지 총 30년간의 언론 기사를 분석하여, 한국 창업생태계의 변화를 고찰하고자 한다. 이는 시대를 비추고 여론을 대변하는 언론 기사를 통해 창업생태계의 다양한 변화를 파악하고, 이를 기초로 정책적 관점이 아닌 시대적 관점과 국민적 인식을 토대로 창업생태계의 트렌드 변화를 고찰한다는 데에 그 의의가 있다. 또한, 단순히 연대별, 정부별로 구분해 온 생태계 발전사에서 벗어나 시대적 변화에 기반을 두어 발전사를 재정립한다는 측면에서 시사점을 제공하고자 한다.

2. 이론적 배경

2.1. 창업생태계와 구성요소

창업생태계 혹은 기업가적 생태계의 개념과 관련하여 많은 연구가 진행되고 있다. 학술적 관심뿐만 아니라 보다 총체적이고 역동적인 접근 방식이 시도되고 있으며(Mason, 2019) 기업가와 창업생태계가 상호간 연계성이 있다는 시각으로 자리 잡고 있다(Spigel, 2017; Cavallo et al., 2018). 또한, 기업가적 활동을 결정하는 지역적 요소의 상호작용 결과로서 신생기업의

〈표 1〉 창업생태와 구성요소에 관한 주요연구

연구자	구성요소	주요연구 내용
Stam and Van de van(2019)	공식적 비공식적인 제도, 사회적 네트워크, 유형자원, 재무자원, 리더십, 인적자원, 지식, 소비, 제조(전환)서비스, 생산적인 기업가정신	게임의 법칙, 문화적 맥락, 사회적 연계 정도, 생태계 내 활동가 간의 물리적 연계, 투자와 관련된 활동, 리더십, 개인으로부터 획득 가능한 지식, 재화와 서비스를 구매할 수 있는 능력 등이 연계되어 기업가적 활동으로 전환되어 부가적 성과를 창출해 내는 것
Spigel(2017)	1) 문화: 문화적 속성, 기업가정신의 역사 2) 사회: 네트워크, 자본투자, 멘토와 협상가, 재능 3) 재료: 대학, 지원시스템, 물리적 인프라, 정치와 지배구조, 시장 등	기업가적 생태계의 속성과 특징을 지역별 사례연구를 통하여 관계성 규명
Stam(2015)	1) 시스템적 요건 : 네트워크, 리더십, 재무, 재능, 지식, 지원서비스 2) 기반구조적 조건: 공식적인 제도, 문화, 물리적 인프라스트럭처, 요구 등	창업생태계 문헌연구를 통하여 인과적 깊이와 증거기반의 연구가 제한적이라는 단점을 제시하고 특정 기업가적 활동을 통한 생태계 내의 새로운 가치창출로 연계되는 체계 및 프레임워크를 제안
Mason and Brown(2014)	기업, 벤처투자자, 엔젤투자자, 은행, 대학, 공공기관, 창업과정(창업기업 수, 스타기업 수준, 연속창업자 수 등)	창업활동을 영위하는 자, 기업가적 조직, 기관, 기업가적 과정 등의 구성요인에서 발생하는 성과를 연결, 확장 및 관리하기 위해 생태계 내에서 공식 또는 비공식적 연합
Feld(2012)	기업가, 정부, 대학, 투자자, 멘토, 서비스제공자, 대기업	생태계 구성요소를 창업기업 커뮤니티의 참여자로 표현하고 리더, 공급자, 그리고 선동자의 중요성을 설명, 생태계 내의 문제점과 역할, 액셀러레이터와 대학의 참여에 대한 사례 위주의 연구
Isenberg(2010)	리더십, 문화, 자본시장, 개방적 고객 등 복잡한 방식으로 결합된 개별요인	해당 요인의 상호 연결된 특성을 통하여 결과를 창출, 해당 특성을 무시할 경우 왜곡된 결과 초래 가능성
서리빈 등(2020)	제도, 자본, 문화, 지원서비스, 인프라, 인재, 교육, 시장, 조직, 기업가	창업생태요인을 투입, 산출, 결과, 성과 등으로 구분하여 고용, 기술사업화, 시장 효율성을 창업생태계의 결과요인으로, 사회경제적 부가가치 및 후생증가를 성과요인으로 정의
변영조 등(2020)	참여주체, 자원, 인프라, 사업계획, 혁신아이템, 지역사회연계 프로그램, 지역인재양성 프로그램 등	산학연관의 창업 성과 활성화를 위해 생태계 내에서의 구성요인을 투입자원-혁신 활동-실적-성가로 구분하여 성과측정지표 및 정의

설립, 즉 창업이 촉진된다는 관점에 입각하여(서리빈 등, 2020) 공식적 또는 비공식적으로 상호 연결된 기업가적인 활동가, 조직, 기관 또는 기업가적 과정의 집합으로 정의할 수 있다.(Brown & Mason, 2017; Shwitzer et al., 2019). 창업생태계를 기업가적 환경 내에서 창업 과정을 상호 연결, 조정 및 관리하는 기업가적 주체, 조직 및 기관의 집합으로 정의한 Mason and Brown(2014)은 생태계를 구성하는 요소 간 상호작용의 결과로 지역 내 창업 성과가 증진된다고 설명하였다.

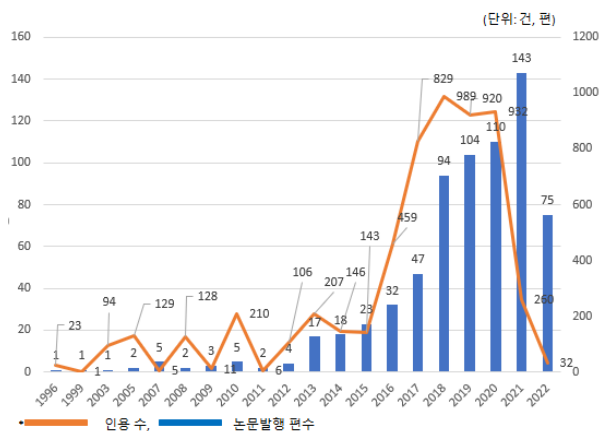
창업생태계의 구성요소에 관한 연구도 다양하게 진

행되어 오고 있다. 자원, 리더십, 제도, 서비스 등의 창업기업을 중심으로 한 내·외부 제도 및 시스템접근을 통한 요인 간의 연계성이 성과를 창출한다는 연구(Stam, 2015; Stam & Van de Van, 2019; 서리빈 등, 2020)와 요인 또는 속성 등에 집중하여 생태계 구성요인을 문화, 사회, 재료 등으로 구성하고 사례를 통해 요인도출과 관계성을 입증하는 연구(Isenberg, 2010; Spigel, 2017), 그리고 생태계 구성요인 자체를 생태계 내에서의 실질적인 활동을 영위하는 기업, 투자자, 금융기관, 대학, 공공기관, 대기업 등으로 제시한 연구

(Feld, 2012; Mason & Brown, 2014; 변영조 등, 2020) 등으로 분류할 수 있다. 선행연구의 공통점은 창업생태계 내에서 요인 간의 연계 활동 또는 시스템적으로 투입-전환-산출의 과정을 통하여 성과를 창출한다는 것이다.

2.2. 창업생태계의 연구동향

창업생태계의 선행연구는 주로 창업생태계의 정의, 성과창출을 위한 구조와 시스템적 접근, 그리고 구성요인 자체에 대한 정의 및 이들 간의 관계를 통한 기업가적 성과들에 대하여 다루고 있음을 알 수 있었다. 관련 연구 동향을 파악하기 위하여 SCOPUS 데이터베이스를 이용하여 ‘Entrepreneurial Ecosystem’을 탐색 단어로 한 발표논문 조사결과, 1996년에 최초로 1편이 출판되었고 2012년까지 5편 이내로 출간되다가 2013년 17편을 시작으로 급속하게 증가하여 2019년 이후에는 매년 100편이 넘는 창업생태계에 관한 연구가 발표되고 있다. 인용 또한 논문의 수가 증가한 2013년 이후 2015년부터 가파르게 상승함을 볼 수 있었다.



〈그림 1〉 창업생태계 논문 발행 및 인용수

연구에서 다른 주요 주제어의 파악을 위해 검색한

총 689편의 논문 제목과 논문 내 키워드를 보이언트(voyant-tools.org)를 사용하여 빈도 분석한 결과, 논문 제목 8,546개의 단어 중 상위 5위 내에 속한 주요 주제어는 ‘ecosystem, startup(s), innovation, entrepreneurial, ecosystems’였으며, 이후 30위 내에서 ‘창업, 생태계’, ‘혁신’ 등의 제목과 관련된 주제어를 제외한 기타 용어로 ‘business, technology, software, digital, international, accelerators’ 등을 확인할 수 있었다. 그리고 논문 내에서 제시된 키워드를 분석한 결과 전체 3,384개의 단어 중 ‘development, ecosystem, entrepreneur, economic, entrepreneurial’이 상위 5위 내 추출 단어였으며, 이후 30위 내에서 ‘창업, 생태계, 혁신’ 등의 제목과 관련된 주제어를 제외한 기타 용어로 ‘education, policy, engineering, sustainable, university, government’가 등장하는 것이 확인되었다.

〈표 2〉 주요주제어의 워드 클라우드 분석결과 상위 빈출 용어

(단위: 개)

순위	논문제목 내(빈도)	키워드 내(빈도)
1	ecosystem(163)	development(105)
2	startup(144)	ecosystems(97)
3	startups(129)	entrepreneurs(81)
4	innovation(106)	economic(67)
5	entrepreneurial(102)	entrepreneurial(63)
6	ecosystem(88)	innovation(58)
7	entrepreneurship(77)	regional(53)
8	business(69)	entrepreneurship(51)
9	study(60)	ecosystem(47)
10	technology(55)	analysis(45)
11	case(53)	education(42)
12	software(47)	business(39)
13	digital(45)	policy(33)
14	development(45)	social(32)
15	start(39)	research(30)
16	analysis(36)	engineering(30)
17	role(31)	industrial(29)
18	new(29)	technology(28)

순위	논문제목 내(빈도)	키워드 내(빈도)
19	conference(28)	sustainable(27)
20	based(28)	economics(27)
21	venture(27)	knowledge(26)
22	fintech(27)	university(25)
23	emerging(27)	framework(24)
24	model(26)	management(22)
25	social(24)	approach(22)
26	international(24)	ecology(20)
27	accelerators(24)	planning(18)
28	industry(22)	government(18)
29	entrepreneurs(21)	information(17)
30	open(19)	growth(17)

이상의 내용에서와 같이 연구의 주요 내용은 창업 생태계 내에서의 ‘스타트업, 혁신, 기업, 기술, 정치, 경제, 사회, 지속가능성과 대학, 정부’ 등 주요 선행연구에서 제시하였던 구성요인의 빈도가 높게 검출되었다.

3. 연구방법

3.1. 자료수집

본 연구의 분석을 위해 언론진흥재단 뉴스빅데이터 분석 서비스 ‘빅카인즈’ 내에 포함된 50개 언론사를 대상으로 1991년 1월 1일부터 2020년 12월 31일까지 30년간의 언론기사 중 ‘창업’, ‘벤처’, ‘스타트업’ 키워드가 포함된 언론 기사를 추출하였으며, 중복을 제외한 총 1,187,953건의 기사가 추출되었다. 그리고 본 연구의 목적에 부합하도록 ‘소상공인’, ‘자영업자’의 유사 키워드가 포함된 79,807건을 제외하여 총 1,108,146건의 언론 기사를 최종 분석데이터로 확정하여 <표 3>과 같이 정리하였다.

이후 정제된 키워드를 활용한 빅데이터 분석을 위해 한글과 띄어쓰기를 제외한 키워드(숫자, 영문 등)

<표 3> 분석대상 언론기사의 연도별 분포현황

(단위: 건)

연도	언론기사 수		
	전체	제외기사	분석대상
1991	820	10	810
1992	1,841	21	1,820
1993	2,049	30	2,019
1994	2,889	37	2,852
1995	3,907	45	3,862
1996	4,974	57	4,917
1997	11,186	238	10,948
1998	14,753	344	14,409
1999	22,451	918	21,533
2000	52,172	990	51,182
2001	45,303	1,355	43,948
2002	35,733	921	34,812
2003	25,557	1,008	24,549
2004	24,953	1,012	23,941
2005	26,149	1,482	24,667
2006	21,059	1,041	20,018
2007	20,066	955	19,111
2008	27,113	1,784	25,329
2009	33,655	3,188	30,467
2010	42,352	3,921	38,431
2011	47,067	3,599	43,468
2012	48,563	5,009	43,554
2013	60,529	4,445	56,084
2014	59,180	4,252	54,928
2015	76,951	4,557	72,394
2016	78,864	4,405	74,459
2017	87,542	6,960	80,582
2018	100,927	8,413	92,514
2019	110,888	8,951	101,937
소계	1,187,953	79,807	1,108,146

및 '인물', '위치', '기관'을 나타내는 고유명사를 불용어로 처리하고, 오·탈자 및 동의어, 유사어에 대한 병합예를 들면 ‘창업자’, ‘창업가’, ‘창업가들’의 명사와 조사가 같이 있는 경우 ‘창업자’로 병합을 실시하였다.

3.2. 분석방법

본 연구는 한국 창업생태계의 트렌드 변화를 분석하기 위해 네트워크 분석(Network Analysis)과 토픽 모델링(Topic Modeling)을 활용하였다. 키워드 네트워크 분석은 사회연결망이라고도 하며 사람이나 조직 등의 객체 간의 관계를 계량화해서 표현(주재홍, 송지훈, 2021)할 수 있으며 텍스트의 내용을 분석하기 위해 키워드의 관계를 토대로 네트워크를 구성하고, 이를 시각화하는 분석방법으로 다양한 영역에서 활용되고 있다(현기순, 2022). 또한, 토픽 모델링은 비구조화된 텍스트에서 새로운 정보를 추출하기 위한 연구에 주로 활용되는 텍스트 마이닝 기법으로, 특정 분야의 주제 및 동향 파악에 효과적인 수단으로 많이 활용되고 있다(박종도, 2019; 이소현 등, 2019; 이금실 등, 2018). 텍스트마이닝이란 비정형화된 텍스트 데이터를 처리하여 의미 있는 패턴을 추출하거나 그 텍스트에서 반영하고 있는 의미를 수치화한 뒤 그 값을 추출하는 과정이다(박명석 등, 2018).

네트워크 분석에서는 먼저, 단어 관계 행렬을 구축하기 위해 기사 내 빈출 단어를 기반으로 문서-단어 행렬(Document-by-Term matrix)을 구축하였다. 이를 위해, 문서 내에 빈출이 높으나 타 문서에서는 빈출이 적은 단어에 높은 가중치를 부여하는 방식으로, 문서 내 단어의 상대적 중요도를 파악하고 중요도가 낮은 무의미한 단어 및 범용단어에 대해 적은 가중치로 보정이 가능해 문서-단어 행렬 구축에 적합한 TF-IDF (Term Frequency-Inverse Document Frequency)를 활용하였다(Qaiser-Ali, 2018). 또한, 유의미한 키워드 네트워크의 구성을 위해 40% 이상의 문서에서 빈출된 보편적인 단어와 1% 이하의 문서에서 빈출된 극단적 단어를 제거하고 연산을 수행하였다.

$$W_{x,y} = tf_{x,y} \times \log \left(\frac{N}{df_x} \right)$$

tf_x : 문서(기사) 내에 빈출하는 단어의 횟수

df_x : 단어를 포함하는 문서의 개수

N : 전체 문서의 개수

다음으로 단어 간 관계를 나타내는 유사도 행렬 구축을 위해 코사인 유사도(Cosine Similarity)를 활용하여 단어-단어 간 유사도를 측정하였다.

$$\begin{aligned} \text{similarity} = \cos(\theta) &= \frac{A \cdot B}{\|A\| \|B\|} \\ &= \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \end{aligned}$$

이후, 단어 관계 행렬을 입력 값으로 하여, 네트워크 시각화 소프트웨어(Gephi)를 활용하여 네트워크 구성하였다. edge weight를 0.1~0.15 사이에서 필터링한 후, node degree가 0~2 사이 값에 대해 필터링하여 node(단어)를 제거하였으며, 반복 분석을 통해 연도별 단어 제거를 최소화하는 필터링 파라미터를 추출하고, 모듈 클래스(modularity)별로 분리하여 시각화하였다. 배열구성(layout)은 Gephi에서 제공하는 것 중, 상대적으로 네트워크 수렴 속도 및 방향이 안정적으로 알려진 Force Atlas 2를 주로 활용하였다.

토픽 모델링은 토픽별 단어의 분포 및 문서별 토픽의 분포를 추정하는 확률 모형으로, 유사한 의미의 단어들을 그룹화하여 잠재된 주제를 추론하는 가장 대표적인 토픽 모델링 방법론인 잠재 디리클레 할당 즉, LDA(Latent Dirichlet Allocation) 알고리즘을 활용하였다(De Battisti et al., 2015; 현기순, 2022).

토픽 모델링을 수행하기에 앞서, 30년간의 이슈를 시기별로 클러스터링하기 위해 연도별 키워드를 대상으로 직전 연도와의 중복비율을 비교하였다. 이는 분

석주제인 창업생태계의 특성상 주요 이슈와 트렌드가 여러 연도에 걸쳐 나타나는 경향을 보이기 때문에, 이를 정량자료에 기초해 정의하기 위함이다. 이를 위해, 직전년도 주요 키워드와의 중복비율을 산출하여 비율이 높은 연도를 병합하고, 병합한 기간에 할당된 주제

별 의미론적 일관성을 의미하는 Coherence score를 0.5 이상이 되도록 반복 분석하여 최적의 토픽 수를 정의하고자 하였다.

직전년도 주요 키워드 중복비율을 산출한 결과는 <표 4>와 같았으며, 전년도와의 중복비율이 현격히

<표 4> 30년간의 직전연도와 비교한 주요키워드 중복비율

(단위: %)

연도	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1991	100.0	58.8	51.3	44.2	44.7	32.7	28.1	19.6	13.1	11.1	12.6	12.6	12.6	18.6	15.1
1992	49.8	100.0	56.6	46.4	43.8	34.5	31.9	22.1	13.2	11.9	13.6	14.0	13.2	20.0	17.9
1993	57.3	74.7	100.0	61.8	57.9	43.8	42.7	27.5	16.9	15.2	16.3	18.0	16.9	23.0	21.3
1994	54.3	67.3	67.9	100.0	69.8	56.2	49.4	33.3	23.5	19.8	21.0	21.0	20.4	24.7	22.8
1995	53.3	61.7	61.7	67.7	100.0	61.1	51.5	34.1	24.6	24.0	25.7	26.3	23.4	29.3	27.5
1996	47.4	59.1	56.9	66.4	74.5	100.0	75.9	50.4	39.4	35.0	36.5	35.8	35.8	40.9	36.5
1997	37.8	50.7	51.4	54.1	58.1	70.3	100.0	55.4	43.2	39.9	37.8	37.2	33.1	39.9	36.5
1998	36.8	49.1	46.2	50.9	53.8	65.1	77.4	100.0	62.3	51.9	50.0	50.0	45.3	55.7	48.1
1999	26.3	31.3	30.3	38.4	41.4	54.5	64.6	66.7	100.0	73.7	61.6	61.6	53.5	60.6	53.5
2000	20.6	26.2	25.2	29.9	37.4	44.9	55.1	51.4	68.2	100.0	75.7	70.1	57.0	60.7	58.9
2001	25.3	32.3	29.3	34.3	43.4	50.5	56.6	53.5	61.6	81.8	100.0	83.8	66.7	69.7	65.7
2002	24.3	32.0	31.1	33.0	42.7	47.6	53.4	51.5	59.2	72.8	80.6	100.0	68.9	72.8	69.9
2003	26.9	33.3	32.3	35.5	41.9	52.7	52.7	51.6	57.0	65.6	71.0	76.3	100.0	89.2	86.0
2004	29.4	37.3	32.5	31.7	38.9	44.4	46.8	46.8	47.6	51.6	54.8	59.5	65.9	100.0	77.8
2005	24.0	33.6	30.4	29.6	36.8	40.0	43.2	40.8	42.4	50.4	52.0	57.6	64.0	78.4	100.0
연도	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
2006	100.0	85.5	82.4	72.5	72.5	61.8	65.6	61.8	67.2	70.2	62.6	59.5	60.3	61.8	61.8
2007	81.8	100.0	79.6	70.1	73.7	62.8	66.4	64.2	66.4	69.3	63.5	60.6	62.8	63.5	63.5
2008	75.5	76.2	100.0	72.7	72.7	60.8	65.7	63.6	64.3	65.0	63.6	61.5	61.5	60.8	60.8
2009	73.1	73.8	80.0	100.0	83.1	70.0	71.5	64.6	68.5	71.5	65.4	61.5	64.6	62.3	63.8
2010	59.7	63.5	65.4	67.9	100.0	74.2	75.5	69.8	71.1	71.7	68.6	65.4	66.0	64.2	64.8
2011	55.5	58.9	59.6	62.3	80.8	100.0	87.7	77.4	77.4	78.1	72.6	69.9	69.9	69.9	69.9
2012	53.1	56.2	58.0	57.4	74.1	79.0	100.0	82.1	83.3	80.9	75.9	72.2	69.1	69.1	66.0
2013	47.4	51.5	53.2	49.1	64.9	66.1	77.8	100.0	86.0	83.0	76.6	73.1	71.3	71.9	67.3
2014	47.1	48.7	49.2	47.6	60.4	60.4	72.2	78.6	100.0	91.4	81.8	77.0	75.9	75.4	71.1
2015	44.7	46.1	45.1	45.1	55.3	55.3	63.6	68.9	83.0	100.0	84.0	76.7	76.2	77.7	74.8
2016	40.6	43.1	45.0	42.1	54.0	52.5	60.9	64.9	75.7	85.6	100.0	86.6	87.1	86.1	83.2
2017	38.8	41.3	43.8	39.8	51.7	50.7	58.2	62.2	71.6	78.6	87.1	100.0	89.1	88.1	83.6
2018	36.2	39.4	40.4	38.5	48.2	46.8	51.4	56.0	65.1	72.0	80.7	82.1	100.0	93.6	88.5
2019	34.3	36.9	36.9	34.3	43.2	43.2	47.5	52.1	59.7	67.8	73.7	75.0	86.4	100.0	86.0
2020	31.9	34.3	34.3	32.7	40.6	40.2	42.1	45.3	52.4	60.6	66.1	66.1	76.0	79.9	100.0

낮아지는 연도를 중심으로 구간을 나누고, 이를 토대로 연도별 구간을 설정하여 Coherence score를 계산하는 방법을 반복 수행하였다. 즉, 이전 연도들과 이슈가 공유되는 현상이 상대적으로 급격히 줄어드는 연도를 기준점으로 하여 구간을 설정하고, 그 구간 내에 의미론적 일관성이 담보되는 구간을 추출하고자 한 것이다. 이는 본 연구가 다루는 자료가 텍스트임을 고려할 때, 군집분석(cluster analysis)을 통한 구간화가 불가능하며, 중복비율의 정도에 따라 구간을 구분하는 객관적 기준이 정의되지 않은 실정으로, 탐색적 구간화를 위해 가장 현실적인 접근방법이라 할 수 있으며, 이에 대한 객관성을 확보하기 위해 키워드 중복비율과 Coherence score의 2가지 지표를 모두 충족하는 방법으로 시기를 구분하고자 한 것이다.

또한, 최종 토픽 수의 선정은 <표 5>와 같이 Coherence score가 0.5 이상이면, 가장 적은 토픽 수를 갖는 지

점을 최적의 토픽 수로 정의하여 추출하였으며, 모든 토픽 수에 대해서 Coherence score가 0.5를 넘지 못할 경우, Coherence score가 가장 높은 토픽 수를 선정하여 추출하였다.

이러한 반복 분석과정을 통해 분석대상 기간인 30년을 아래의 <표 6>과 같이 총 10개의 기간으로 병합하였다.

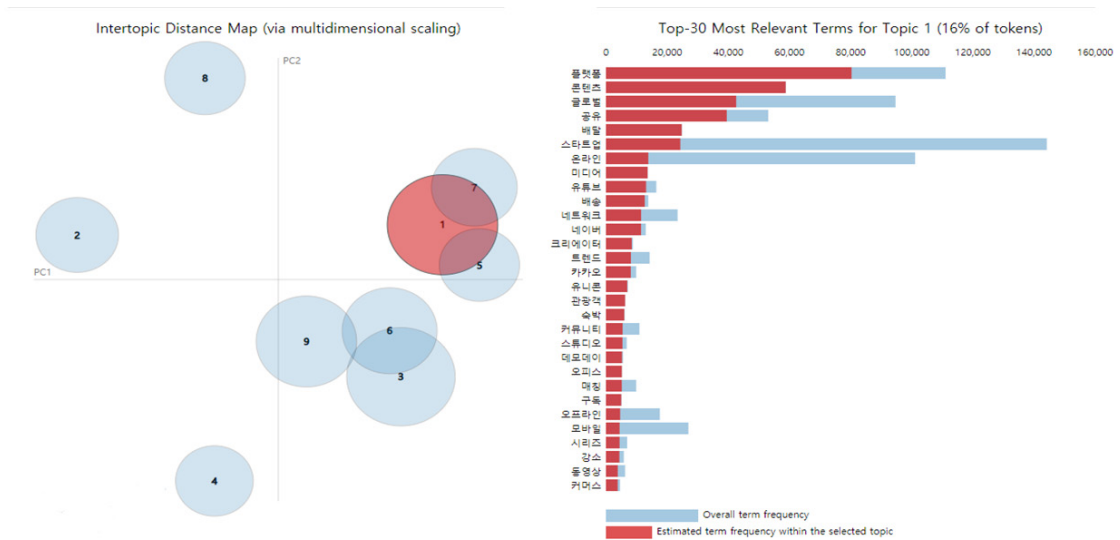
이후, LDA 모델을 활용하여 토픽별 유사도(거리) 지도와 구성 키워드 분포를 추출하고 Interactive Plot을 통해 시각화하였으며, <그림 2>와 같이 토픽 간의 거리지도(Intertopic Distance Map)를 통해 토픽의 수와 토픽 간 거리를 확인하고, 상위 30위의 토픽 관련 용어 그래프(Top-30 Most Relevant Terms for Topic)를 통해 개별 토픽에 포함된 키워드를 확인할 수 있도록 하였다. 그리고 시각화한 토픽 모델링을 기반으로 분석 결과의 주요 키워드 및 시기별 이슈를 고려하여, 정성 분석을 통해 토픽을 정의하였다.

<표 5> 시기별 'Coherence score > 0.5'의 추출결과



<표 6> 분석자료의 기간 병합결과

구분	기 간	구분	기 간
1	1991-1995	6	2009-2010
2	1996-1998	7	2011-2012
3	1999-2002	8	2013-2015
4	2003-2005	9	2016-2017
5	2006-2008	10	2018-2020



〈그림 2〉 토픽 모델링 시각화 예시화면

4. 창업생태계 트렌드 분석결과

앞서 제시한 바와 같이, 본 연구는 언론기사를 토대로 한국 창업생태계의 트렌드 변화를 탐색적으로 분석하고자 하였다. 이를 위해 30년간의 언론기사 데이터를 기초로 토픽 모델링을 실시하여 10개의 기간으로 구분하였으며, 기간별 이슈를 분석하였다. 또한, 연도별 키워드 네트워크 분석을 통해 기간별 주요 이슈와 트렌드를 도출하였다.

4.1. 분석기간 1기(1991-1995년)

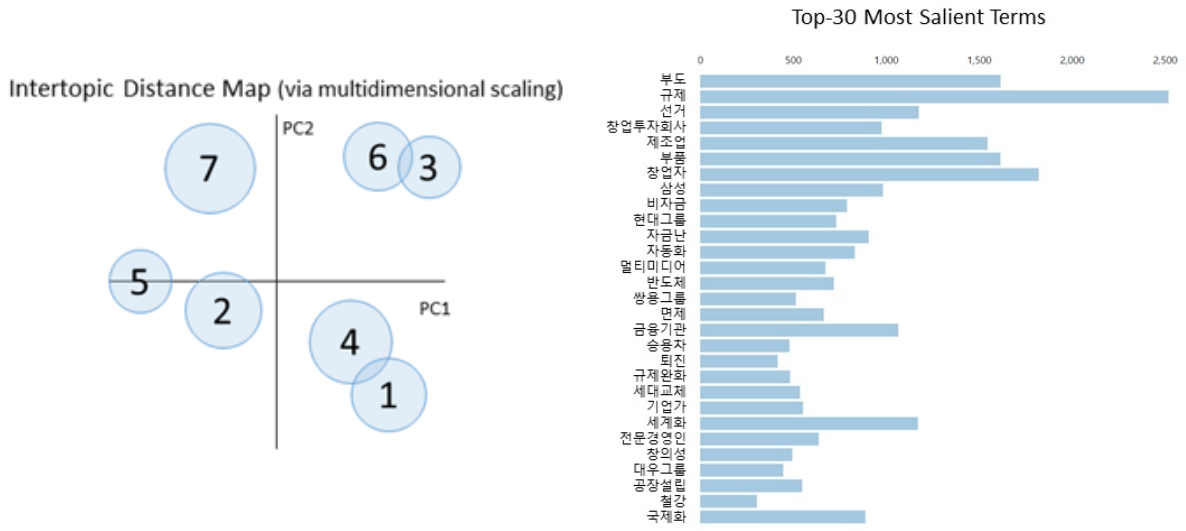
1991에서 1995년은 1990년 동서독 통일 및 1991년 소비에트 연방 붕괴 이후 냉전체제 종식으로 세계질서가 미국 주도로 재편되었으며, 1994년 우루과이 라운드 타결 및 1995년 세계무역기구(WTO) 출범 등 세계 경제장벽 완화를 통해 금융·경제 영역에서의 활발한 세계화 이루어진 시기였다. 한국은 1994년 신한국 창조를 목표로 세계화를 선언하였으며, 1인당 GDP 1

만 달러 돌파와 1995년 국내총생산 세계 11위를 차지하는 성과를 거두었다. 한편, 창업생태계는 1986년 「중소기업창업법」 제정 이후 처음으로 창업보육정책이 추진되었으며, 1995년 창업지원기금이 중기진흥기금으로 통합되었다.

토픽 분석결과, <그림 3>과 같이 총 7개의 트렌드가 도출되었다.

먼저, 삼성 반도체의 세계화를 시작으로 멀티미디어에 관한 관심이 증가하며, 관련 기기 생산 및 소프트웨어 개발 분야에서의 창업이 확대되는 ‘멀티미디어 중심의 벤처 확산’ 트렌드가 나타났다. 또한, 개도국의 부상과 인건비 상승 및 인력난으로 섬유산업이 축소되며, 소재·부품 산업을 육성하기 위해 수입 선 다변화 및 국산화를 추진하며, ‘섬유산업 축소 및 소재·부품산업 육성’이라는 산업적 트렌드가 확인되었다.

다음으로 제조기업의 자동화 투자 확대에 의해 부채비율이 상승한 시기에 정부의 기업 및 산업 구조조정 촉진으로 자금난이 가중되며, ‘산업 구조조정으로 인한 부도 증가’가 하나의 트렌드로 확인되었다. 반면,



〈그림 3〉 토픽 모델링 결과(1991~1995)

경제성장 둔화로 정부가 노사관계 안정을 위해 불법 파업을 규제함에 따라 ‘정부 규제로 인한 노사분규 격감’의 트렌드가 나타났으며, 1993년 「국토이용관리법」 개정을 통해 개별입지의 공장설립을 촉진함으로써 ‘제조업 중심의 고속성장을 위한 규제완화’라는 트렌드를 나타내며, 정부의 산업 관련 정책·규제 이슈가 다양한 트렌드를 양산한 것을 확인할 수 있었다.

이 외에 대우그룹 등이 노태우 전 대통령 비자금 사건에 연루되며 정경유착이 사회문제로 대두되었으며, 이를 통해 ‘대기업의 비자금 문제 확산 및 세대교체’의 트렌드를 양산하였다. 그리고 마지막으로 실리콘 밸리를 중심으로 한 글로벌 벤처 열풍과 함께 벤처기업 ‘메디슨’의 1995년 코스피 상장으로 인해 창업투자 회사의 입지가 미미했던 창업생태계에 벤처투자에 관한 관심이 촉발되며, ‘벤처투자의 시작과 창업투자회사’에 대한 새로운 트렌드를 일으키기 시작하였다.

4.2. 분석기간 2기(1996-1998년)

1997년 아시아 금융위기로 세계 경제 불안을 일으

켰으며, 1998년 OECD를 중심으로 기업지배구조 가이드라인 채택을 위한 논의가 개시된 시기로, 한국은 1996년 OECD 가입을 계기로 외형적 경제발전을 이루었으나, 외환위기(1997-1998년)로 국가 부도의 위험을 겪었다. 한편, 창업생태계 측면에서는 1997년 벤처기업창업활성화 종합대책이 마련되었으며, 특히 세계 최초로 벤처기업 육성을 위한 특별법인 「벤처기업육성에 관한 특별조치법」이 제정되며 벤처 활성화를 위한 정책지원이 추진되기 시작한 시기이다.

토픽 분석결과, 1996에서 1998년에는 <그림 4>와 같이 3개의 트렌드가 도출되었다. 먼저, 앞서 제시한 ‘벤처기업법 제정’이 하나의 트렌드로 확인되었으며, IMF 여파로 인한 대규모 경제 구조조정으로 대기업, 은행 등이 연이어 도산하는 ‘외환위기로 인한 부도·구조조정 가속’이 트렌드로 확인되었다. 이와 달리, 다른 하나의 트렌드는 ‘멀티미디어, 게임산업 발전’으로, 인터넷 시대의 개막과 함께 멀티미디어, 인터넷 콘텐츠에 관한 관심이 확대됨과 동시에 ‘스타크래프트’의 등장으로 네트워크 게임산업이 발전하는 트렌드를 양산하였다.

전하는 인터넷을 중심으로 벤처투자가 확대됐지만, 닷컴버블 붕괴와 함께 벤처시장에서도 버블이 붕괴되며 ‘벤처투자 확대와 버블 붕괴’라는 상반된 트렌드가 동시에 확인되었다. 반면, 정보통신 분야가 발전하며, 컴퓨터, 휴대폰, 네트워크 등 ‘하이테크·신기술 상용화’라는 긍정적 트렌드를 양산하였으며, IT 제품의 발전과 오픈마켓의 성장을 계기로 ‘멀티미디어, 전자상거래 활황’의 시기라는 트렌드가 나타났다.

4.4. 분석기간 4기(2003-2005년)

2003년 중증급성호흡기중후군(SARS)으로 인해 글로벌 경제적 손실을 가져왔지만, 중국을 중심으로 빠른 경제회복 국면에 접어들었던 시기이다. 한국에서는 외환위기로 인한 금융산업 구조개편이 지속함에 따라 실업 문제가 심화되었으며, 벤처 버블 붕괴로 인해 기업 구조조정이 벤처업계로 확산되었다. 이에, 정부는 2004년 벤처활성화 대책을 마련하고, 2005년 창업진흥전담조직에 대한 설치 근거를 마련하는 등 벤처 건전화를 위한 정책을 추진하였다.

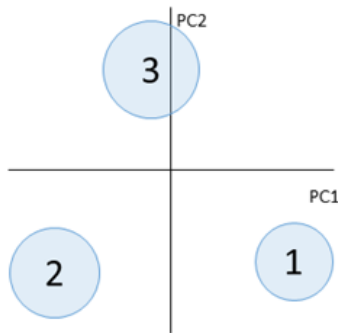
이 시기에 대한 토픽 분석 결과 <그림 6>과 같이 3

개의 토픽이 도출되었다. 먼저, 앞서 제시한 기업 구조조정이 벤처기업으로 확대되며, ‘벤처 건전화’와 관련된 트렌드가 확인되었다. 한편, 다른 두 토픽은 사회적 현상 및 산업적 이슈에 기초한 트렌드가 나타났다. 당시 가장 큰 산업적 이슈였던 IT산업과 관련하여, 유비쿼터스 기술을 활용한 정보통신 분야에 새로운 서비스 모델 창출이 확대되며, ‘IT산업의 발전’이 트렌드로 나타났다. 또한, 웰빙과 즐기세포가 사회적으로 큰 반향을 일으키며, ‘바이오산업 영역에서 벤처기업이 증가’하는 현상을 보였다.

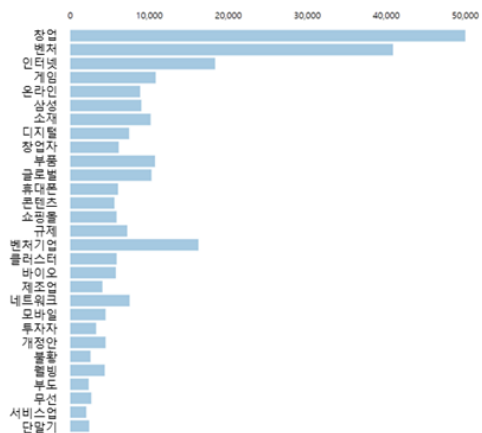
4.5. 분석기간 5기(2006-2008년)

2006년까지 활황을 기록하던 글로벌 증시가 2007년 유가 폭등에 이은 글로벌 금융위기(2007-2008년)를 맞이하며 글로벌 경기침체 국면에 접어들었던 시기이다. 한국에서는 2006년 한미 FTA에 반대하는 시위와 더불어 북한의 핵실험 등으로 사회적 불안이 가중되었으며, 글로벌 금융위기 여파로 경기침체가 가중되었다. 한편, 창업생태계는 벤처기업 중심에서 벗어나 기술창업 활성화로 패러다임을 전환하며, 2007년 제

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Salient Terms



<그림 6> 토픽 모델링 결과(2003~2005)



〈그림 7〉 토픽 모델링 결과(2006~2008)

조 창업기업의 부담금 면제, 2008년 창업절차 간소화 등 창업지원을 위한 정책기반을 마련하였다.

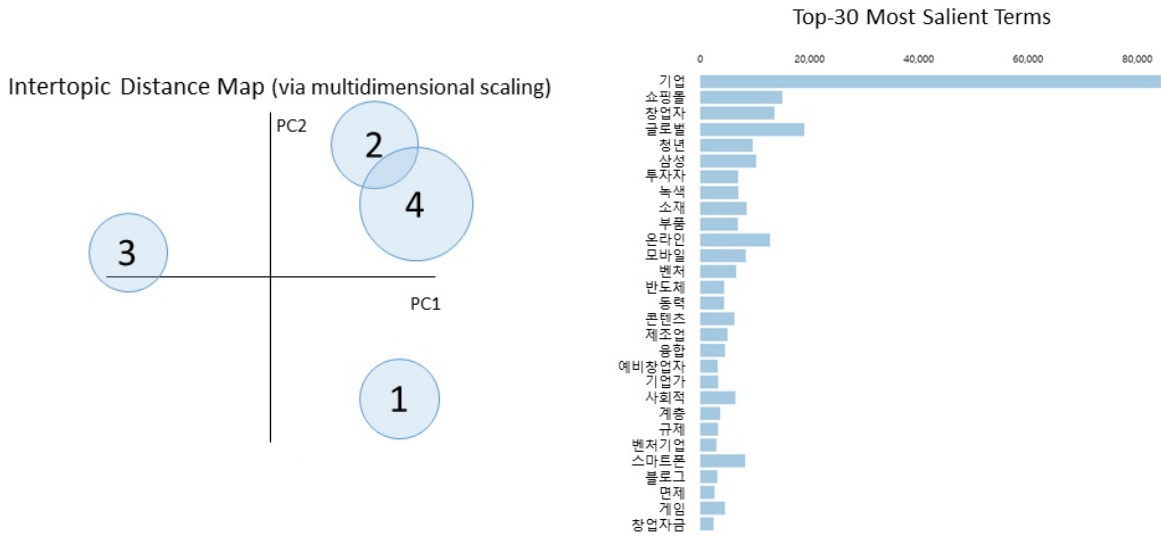
이 시기에는 <그림 7>과 같이 총 5개의 토픽이 도출되었다. 먼저, 감세, 규제완화 등의 ‘창업생태계의 정치적 이슈’가 트렌드로 확인되었다. 다음으로 산업적 이슈가 토픽으로 도출되었다. ‘삼성 휴대폰 세계 최초 최다 판매 기록을 달성’하며 글로벌 진출에 성공하며 큰 반향을 가져왔다. 이와 더불어 IT 기술의 발전과 스마트폰의 확대는 ‘온라인·모바일 콘텐츠의 확산’을 이끌어, UCC(사용자제작콘텐츠) 돌풍을 일으켰다. 또한, 웰빙 트렌드가 지속되는 한편, 명품, 퓨전, 여행 등 다양한 ‘사회·문화적 이슈와 함께 관련 아이템을 중심으로 한 창업이 활성화’되는 현상이 발생했다. 이러한 긍정적 토픽과 반대로, 글로벌 금융위기로 인해 ‘벤처업계의 불황 및 대기업-중소기업 간 양극화’가 심화하는 부정적 토픽이 확인되었다.

4.6. 분석기간 6기(2009-2010년)

2003년 SARS 이후 6년 만에 2009년 ‘신종플루 대유행’이 발생하며, 다시 감염병 위기상황이 되풀이되었

다. 이에 이어, 2010년에는 유럽의 금융·재정 위기가 확산하였고, 미·중 통화무역 갈등이 심화하며 국제경제의 혼란이 지속되었던 시기이다. 반면, 한국은 대기업의 경영실적이 개선되고, 2010년 G20 정상회의를 유치하며 경제회복에 대한 기대감이 조성되었다. 또한, 창업생태계에서는 2009년 창업진흥전담조직인 창업진흥원의 설립과 함께, 사업화 자금을 지원하기 시작하며 정부의 본격적인 창업지원정책이 추진되었다.

이 시기에 대한 토픽 분석결과 <그림 8>과 같이 4개의 토픽이 도출되었다. 먼저, 정부 영역에서는 2008년 저탄소 녹색성장 국가비전 선포 이후, 녹색성장위원회가 설립되며 ‘신재생 에너지 등 녹색 기술 및 산업 발전을 위한 국가주도의 정책’이 하나의 트렌드를 형성하였다. 또한, 앞서 제시한 바와 같이, (예비)창업자 육성을 위해 정부가 창업자금을 지원하는 프로그램을 시작으로 ‘정부의 창업기업 육성 본격화’ 트렌드가 나타났다. 그리고 산업 영역에서는 미국에서 시작된 인터넷 쇼핑몰의 국내 확산으로 쇼핑몰 창업이 확산하였으며, 대기업의 바이오 사업 참여와 정부의 신성장 동력 육성정책에 힘입어 바이오 벤처가 성장하



〈그림 8〉 토픽 모델링 결과(2009~2010)

는 등 ‘벤처업계의 약진’이 이어지는 현상을 나타냈다. 또한, 미국의 아이폰 출시를 기점으로 페이스북, 트위터에 대한 대중의 인기가 높아지며 ‘SNS 활용이 급속도로 확대’되는 트렌드가 나타났다.

4.7. 분석기간 7기(2011-2012년)

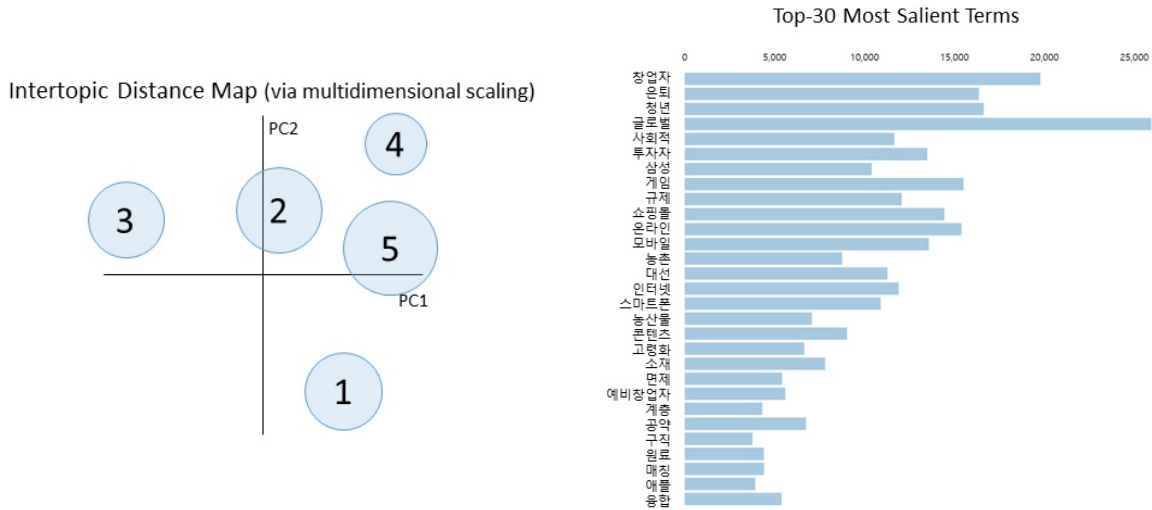
2011년 글로벌 인플레이션과 유럽발 재정위기로 인해 세계 경제 회복이 둔화되고, 주요국들이 양적 완화 정책을 추진(2012년)하며 글로벌 환율전쟁이 가시화되었던 시기이다. 같은 시기에 한국 역시, 산업화 이후 처음으로 2011년 경제 성장률이 세계경제 성장률을 밑도는 등 경제회복이 둔화하는 상황이 전개되었다. 한편, 경제성장 둔화를 타개하기 위한 신성장 동력의 원천으로 창업이 부상하며, 청년전용 창업자금을 신설하고, 2012년에 창업선도대학 지원사업을 신설하는 등 청년창업 및 예비창업자 지원을 확대하는 정책이 추진되었다.

이 시기에는 <그림 9>와 같이 총 5개의 토픽이 도출되었다. 앞서 제시한 청년·예비창업자 육성을 위한

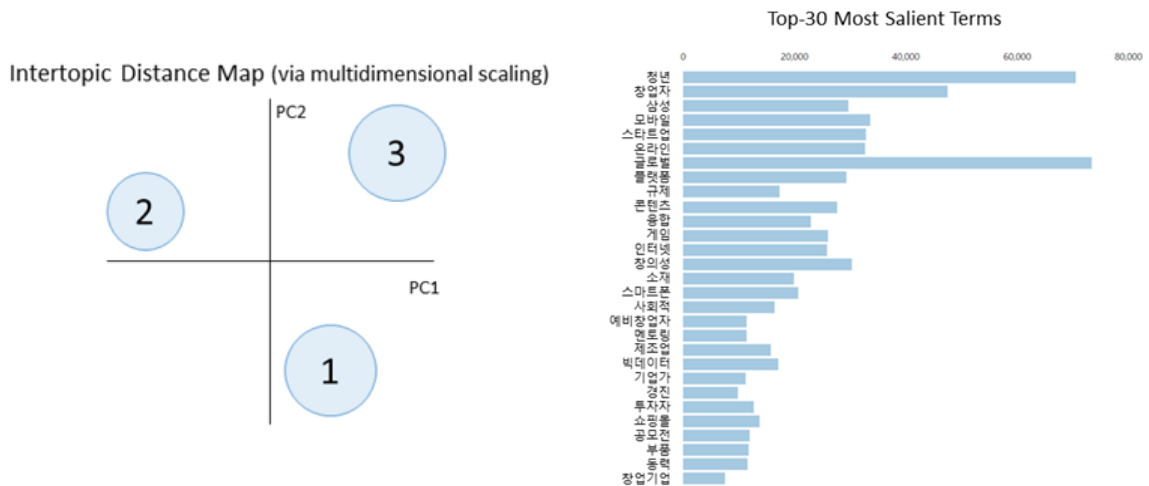
경진대회, 캠프, 교육 등 ‘다양한 지원 프로그램을 확충하며, 창업자 육성이 확대’되는 트렌드가 나타났으며, ‘창조경제’를 강조함에 따라, ‘벤처·창업 활성화’가 그 대안으로 주목받는 트렌드가 나타났다. 반면, 베이비붐 세대의 은퇴가 시작되며, 기술창업 활성화와 대비되는 ‘도·소매, 숙박·음식점업 창업이 증가’하는 현상을 보였다. 이 외에 IT 산업에서는 삼성, 애플, 구글 등 스마트폰 및 모바일 운영체제와 관련된 ‘글로벌기업 간의 소송이 본격화되며, 글로벌 경쟁이 심화’되는 한편, ‘스마트폰 게임산업이 급성장하며, 관련 창업이 증가’하는 트렌드가 확인되었다.

4.8. 분석기간 8기(2013-2015년)

2014년 전 세계적인 디플레이션이 나타나고, 이어서 2015년 파리에서의 테러 사건 이후 세계 각국이 테러 전쟁에 돌입하며 불안감이 고조되었던 시기이다. 한국에서는 창조경제 활성화를 위해 첨단기술 융합 중심의 창조산업 육성정책을 추진하였다. 이와 더불어, 2013년 TIPS 프로그램 도입, 2014년 창조경제혁신



〈그림 9〉 토픽 모델링 결과(2011~2012)



〈그림 10〉 토픽 모델링 결과(2013~2015)

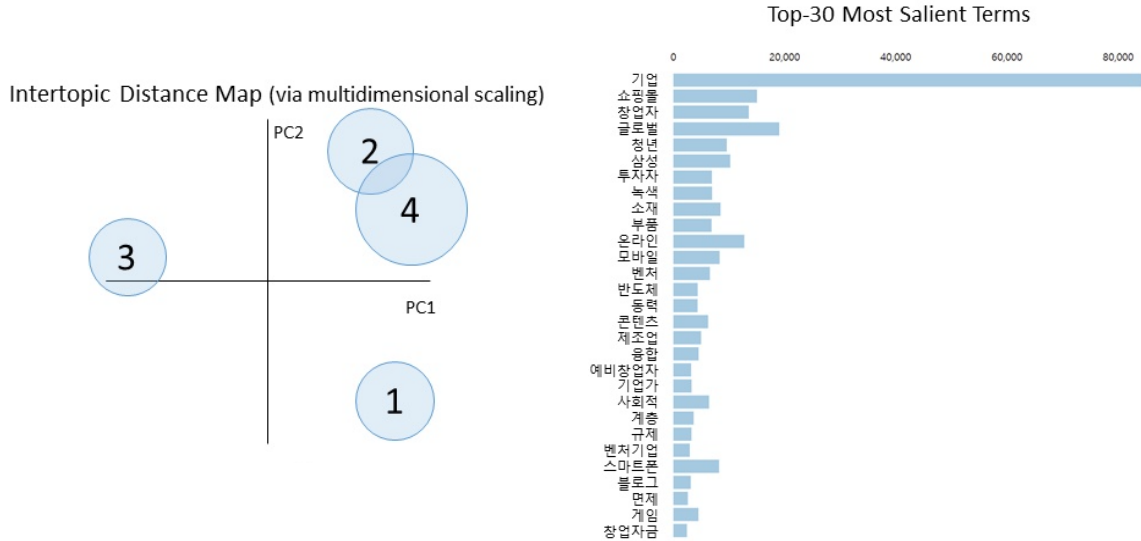
센터 설립 등 창업지원이 확대되었으며, 2013년 자금 생태계 선순환 방안, 2015년 투자활성화 대책 등 자금 생태계 조성을 위한 창업정책을 추진하며, 창조경제 실현을 위해 창업·벤처 활성화에 초점을 둔 시기였다.

2013년~2015년의 토픽 분석결과 <그림 10>과 같이 3개의 토픽이 도출되었다. 앞서 제시한 ‘창조경제’ 이슈를 중심으로, 첨단과학기술에 대한 산업 전반으로의 ‘융합을 강조하는 창조산업 육성’ 트렌드와 함께,

‘투자생태계를 중심으로 한 정부의 창업지원 강화’가 트렌드로 확인되었다. 그리고 스마트폰 보급률이 급 상승함에 따라, ‘모바일 앱을 중심으로 한 콘텐츠 창업이 확산’ 되는 트렌드가 나타났다.

4.9. 분석기간 9기(2016-2017년)

2016년 영국의 브렉시트(Brexit)로 인해 대외경제



〈그림 12〉 토픽 모델링 결과(2018~2020)

던 공유경제와 O2O 등이 인터넷, 모바일을 활용해 플랫폼 생태계를 구성하며, ‘플랫폼 스타트업이 대거 출현’하는 트렌드를 나타냈다. 또한, 지속적인 청년실업과 경기침체를 극복하고, 글로벌 경제를 선도하기 위한 국가발전전략인 ‘한국판 뉴딜의 추진’이 하나의 트렌드로 나타났다. 마지막으로 코로나19의 역사적 위기상황으로 인해 비대면 트렌드가 급속히 확산되며, 4차 산업혁명으로 촉발된 ‘디지털 전환이 가속화’되는 트렌드를 보였다.

5. 결론 및 시사점

이상과 같이 30년간의 분석을 통하여 주요 키워드와 이슈 등을 살펴보았다. 1990년대 초, 산업구조조정 등으로 인한 기업부도 상황이 증가하고 제조기업 중심의 고속성장을 위한 규제가 완화되면서 벤처기업의 본격적 확산을 위한 태동이 시작되었다. 이후 1990년대 후반기, 제도적인 지원 및 외환위기로 인한 불황극복, 그리고 2000년 초기, 무분별한 지원을 통한 벤처

1세대의 실패를 겪은 이후 벤처 건전화에 위한 노력이 시작되었고, 멀티미디어와 전자상거래를 중심으로 온라인 기반의 벤처기업이 성장하기 시작했다. 2000년도 후반기에 들어 한국 대기업의 휴대폰 분야 세계 석권 및 이를 바탕으로 모바일 기반의 콘텐츠가 확대되면서 정부는 또다시 신규창업을 위한 대규모 자금 지원을 시작하게 되었다. 이 시기는 정부가 창업, 연구개발, 투자 등 체계적이고 생애 주기적 지원을 본격화한 시기이며 2010년대 초기, 창업기업이 급속도로 증가하며 정부의 청년기업 육성이 본격화되었다. 2010년대 중반 이후는 융합을 강조하는 창조산업을 육성하고 지원사업뿐만 아니라 투자생태계 중심의 정부창업지원 사업을 강화하게 되어 창업생태계의 회수뿐만 아니라 자생적 성장을 도모하게 되었으며, 2010년대 후반기는 4차 산업혁명 중심의 핵심기술 창업지원 사업이 급부상 되고, 사회적 경제, 공유경제 관련 스타트업의 확산 및 이들의 글로벌 진출에 대한 지원이 확대되었다. 또한, 청년실업의 장기화가 지속되어 창업을 청년실업급증의 해소방안으로 활용하고 데이터 기반의 한국판 뉴딜을 추진하였다.

연도별 약 110여만 개의 키워드를 10개의 기간 분류 및 주요 키워드와 주요 이슈를 바탕으로 정리한 결과는 <표 7>과 같다. 한국의 창업정책 30년의 통합적인 시대분석을 위해

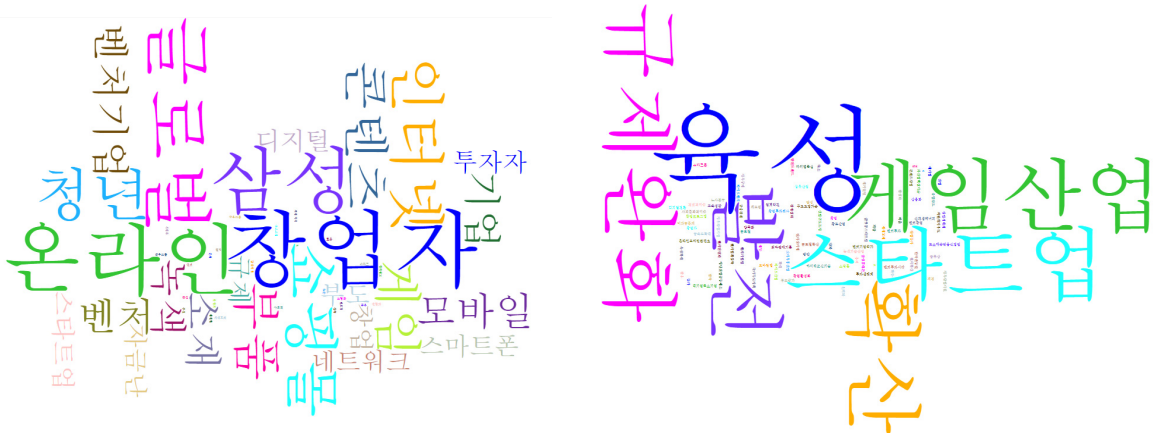
<표 7> 기간별 창업생태계 키워드 및 주요 이슈

구분	기간	시대배경	주요 키워드	주요 이슈	
				개수	내용
1	1991-1995	창업보육정책 태동	규제, 창업자, 부도, 부품, 제조업, 선거, 세계화, 금융기관, 창업투자회사, 자금난, 세대교체, 퇴진, 현대그룹, 쌍용그룹, 대우그룹, 삼성	7	① 멀티미디어 중심의 벤처 확산 ② 섬유산업 축소 및 소재·부품산업 육성 ③ 산업 구조조정으로 인한 부도 증가 ④ 정부 규제로 인한 노사분규 격감 ⑤ 제조업 중심의 고속성장을 위한 규제완화 ⑥ 대기업의 비지금 문제 대두 및 세대교체 ⑦ 벤처투자의 시작과 창업투자회사
2	1996-1998	벤처육성 법제 마련	기업, 벤처, 실업자, 벤처기업, 인터넷, 창업자, 불황, 게임, 구조조정, 부도, 자금난, 정리해고	3	① 벤처기업법 제정 ② 외환위기로 인한 부도·구조조정 가속 ③ 멀티미디어, 게임 산업 발전
3	1999-2002	벤처 건전화	창업, 인터넷, 벤처기업, 게임, 온라인, 네트워크, 디지털, 콘텐츠, 무선, 창업자, 단말기	3	① 벤처투자 확대와 버블 붕괴 ② 하이테크·신기술 상용화 ③ 멀티미디어, 전자상거래 활황
4	2003-2005	창업·벤처육성 강화	창업, 벤처, 인터넷, 벤처기업, 게임, 부품, 소재, 글로벌, 삼성, 온라인, 클러스터	3	① 벤처 건전화 ② IT산업의 발전이 트렌드 ③ 바이오 산업 영역에서 벤처기업이 증가
5	2006-2008	기술창업 육성 전환	글로벌, 인터넷, 규제, 벤처, 창업자, 삼성, 네트워크, 온라인, 쇼핑물, 부품, 콘텐츠, 웰빙	5	① 창업생태계의 정치적 이슈 ② 삼성 휴대폰 세계 최초 최다 판매 기록을 달성 ③ 온라인·모바일 콘텐츠의 확산 ④ 사회·문화적 이슈와 함께 관련 아이템을 중심으로 한 창업이 활성화 ⑤ 벤처업계의 불황 및 대기업-중소기업 간 양극화
6	2009-2010	기술창업 자금지원 확대	기업, 글로벌, 쇼핑물, 창업자, 온라인, 삼성, 청년, 스마트폰, 소재, 녹색, 사회적, 녹색	4	① 신재생 에너지 등 녹색 기술 및 산업 발전을 위한 국가주도의 정책 ② 정부의 창업기업 육성 본격화 ③ 쇼핑물, 바이오 벤처업계의 약진 ④ SNS 활용이 급속도로 확대
7	2011-2012	모바일 창업생태계 전환	글로벌, 창업자, 청년, 은퇴, 게임, 온라인, 쇼핑물, 모바일, 투자자, 인터넷, 대선, 농촌, 고령화	5	① 창업프로그램 확충 및 창업자육성확대 ② 벤처·창업 활성화 ③ 도·소매, 숙박·음식점업 창업이 증가 ④ 글로벌 경쟁 심화 ⑤ 스마트폰 관련 게임산업 증가
8	2013-2015	창업·벤처 중심 창조경제 실현	글로벌, 청년, 창업자, 모바일, 스타트업, 온라인, 창의성, 삼성, 플랫폼, 콘텐츠, 빅데이터	3	① 융합강조 창조산업 육성 ② 투자생태계 중심 정부창업지원 강화 ③ 모바일앱 중심의 콘텐츠 창업확산
9	2016-2017	신기술·신산업 창업 중심 4차 산업혁명	청년, 글로벌, 혁명, 창업자, 온라인, 공유, 쇼핑물, 스마트, 콘텐츠, 모바일, 삼성, 로봇, 멘토링	4	① 4차산업 핵심기술의 급부상 ② 공유경제 관련 스타트업 글로벌 확산 ③ 스타트업 규제완화 이슈대두 ④ 창업이 청년실업증의 해소방안
10	2018-2020	K-스타트업의 디지털 대전환 주도	기업, 글로벌, 쇼핑물, 창업자, 온라인, 삼성, 청년, 소재, 스마트폰, 녹색, 투자자, 부품	3	① 플랫폼 스타트업의 대거출현 ② 한국판 뉴딜의 추진 ③ 디지털전환 가속화

주요 키워드와 주요 이슈로 워드 클라우드 분석을 한 결과는 <그림 13>과 같다. 워드 클라우드 분석은 문서 주제설명과 분류를 통한 설명 작업으로 유용하고, 이러한 분류를 통하여 추출된 단어를 작업은 의미별 구역으로 나누고 색상을 구별하여 시각적으로 쉽게 인식할 수 있게(Hearst et al., 2020)하여, 연구자가 의미 있는 방식으로 분석결과를 묘사할 수 있다(De Paolo & Wilkinson, 2014). 주요 키워드는 총 127개 단어로 추출되었으며, 상위 10개의 용어는 ‘창업자>온라인>삼성>글로벌>청년>인터넷>쇼핑몰>콘텐츠>부품>게임’의 순으로 나타났다. 한국의 창업생태계가 대기업의 영향을 받고 온라인과 인터넷 쇼핑몰 중심의 플랫폼기업의 확대와 글로벌 진출 노력과 청년실업 해소 방안을 위해 노력하였음을 알 수 있었다. 다음으로 주요 이슈는 총 105개 단어로 추출되었고 상위 10개의 용어는 ‘육성>멀티미디어>확산>스타트업>발전>규제완화>게임산업>활황>활용>활성화’로 나타났다. 창업생태계 내에서 언론에서 다루었던 주요 이슈는 ‘육성, 확산, 규제완화, 활황’ 등으로, 용어 내에서도 볼 수 있듯이 한국은 창업생태계 발전을 위해 정부주도형 육성정책을 펼쳤음을 확인할 수 있다.

이상과 같이 선행연구의 분석을 통해 창업생태계의 용어 정의와 구성요인을 도출하고 선행연구 내에서 생태계, 스타트업, 혁신, 기술, 소프트웨어, 디지털, 기술 및 정치, 경제, 사회, 정부 등에 대한 다양한 주제를 다루고 있음을 확인할 수 있었다. 한국의 창업과 관련된 30년간의 언론기사 분석을 통해 창업정책, 법제도, 창업기업의 흥망과 성쇠, 그리고 글로벌 경제의 변화와 4차 산업혁명의 세계적 큰 변화에 따른 정부와 산업 내의 변화 등을 본 연구를 통해 알 수 있었다. 한국 창업생태계의 가장 큰 특징은 시대적 배경에서도 확인할 수 있듯이 정부주도형 창업생태계 구축 및 발전이다. 생태계의 육성을 위해 정책을 수립 후 법 제도의 마련을 통하여 성장을 가속화하고, 성장과 실패의 경험을 바탕으로 제도와 지원 프로그램의 보완을 통해 성장을 가속했다.

본 연구는 한국의 창업생태계 변화를 파악하고 주요 키워드를 도출하는 목적에 따라 30년간의 기간을 10개의 구간으로 압축하여 그 트렌드를 분석하였다. 이는 유사한 현상이 다년간에 걸쳐 나타나는 정책 트렌드의 성격을 반영하는 방법이었으나, 객관적인 분석을 통해 구간을 병합하는 방법에 관한 선행연구가



<그림 13> 주요키워드, 이슈의 워드 클라우드 분석 결과

부재한 상황에서 탐색적인 방법을 활용함에 따른 한계점이 발생하였다. 또한, Interactive Plot을 활용한 분석기법의 제약으로 토픽분석 결과에 나타난 각각의 토픽에 대한 키워드 분석결과를 모두 수록하지 못한 한계를 가지고 있다. 하지만, 본 연구는 창업정책을 포함한 다양한 정책 변화를 연구한 그간의 선행연구에서 제한적으로 연구되었던 언론기사를 대상으로 30년이라는 장기간에 걸친 트렌드 변화를 텍스트마이닝 기법을 통해 분석하였다. 이는 정책입안자의 관점이 아닌 정책의 대상이 되는 국민의 시각에서 접근한 것으로, 선행연구(하규수 등, 2016; 김만수, 강재원, 2021)가 창업정책의 변화를 정부와 정책의 변경 시기를 기준으로 변화를 분석한 것과 달리, 언론기사의 키워드와 토픽 변화를 기준으로 창업생태계의 트렌드 변화를 분석하였다는 점에서 학술적 의의를 지니고 있다. 뿐만 아니라 시대별로 나타난 한국 창업생태계의 특징을 바탕으로 창업성과가 우수한 주요 국가와의 비교, 그리고 연구에서 나타나는 다양한 창업생태계 구성요인들에 대한 분석 등으로 발전된다면 향후 정부의 정책수립, 연구자들의 연구주제 등 다양한 분야에서 활용할 수 있는 실무적 시사점을 제공하였다.

〈참고문헌〉

[국내 문헌]

1. 김만수, 강재원 (2021). 정책 패러다임 관점에서 살펴본 창업정책 변화. **벤처창업연구**, 16(3), 43-58.
2. 박명석, 권영진, 이상용 (2018). 댓글이 음원 판매량에 미치는 차별적 영향에 관한 텍스트마이닝 분석. **지식경영연구**, 19(2), 91-108.
3. 박종도 (2019). 토픽 모델링을 활용한 다문화 연구의 이슈 추적 연구. **한국문화정보학회지**, 53(3), 279-289.
4. 변영조, 이상한, 김재영 (2020). 대학의 지속가능성을 위한 지역연계 창업혁신 활동 성과지표 개발에 관한 탐색적 연구. **인터넷전자상거래연구**, 20(3), 33-47.
5. 서리빈, 최경철, 변영조 (2020). 「중소기업연구」 40년 동안의 창업생태계 연구 동향 고찰 및 측정모형 개발을 위한 탐색적 연구. **중소기업연구**, 42(4), 69-102.
6. 손희영, 이명중 (2021). '91~' 20년 언론에 비친 창업·벤처의 모습. 조사연구-2021, 세종: 창업진흥원.
7. 이금실, 이인주, 이영경 (2018). LDA(Latent Dirichlet Allocation) 기반 토픽 모델링 기법을 활용한 관광분야의 가상현실(Virtual Reality) 연구동향 분석. **2018년도 제50차 한국관광레저학회 학술발표대회 논문집**, 425-432.
8. 이소현, 손새아, 김희웅 (2019). 텍스트마이닝을 이용한 미용성형 주요 요인에 관한 연구. **지식경영연구**, 20(1), 45-75.
9. 이우진, 오혜미 (2019). 2013~2016년 한국 벤처창업정책의 특징에 관한 연구. **한국창업학회지**, 14(1), 129-159.
10. 주재홍, 송지훈 (2021). 키워드 네트워크 분석을 활용한 지식은폐 연구동향 분석. **지식경영연구**, 22(1), 217-242.
11. 중소벤처기업부 보도자료 (2022). **실질 창업 증가 속, 기술창업 역대 최대 23만개 달성**. 제142009호.
12. 하규수, 홍길표, 이춘우 (2016). 2002-2012년의 한국 벤처창업정책의 특징 연구. **벤처창업연구**, 12(1), 15-24.
13. 현기순 (2022). 지속가능성에 대한 지리학의 연구 동향 분석-토픽 모델링과 키워드 네트워크 분석을 중심으로-. **대한지리학회지**, 57(1), 1-17.
14. 황보운, 김재형 (2016). 창업기업 성과에 미치는 창업보육지원 서비스 연구. **벤처창업연구**, 11(5), 57-66.

[국외 문헌]

15. Cavallo, A., Ghezzi, A., & Balocco, R. (2018). Entrepreneurial ecosystem research: Present debates and future directions. **International Entrepreneurship and Management Journal**, 3(2), 1-31.
16. De Battisti, F., Ferrara, A., & Salini, S. (2015). A decade of research in statistics: A topic model approach. **Scientometrics**, 103, 413-433.
17. DePaolo, C. A., & Wilkinson, K. (2014). Get your head into the clouds: Using word clouds for analyzing qualitative assessment data. **Tech Trends**, 58, 38-44.
18. Feld, B. (2012). **Startup communities: Building an entrepreneurial ecosystem in your city**. New York, NY: Wiley.
19. GERA. (2019-2021). **Global entrepreneurship monitor global report**.
20. Hayton, J. C., George, G., & Zahra, S. A. (2002). National culture and entrepreneurship: A review of behavioral research. **Entrepreneurship Theory and Practice**, 26(4), 33-52.
21. Hearst, M. A., Pedersen, E., Patil, L., Lee, E, Laskowski, P., & Franconeri, S. (2020). An evaluation of semantically grouped word cloud designs. **IEEE Transactions on Visualization and Computer Graphics**, 26(9), 2748-2761.
22. Isenberg, D. (2010). How to start an entrepreneurial revolution. **Harvard Business Review**, 88(6), 40-50.
23. Mason, C., & Brown, R. (2014). **Entrepreneurial ecosystems and growth oriented entrepreneurship**. Paris: OECD.
24. Qaiser, S., & Ali, R. (2018). Text mining: Use of TF-IDF to examine the relevance of words to document. **International Journal of Computer Applications**, 181(1), 25-29.
25. Shwetter, C., Maritz, A., & Nguyen, Q. (2019). Entrepreneurial ecosystems: A holistic and dynamic approach. **Journal of Industry-University Collaboration**, 1(2), 79-95.
26. Spigel, B. (2017). The relational organization of entrepreneurial ecosystems. **Entrepreneurship Theory and Practice**, 41(1), 49-72.

27. Stam, E. (2015). Entrepreneurial ecosystems and regional policy: A sympathetic critique, *European Planning Studies*, *23*(9), 1759–1769.
28. Stam, E., & Van de Ven, A. (2021). Entrepreneurial ecosystem elements, *Small Business Economics*, *56*, 809–832.

저 자 소 개



손 희 영 (Heeyoung Son)

현재 창업진흥원 대외협력실 실장으로 재직 중이다. 강원대학교에서 경영학 박사 학위를 취득하였고, 창업진흥원 조사연구실에서 연구위원으로 근무하였다. 주요 관심분야는 국내외 창업정책 및 창업생태계, 기업가정신 등 창업 관련 분야와 소상공인, 서비스마케팅 분야이며, 지금까지 벤처창업연구, 한국창업학회지, 예술인문사회 융합 멀티미디어 논문지, 한국경영과학회지, 산업혁신연구, 마케팅관리연구, MANAGING SERVICE QUALITY, International Journal of Software Engineering and its Application 등 주요 학술지에 논문을 발표하였다.



이 명 종 (Myungjong Lee)

현재 창업진흥원 창업규제정책실 실장으로 재직 중이다. 충북대학교에서 마케팅을 전공하여 경영학 박사 학위를 취득하였고, 창업진흥원 창업정책연구실에서 연구위원으로 근무하였다. 주요 관심분야는 스타트업 성장전략, 스타트업 생태계 개발전략, 스타트업 규제혁신 등 창업정책 분야와 계량 마케팅, 하이테크 마케팅 분야이다. 지금까지 한국경영과학회지, 벤처창업연구 등 주요 학술지에 논문을 발표하였다.



변 영 조 (Youngjo Byun)

현재 국립한밭대학교 공과대학 산학융합학부 부교수로 재직 중이다. 건국대학교에서 경영학 박사 학위를 취득하였고, 온라인 기업교육 전문회사인 ㈜크레듀(현-삼성멀티캠퍼스)에서 책임연구원, 현재 한밭대학교 창업보육센터장으로 근무중이다. 주요 관심분야는 창업정책, 기업가정신교육 및 기업가적 대학의 지속가능 전략 수립, 벤처기업의 ESG 등이며 지금까지 중소기업연구, 벤처창업연구, 한국창업학회지, 인터넷전자상거래연구, 표준인증안전학회지, Clean Technology 등 주요 학술지에 논문을 발표하였다.

〈 Abstract 〉

Analysis on Dynamics of Korea Startup Ecosystems Based on Topic Modeling

Heeyoung Son^{*}, Myungjong Lee^{**}, Youngjo Byun^{***}

In 1986, Korea established legal systems to support small and medium-sized start-ups, which becomes the main pillars of national development. The legal systems have stimulated start-up ecosystems to have more than 1 million new start-up companies founded every year during the past 30 years. To analyze the trend of Korea's start-up ecosystem, in this study, we collected 1.18 million news articles from 1991 to 2020. Then, we extracted news articles that have the keywords “start-up”, “venture”, and “start-up”. We employed network analysis and topic modeling to analyze collected news articles. Our analysis can contribute to analyzing the government policy direction shown in the history of start-up support policy. Specifically, our analysis identifies the dynamic characteristics of government influenced by external environmental factors (e.g., society, economy, and culture). The results of our analysis suggest that the start-up ecosystems in Korea have changed and developed mainly by the government policies for corporation governance, industrial development planning, deregulation, and economic prosperity plan. Our frequency keyword analysis contributes to understanding entrepreneurial productivity attributed to activities among the networked components in industrial ecosystems. Our analyses and results provide practitioners and researchers with practical and academic implications that can help to establish dedicated support policies through forecast tasks of the economic environment surrounding the start-ups. Korean entrepreneurial productivity has been empowered by growing numbers of large companies in the mobile phone industry. The spectrum of large companies incorporates content startups, platform providers, online shopping malls, and youth-oriented start-ups. In addition, economic situational factors contribute to the growth of Korean entrepreneurial productivity the economic, which are related to the global expansions of the mobile industry, and government efforts to foster start-ups. Our research is methodologically implicative. We employ natural language processes for 30 years of media articles, which enables more rigorous analysis compared to the existing studies which only observe changes in government and policy based on a qualitative manner.

Key words: Startup ecosystem, Startup trend, Startup support program, Text mining analysis, Korean startup development history

* KISED, Korea Institute of Entrepreneurship Development

** KISED, Korea Institute of Entrepreneurship Development

*** Hanbat National University