Journal of the Korea Academia-Industrial cooperation Society
Vol. 23, No. 10 pp. 303-313, 2022

https://doi.org/10.5762/KAIS.2022.23.10.303
ISSN 1975-4701 / eISSN 2288-4688

데이터 마이닝을 활용한 대학생과 교원의 창업 아이템 분석

김재훈, 정혜진*
부산대학교 행정학과

An Analysis on Items of Startups by Undergraduates and Faculties Using Data Mining

Jae-Hoon Kim, Hyejin Jung*
Department of Public Administration, Pusan National University

요 약 대학은 교육과 지식 창출 뿐만 아니라, 새로운 기술을 통한 경제적 가치 창출을 위한 역할이 강조되고 있다. 최근 COVID-19로 인해 비대면 거래가 늘어나면서 대학생과 교원의 창업 아이템은 창업 기업의 성공과도 많은 관련이 있다. 이에 본 연구는 R 프로그램 기반 데이터 마이닝 기법을 활용하여 대학생과 교원의 창업 아이템을 분석하였다. 2020년부터 2021년 동안 3,422개 대학생 창업 기업과 615개 교원 창업 기업의 아이템에 대하여 TF-IDF분석, 동시출현 네트워크분석, LDA 토픽분석을 실시하였다. 분석 결과, 대학생과 교원의 창업 아이템 가운데 빈출 키워드와 중요 키워드 간의 차이가 존재하였다. 뿐만 아니라, 네트워크 분석과 토픽 모델링 분석에서도 대학생은 주로 온라인 플랫폼, 의류, 전자상거래의 키워드가 중요한 반면, 교원 창업은 치료제, 시스템, AI 등이 중요한 키워드로 도출되었다. 본 연구는 향후 대학생과 교원의 창업의 성공 가능성을 높일 수 있는 정책적 시사점을 도출하였다.

Abstract The roles of universities include not only education and new knowledge creation, they also include creation of economic value through new technology. The startup items for new firms created by undergraduates and faculties are closely related to the success of those startup companies as the so-called untact businesses increase due to the outbreak of COVID-19. This study analyzes such startup items of undergrads and faculty members who use data mining methodologies based on the R program. To be specific, we simultaneously dealt with 3422 new firms and 615 new firms created by undergraduates and faculties, respectively, from 2020 to 2021 by using three different methodologies, including TF-IDF, a co-occurrence network, and LDA topic modeling. The results found differences in keywords for frequency and importance in both groups. Furthermore, we also found that important keywords and topics were different between startups by undergraduates and those by faculties. Online platforms, fashion items, and online transfer took center positions and were frequent topics in the networks of the former group, whereas remedies, systems, and AI were found to be critical in the networks of the latter group. We suggest various policy implications to increase the probability of success for new firms created by undergrads and faculties.

Keywords : Entrepreneurial University, University Startup, Items of University Startup, Data Mining, Topic Modeling

본 연구는 대학 창업교육 및 창업문화 활성화를 위한 성과분석 및 발전방향 연구(한국연구재단, 2022)의 연구 내용을 일부 발췌하여 보완한 연구임.
*Corresponding Author : Hyejin Jung(Pusan National University)
email: hjung@pusan.ac.kr
Received July 19, 2022 Revised August 18, 2022
Accepted October 7, 2022 Published October 31, 2022

1. 서론

최근 자본과 노동과 더불어 지식을 기반으로 하는 사회로 변화함에 따라, 지식을 생산하는 대학은 산업발전의 원동력으로 인식되고 있다[1]. 이러한 맥락에서 대학은 연구를 통해 지식을 생산하는 역할 뿐만 아니라, 창업과 기술 이전 등을 통해 새로운 지식과 기술을 경제적 가치를 지닌 지식으로 전환할 수 있는 기업가적 역할이 강조되고 있다[2]. 특히 대학 창업은 기술과 시설, 자금 및 인력 등을 활용하여 새로운 기업을 설립하는 활동으로서[3], 대학의 기술과 지식은 창업을 위한 핵심적인 기반이라고 볼 수 있다[4]. 대학 창업을 활성화하기 위한 국가적 지원 역시 이러한 맥락에서 지속적으로 이어지고 있다.

한편, 대학생과 교원의 창업은 서로 다른 특징을 가지고 있다. 대학생 창업의 폐업에 대한 공식적인 통계는 없지만, 2019년을 기준으로 30대 이하 창업주의 경우 3년 미만 폐업률이 42.7%에 달할 정도로 전 연령대에 비해 사업이 실패할 확률이 매우 높은 편이다[5]. 즉, 청년 실업률 해소와 더불어 국가의 경제 발전을 위해 대학생 창업에 대한 정부 차원의 다양한 지원이 이루어졌으나, 청년층의 창업 생존율은 낮다고 볼 수 있다. 이와 달리 교원 창업은 연구를 통해 기반으로 하는 경우 원천기술 산업화 가능성이 학생 창업에 비해 더 크며, 창업 성공 가능성과 일자리 창출 가능성 역시 더 높은 편이다[6].

그동안 대학생과 교원의 창업 의지와 성과에 대한 영향 요인에 대한 연구는 매우 활발하게 이루어져 왔다. 예컨대, 대학생의 창업 성과에는 창업휴학제, 창업학점교류제 등 창업친화적 학사제도와 학업 교육과정이 긍정적인 영향을 미치는 것으로 알려져왔다[7]. 또한 교원 창업에는 대학의 특허와 논문 수, 기술개발 등 대학의 산학협력 역량이 영향을 주는 것으로 나타났다[8]. 한편, 창업 아이템은 기업의 재무적, 비재무적 성과에 미치는 중요한 요소가 될 수 있다[9,10]. 그러나 아직까지 대학생과 교원의 창업 아이템에 대한 분석은 체계적으로 이루어지지 않았다. 앞서 언급한 바와 같이 대학생과 교원의 창업 성공률이 서로 다르다면, 각 그룹의 창업 아이템이 기업의 성과와 폐업에도 많은 영향을 미칠 수 있음을 유추할 수 있다.

이에 본 연구는 데이터 마이닝을 활용하여 대학 창업 아이템을 분석하고자 한다. 통계프로그램 R을 이용하여 2020년부터 2021년 동안 3,422개 대학생 창업 기업과 615개 교원 창업 기업의 아이템에 대하여 TF-IDF분석을 수행하여 대학생과 교원의 창업 아이템의 주요 키워드를 파악하였다. 이와 더불어 네트워크 분석을 통해 주요 창업 아이템의 키워드 간 관계를 살펴보았다. 마지막으로 창업 아이템 빈도를 바탕으로 토픽을 추출하는 LDA(Latent Dirichlet Allocation, 이하 LDA)분석을 통해 대학생과 교원 창업 아이템의 주요 경향을 분석하였다. 이를 통해 본 연구는 대학생과 교원의 창업 아이템의 경향과 특성을 파악하는 한편, 향후 대학생의 창업 성공을 위한 지원 정책 및 교원의 창업을 활성화하기 위한 정책적 시사점을 도출하고자 하였다.

이하 본 연구는 다음과 같이 구성된다. 제2장에서는 대학 창업의 현황과 특성을 살펴보고, 제3장에서는 분석 대상 및 연구 방법을 설명한다. 제4장에서는 데이터 마이닝을 활용한 대학생과 교원의 창업 아이템 분석 결과를 제시하고, 마지막으로 제5장에서는 정책적 시사점과 연구의 한계를 논의한다.

2. 대학 창업 현황 및 특성

2.1 기업가적 대학과 대학 창업

지식 기반의 경제가 활성화되면서, 대학은 연구와 교육을 넘어 경제적 가치를 지닌 새로운 지식을 창출할 수 있는 역량이 강조되는 기업가적 대학의 필요성이 증대되었다[11]. 기업가적 대학(entrepreneurial university)이란 대학의 지식을 기반으로 생성된 성과를 기술 이전이나 창업 등을 통해 사업화하여 경제적 가치를 창출하는 대학의 역할을 의미한다[12].

기업가적 대학은 전통적인 대학 즉, 이른바 상아탑으로서의 대학과 비교하여 다음과 같이 네 가지 전제 조건을 가지고 있다[13]. 첫째, 기업가적 대학은 산업계와 정부와 긴밀하게 상호작용하는 특징을 나타낸다(interaction). 둘째, 특정한 제도적 산출이라기보다는 독립적인 기관의 특성을 가지고 있다(independence). 셋째, 기업가적 대학은 연구 중심에서 기업가적 모델로 전환을 하는 과정에서 앞서 언급한 상호작용성과 독립성을 동시에 보유하고 있다(hybridization). 마지막으로 대학이 산업계와 정부와의 지속적인 상호작용을 통해 대학 내부의 구조도 서서히 혁신적으로 변화한다는 것이다(reciprocality). 이러한 가정을 토대로 Table 1은 상아탑으로서의 대학과 기업가적 대학의 차이를 제시하고 있다.

Table 1. Ivory Tower university vs Entrepreneurial university[13]

Ivory Tower university	Entrepreneurial university
Isolated form the society	Open and serve to the external society
Teaching on campus	Teaching on/off campus
Knowledge production for own sake	Polyvalent knowledge produced
Meandering stream of basic research	Multiple sources of input into research direction
Useful knowledge as accident	Useful knowledge sought
No organizational technology transfer capability and no firm formation	TTO, Incubator integrated into innovation strategy to foster start-ups
Discipline-based departments as primary units	Departments and inter-disciplinary centers have equal status
Single internal stakeholder	Multiple stakeholders –internal and external
University administration only from academia	University administration from multiple sources, including industry and government
Funding as matter of right	Funding as matter of exchange, something to be earned
Operation for self sustainability	Make significant contribution to regional development as well
Only academic mind-set	With entrepreneurial ethos

창업과 사업화 등을 통한 기업가적 대학의 역할은 최근 COVID-19 발생에 따른 비대면 사회와 가벼운 창업(lean startup)의 맥락에서 더욱 증대되었다[14]. 먼저 COVID-19로 인해 비대면 거래가 확산되면서, 새로운 비즈니스 모델과 신산업 창출의 가능성이 확대되고 있다. 특히 최근 온라인 거래가 폭증함에 따라 다양성, 맞춤형, 실감형, 디지털화, 스마트화로 요약되는 새로운 니즈가 발굴되고 있다[15]. 이로 인해 언택트, 스마트 팩토리(Smart Factory), AI/빅데이터, 게임, OTT(Over the Top), 보안, 디지털 전환, 디지털 헬스, 에듀 테크(Edu Tech), 로봇, 클라우드 컴퓨팅 영역에 대한 관심이 증대되었다[16]. 중요한 점은 이러한 분야에서 스타트업 기업이 빠르게 성장하고 있을 뿐만 아니라, 양질의 일자리도 증가하고 있다는 점이다[14].

둘째, COVID-19로 인해 가벼운 창업을 통한 속도감 있는 창업이 확산되고 있다[14]. 가벼운 창업은 물건을 생산하거나 해결책을 고안하기 위해 소비자로부터 학습하고, 기민하고 민첩하게 실험하는데 초점을 두고 있다[17]. 이러한 창업 전략은 창업가들이 제품이나 서비스를 고안 및 개발하는데 있어서 비용과 시간을 최소화할 수 있도록 하는 프로토타입이 중요하다[18]. 더욱이 COVID-19로 인해 취업 기회가 줄어들고 온라인 창업 플랫폼이 확대 되면서 대학생들이 실패하더라도 상대적으로 소규모로 창업할 수 있는 로우리스크(low risk) 창업에 대한 시도가 늘어나게 되었다[19].

2.2 대학생과 교원 창업의 현황과 특성

Fig. 1은 학생과 교원 창업의 수를 나타내고 있다. 대학생의 창업 기업 수는 2016년 1,191개에서 2020년 1,805개로, 교원의 창업 기업 수는 같은 기간 195개에서 333개로 최근 5년간 계속해서 증가하였다. 또한, 대학생의 창업 기업 수는 교원의 창업 기업 수보다 매년 5배 이상 많은 편이다.

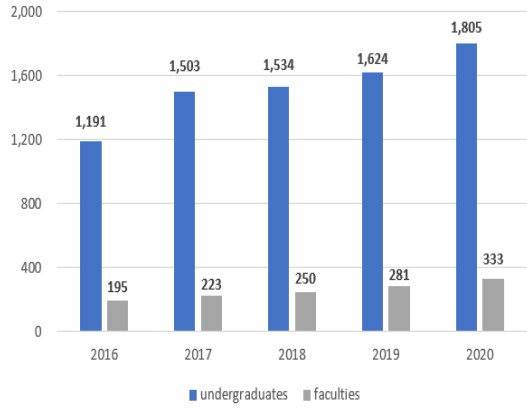


Fig. 1. Number of startups by undergraduates and faculties[20]

앞에서 언급한 바와 같이 COVID-19로 인해 차별화된 아이디어를 사업화할 수 있는 가벼운 창업이 강조되었으나, 대학생의 경우 창업 초기 자금의 어려움이 큰 편이다. 2019년 창업실태조사[21]에 따르면, 창업 시 장애요인에는 '창업자금 확보의 어려움'이 70.9%로 가장 높게 나왔고, '창업 초기부터 성공시까지 경제활동 문제'가 23.1%로 나타났다. 중요한 점은 이처럼 창업 자본이 충분하지 않은 경우 사업의 실패로 이어지기 쉽다는 점이다. 예컨대, 20대에 창업한 신생기업의 53.4%는 3년까지 생존하고, 2년까지 생존한 기업은 35%, 3년까지 생존한 기업은 26.6%에 불과하다[22]

이와 달리 대학 교원 창업은 높은 기술력을 토대로 기술 기반 창업이 많은 편이다. 기술 기반 창업이란 기술이나 지식의 집약 정도가 상대적으로 높은 창업을 의미하며[23], 일반 창업보다 성장성이 높다. 따라서 교원 창업은 학생 창업에 비해 그 수가 적음에도 불구하고, 경제발

전에 더 긍정적인 영향을 미칠 수 있다[24]. 뿐만 아니라 교원 창업은 일자리를 생성하는 경제 주체로서, 대학 근처에 창업을 함으로써 지역 경제에 긍정적인 영향을 준다[25].

교원 창업을 증진하기 위해 벤처기업육성에 관한 특별조치법에 휴직과 겸직 허용에 관한 규정이 제정되어 있다. 2021년을 기준으로 창업 휴직제를 시행하는 일반 대학과 2년제 대학은 각각 49.4%와 36.3%에 달한다. 그러나 Fig. 2에서 보는 것과 같이 2017년부터 2021년까지 일반 대학의 휴직 교원 수는 4~8명으로 많지 않고, 2년제 대학에서 해당 제도를 이용한 교원 수는 존재하지 않는다[26].

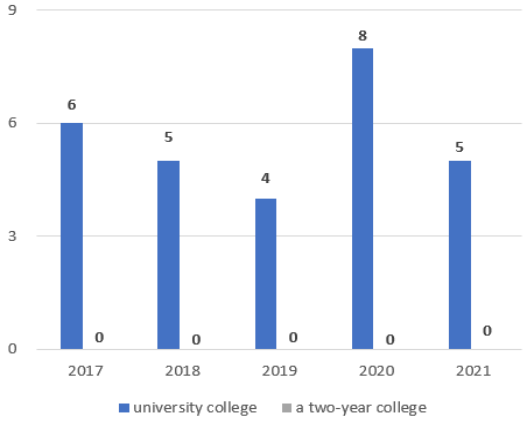


Fig. 2. Number of facilities using layoff system[26]

이와 같이 휴직 교원 수가 저조한 이유는 다양하지만, 대표적인 이유 중의 하나는 업무 공백 대체에 대한 부담이 따르기 때문으로 해석할 수 있다[25]. 뿐만 아니라, 교원은 업무를 유지하는 경우 높은 보수와 직업 안정성이 보장되지만, 창업으로 인해 휴직하게 되는 경우 실패 위험에 대한 기회비용이 높기 때문에 창업 휴직제도를 이용할 동기가 높지 않다는 문제점이 지적되고 있다[14].

3. 연구 설계

3.1 분석 대상

본 연구는 대학 내 창업 기업의 아이템 분석을 위해 대학생과 교원의 창업으로 나누어 연구를 수행하였다. 연구의 시간적 범위는 2020년부터 2021년까지 2개년을 선정하였다. 이는 대학별로 창업 기업에 대한 아이템 분석이 2020년부터 이루어졌으며, 2021년이 가장 최근 데이터이기 때문이다. 분석을 위한 자료 수집은 교육부와 한국연구재단의 산학협력지원팀으로부터 지원을 받았다.

먼저 대학생 창업의 아이템 분석은 분석 기간 동안 4년제 대학생과 2년제 대학생이 창업한 3,438개의 기업 가운데 창업 아이템이 존재하지 않는 16개 기업을 제외하고, 4년제 대학생 창업기업 2,936개와 2년제 대학생 창업기업 486개, 총 3,422개를 대상으로 분석하였다. 또한 교원 창업 아이템 분석은 일반 대학 교원 창업기업 580개와 2년제 대학 창업기업 35개, 총 615개를 분석하였다.

3.2 분석 방법

3.2.1 데이터 전처리

데이터 전처리 단계는 수집된 비정형 텍스트 데이터를 분석이 쉬운 형태로 만드는 과정이다. 전처리 방식에 따라 분석 결과가 달라지므로 전처리 과정은 복잡하고 중요한 작업이다[27]. 본 연구에서는 R-Studio 프로그램을 활용하여 전처리 작업과 형태소 분석을 하였다. 전처리 작업에서는 R의 stringr 패키지를 사용하여 공백을 제거하고 한글만 남겼다. 형태소 분석은 텍스트에서 형태소를 추출하여 명사, 형용사 등 품사로 나누는 과정이다[28]. tidytext 패키지와 한글 자연어 분석 패키지인 KoNLP(Korean Natural Language Processing) 패키지를 활용하여 창업 아이템에서 명사만 추출하였다. 명사 중에서 아이템에 속하지 않는 불용어를 제거하고 유사한 의미를 가진 단어는 하나의 단어로 통일시켜 유의어를 수정하였다. 명사 추출 시 분석에 있어 의미가 없는 불필요한 한 글자(예: 등, 및) 명사는 제거하고 두 글자 이상 명사만 추출하였다.

3.2.2 중요 키워드

본 연구에서는 대학생과 교원의 창업 아이템 가운데 중요 키워드를 추출하기 위해서 TF-IDF(Term Frequency-Inverse Document Frequency)를 적용하였다. TF-IDF는 단어 빈도를 나타내는 TF와 역 문서 빈도를 나타내는 IDF를 곱한 값으로 키워드가 특정 텍스트에서 얼마나 사용되었는지 계산한다[29]. 따라서 TF-IDF는 키워드가 흔하지 않으며, 특정 텍스트에서 빈번하게 사용되는 정도를 나타낸다. TF-IDF의 값은 텍스트의 개성을 나타내는 키워드의 중요도를 판단할 수 있으며[30], 본 연구에서는 학생과 교원 창업 아이템에서 각각 상위 10개의 키워드를 추출한 후 비교하였다.

3.2.3 동시 출현 네트워크

학생과 교원 창업 아이템에서 텍스트의 맥락을 이해하기 위해 동시 출현 네트워크(Co-occurrence network) 분석을 하였다. 동시 출현 네트워크는 각 키워드가 동시에 출현하는 빈도를 활용하여 단어의 관계를 네트워크 형식으로 표현한 것이다[28].

본 연구에서는 학생과 교원 창업 아이템에 대하여 동시 출현 빈도를 구한 후, tidygraph 패키지를 활용하여 네트워크 그래프 데이터를 생성하였다. 네트워크의 복잡성을 고려하여 학생 창업 아이템에서는 5회 이상 사용된 단어만 추출하였고, 교원 창업 아이템에서는 2회 이상 사용된 단어만 추출하였다.

3.2.4 LDA 토픽모델

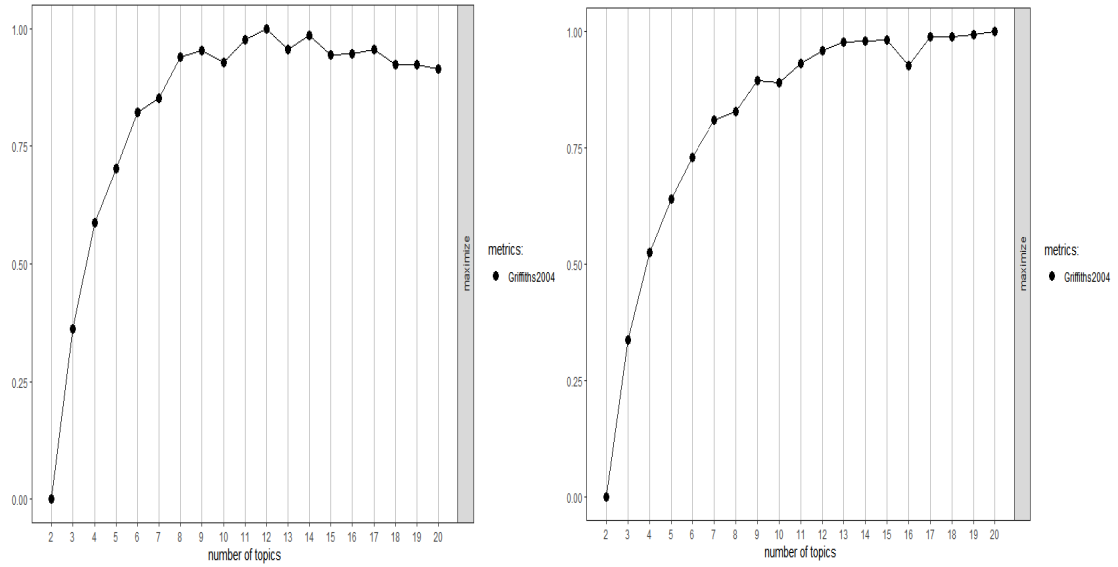
토픽 모델링은 방대한 텍스트 자료를 분석하여 숨겨진 토픽을 추출하는 알고리즘으로서[31], 문서와 단어로 구성된 행렬을 사용하여 문맥 속에 있는 토픽을 추론하는 방식이다[32]. 토픽 모델링의 종류에는 구조적 토픽 모형, 상관 토픽 모형, 잠재적 디리클레 할당 모형이 있다[32]. 본 연구에서는 이 가운데 가장 많이 사용되는 잠재적 디리클레 할당 LDA 분석을 활용한다.

LDA 분석은 토픽에 대해 단어가 등장할 확률을 디리클레 분포를 통해서 모델링 하는 방법이다[29]. 해당 분석은 다른 기법보다 결과 해석이 쉽고, 방대한 말뭉치를 요약하는데 도움을 준다[31]. 또한 문서자료, 텍스트 자료 등에서 유사한 의미를 가진 단어가 동시에 출현하는 확률기법을 활용하여 의미 있는 주제를 추론하는 모형이다[33]. 특히 많은 양의 빅데이터를 토대로 데이터가 내포한 핵심 주제를 도출하여 주요 개념의 분포를 비교·분석할 수 있다[34].

전처리 과정을 마친 후 DTM(Document-Term Matrix)을 이용하여 LDA 모델을 생성하였다. DTM은 행은 문서, 열은 단어로 구성된 것으로 빈도를 나타낸다. 본 연구에서 학생 창업 아이템은 문서 2,910개와 단어 2,548개를 분석하였고, 교원 창업 아이템은 문서 526개와 단어 784개를 분석하였다.

한편, LDA를 사용하기 위해서는 사전에 적절한 토픽의 수가 결정되어야 한다[28]. 최적의 토픽 수를 정하는 대표적인 방법은 이론을 바탕으로 연구자의 판단으로 토픽의 수(k)를 정하는 것이다[27]. 하지만 이러한 방법은 텍스트 자료에 대한 연구자의 전문성에 따라 판단의 한계가 있다[28]. 따라서 본 연구에서는 연구자의 재량적 판단에 의한 k값을 정하지 않고 텍스트 데이터를 통하여 최적의 k값을 정하는 방법을 사용하였다. 최적의 토픽 수를 정하기 위해 R의 ldatuning 패키지의 FindTopicsNumber() 함수를 활용하여 하이퍼파라미터 튜닝을 하였다.

토픽 수는 2에서 20까지 비교하였고, 반복 실행해도



(a) undergraduates students

(b) faculties

Fig. 3. Number of Topics

동일한 결과를 갖도록 난수를 고정하였다. 모델의 성능 지표인 Griffiths2004는 k값의 적합도를 나타낸다.

텍스트에 대한 설명력이 높을수록 Griffiths2004값이 크며, 성능 지표 그래프가 0~1중 1에 가까울수록 성능이 우수하다[28]. 하이퍼파라미터 튜닝 실시 후 학생 창업 아이템과 교원 창업 아이템의 최적 토픽 수는 Fig. 3에 제시된 바와 같이 각각 9개로 선정하였다.

4. 분석 결과

4.1 주요 키워드 분석

4.1.1 대학생 창업 아이템 주요 키워드

Table 2는 학생 창업 아이템에서 자주 사용된 키워드를 추출하고 TF-IDF 가중치를 적용하여 도출된 상위 10개 키워드를 나타내고 있다.

대학생 창업 아이템 가운데 빈도가 높은 키워드는 플랫폼, 전자상거래, 의류, 서비스, 디자인업, 쇼핑, 생활용품, 패션, 소프트웨어, 시스템 순으로 빈출하였다. 특히 플랫폼은 다른 키워드에 비하여 빈도수가 높은 것으로 보아 플랫폼과 관련된 창업 아이템이 많은 것을 알 수 있다.

반면 중요 키워드는 쇼핑, 음식, 사진, 인터넷, 굿즈, 브랜드, 안경, 애견용품, 가방, 생필품 순으로 추출되었다. 쇼핑을 제외한 나머지 키워드는 빈도 분석에서 등장하지 않았던 새로운 키워드임을 확인할 수 있다. 특히 TF-IDF가 가장 높게 나온 쇼핑이 학생 창업 아이템의 특징을 드러내는 주요 키워드라고 볼 수 있다. TF-IDF 키워드를 종합해 보면 일상에서 쓰이는 물품과 관련된 아이템이 중요하다는 것을 유추할 수 있다.

Table 2. Frequency & TF-IDF Comparison of Startup Items of Undergraduate Students

Keyword	frequency	Keyword	TF-IDF
Platform	269	Shopping	0.00710
E-commerce	118	Food	0.00327
Clothing	110	Photo	0.00270
Service	75	Internet	0.00225
Design Industry	67	Goods	0.00203
Shopping	63	Brand	0.00180
Household Goods	60	Glasses	0.00180
Fashion	53	Pet Supplies	0.00180
Software	51	Bag	0.00169
System	44	Daily Necessities	0.00169

4.1.2 교원 창업 아이템 주요 키워드

Table 3은 교원 창업 아이템에서 자주 사용된 키워드를 추출하고 TF-IDF 가중치를 적용하여 도출된 상위 10개 키워드를 나타내고 있다.

교원 창업 아이템 가운데 빈도가 높은 키워드는 치료제, 시스템, 인공지능, 소프트웨어, 신약, 빅데이터, 전기제품, 서비스, 플랫폼, 콘텐츠 순으로 빈출하였다. 특히 치료제의 빈도수가 가장 높은 것으로 보아 치료제와 관련된 창업 아이템이 많은 것을 알 수 있다.

반면에 중요 키워드는 신약, 줄기세포, 미생물, 세포치료, 유전자, 백신, 항생제, 파마바이오틱스, 액정, 살균제 순으로 추출되었다. 신약을 제외한 나머지 키워드는 빈도 분석에서 등장하지 않았던 새로운 키워드임을 확인할 수 있다. 특히 TF-IDF가 가장 높게 나온 신약이 교원 창업 아이템의 특징을 드러내는 주요 키워드라고 볼 수 있다. TF-IDF 키워드를 종합해 보면 대체로 의료와 관련된 아이템이 중요하다는 것을 유추할 수 있다.

Table 3. Frequency & TF-IDF Comparison of Startup Items of Faculties

Keyword	frequency	Keyword	TF-IDF
Medicine	31	New Drug	0.00758
System	22	Stem Cells	0.00292
Artificial Intelligence	21	Microorganism	0.00233
Software	13	Cell Therapy	0.00233
New Drug	13	Gene	0.00233
Big Data	12	Vaccine	0.00175
Electrical Appliances	12	Antibiotics	0.00117
Service	11	PharmaBiotics	0.00117
Platform	10	Display Screen	0.00117
Content	9	Sanitizer	0.00117

4.2 네트워크 분석

4.2.1 대학생 창업 아이템 네트워크 분석

네트워크 그래프에서 키워드의 노드가 다른 노드들과 빈번하게 연결되면 노드 크기가 크고, 노드의 크기가 작으면 노드와의 연결 빈도가 낮다는 것을 의미한다[35]. 또한 키워드 간의 관계가 가까운 노드는 같은 색으로 표현되어 있다[28].

Fig. 4에서 보는 바와 같이 동시 출현 네트워크 결과 키워드를 나타내는 노드 25개와 키워드를 연결하는 엣지 60개가 구성되어 있다.

노드의 크기가 가장 큰 플랫폼은 '딥러닝, 콘텐츠, 광고, 패션, 온라인, 모바일, 전자상거래, 빅데이터, 서비스' 등 키워드들과 함께 밀접하게 연결되어 있어 연결 중심성이 높은 키워드로 볼 수 있다.

키워드 간의 관계 빈도를 살펴보면 같은 색으로 표현된 '플랫폼, 딥러닝, 빅데이터, 여행용품'은 주로 컴퓨터 시스템에 관한 내용이다. 한편, '서비스, 스포츠, 어플리케이션, 광고'는 앱 서비스에 관한 내용으로 해석할 수 있다. '콘텐츠, 영상, 미디어, 사진'은 주로 영상과 사진 콘텐츠에 관한 내용이며, '의류, 패션, 잡화'는 주로 패션 중심의 내용으로 해석할 수 있다.

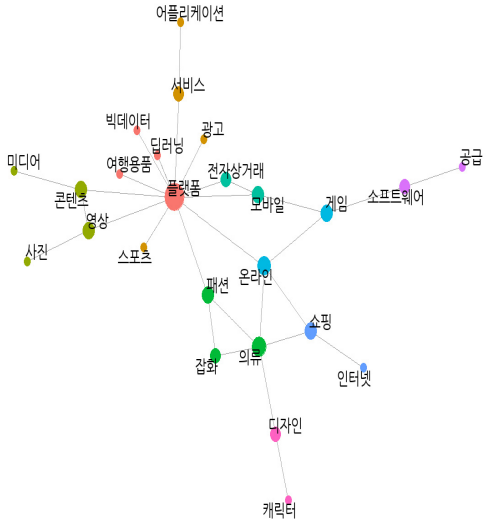


Fig. 4. Co-occurrence matrix of Student Startup Items

4.2.2 교원 창업 아이템 네트워크 분석

Fig. 5은 교원 창업 아이템의 동시 출현 네트워크 분석 결과, 키워드를 나타내는 노드 18개와 키워드를 연결하는 엣지 32개가 구성되어 있다.

노드의 크기가 가장 큰 치료제는 '세포, 근육, 백신, 단백질, 줄기세포' 키워드들과 밀접하게 연결되어 있어 연결 중심성이 높은 키워드로 볼 수 있다.

키워드 간의 관계 빈도를 살펴보면 '치료제, 근육, 백신, 줄기세포, 단백질, 세포'는 주로 의료용 단백질과 줄기세포에 관한 내용인 반면, '소프트웨어, 딥러닝, 서비스'는 주로 소프트웨어개발에 관한 내용으로 파악된다. 한편, '시스템, 광고, 플랫폼'은 주로 광고 시스템에 관한 내용이며, '인공지능, 빅데이터'는 주로 AI 기반 빅데이터에 관한 내용으로 볼 수 있다.

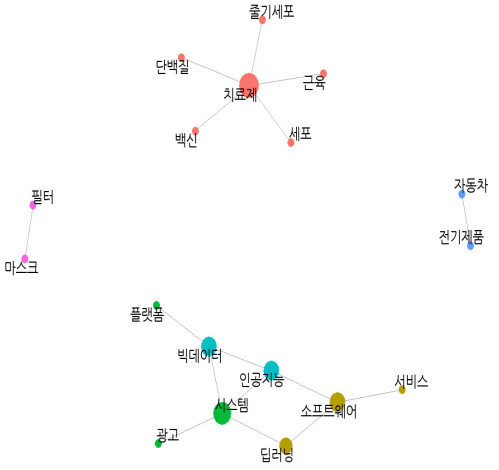


Fig. 5. Co-occurrence matrix of teacher Startup Items

4.3 LDA 토픽 분석

4.3.1 학생 창업 아이템 LDA 분석

LDA 분석을 통해 분석된 9개 토픽의 주요 키워드는 Table 4에 제시하였으며, 각 토픽별로 베타값이 높은 상위 5개 주요 키워드가 제시되어 있다. 베타값은 토픽마다 각 토픽에 단어가 등장할 확률을 나타내는 값이므로, 토픽별로 등장할 가능성이 높은 핵심어라고 할 수 있다[28]. 따라서 각 토픽에 등장할 확률이 높은 키워드를 통해 토픽의 특징을 알 수 있다.

Table 4. LDA Topic Model of Startup Items of Undergraduate Students

Topics	Keywords
Fashion platforms	System, platform, accessory, fashion, daily supplies
Product advertising	Advertising, software, bags, visual design, pet products
Online sales	Online, mobile, cosmetics, systems, food
E-commerce	e-commerce, shopping, characters, daily necessities, medical devices
Content creation and platform	Content, household goods, big data, e-commerce, gaming
AI-based services	Services, video, applications, pets, artificial intelligence
Software development	Software, clothing, sports, health, content
Clothing sales	Clothing, e-commerce, online, drone, goods
Design development	Design, platform, fashion, internet, household goods

첫 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 시스템(0.0401)이 가장 높고 플랫폼(0.0390), 액세서리(0.0370), 패션(0.0278), 생활용품(0.0257) 순으로 나타났다. 이를 바탕으로 첫 번째 토픽은 “패션 플랫폼”으로 명명하였다.

두 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 광고(0.0222)가 가장 높고, 소프트웨어(0.0167), 가방(0.0167), 시각디자인(0.0156), 애견용품(0.0123) 순으로 나타났다. 이를 바탕으로 두 번째 토픽은 “제품 광고”로 명명하였다.

세 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 온라인(0.0454)이 가장 높고, 모바일(0.0254), 화장품(0.0201), 시스템(0.0201), 음식(0.00958) 순으로 나타났다. 이를 바탕으로 세 번째 토픽은 “온라인판매”로 명명하였다.

네 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 전자상거래(0.163)가 가장 높고, 쇼핑(0.0528), 캐릭터(0.0214), 생필품(0.0123), 의료기기(0.00922) 순으로 나타났다. 이를 바탕으로 네 번째 토픽은 “전자상거래”로 명명하였다.

다섯 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 콘텐츠(0.0538)가 가장 높고, 생활용품(0.0404), 빅데이터(0.0280), 전자상거래(0.0249), 게임(0.0249) 순으로 나타났다. 이를 바탕으로 다섯 번째 토픽은 “콘텐츠 제작 및 플랫폼”으로 명명하였다.

여섯 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 서비스(0.0560)가 가장 높고, 영상(0.0397), 어플리케이션(0.0397), 반려동물(0.0224), 인공지능(0.0214) 순으로 나타났다. 이를 바탕으로 여섯 번째 토픽은 “AI 기반 서비스”로 명명하였다.

일곱 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 소프트웨어(0.0407)가 가장 높고, 의류(0.0385), 스포츠(0.0193), 헬스(0.0172), 콘텐츠(0.0161) 순으로 나타났다. 이를 바탕으로 일곱 번째 토픽은 “소프트웨어개발”로 명명하였다.

여덟 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 의류(0.0606)가 가장 높고, 전자상거래(0.0298), 온라인(0.0287), 드론(0.0171), 굿즈(0.0150) 순으로 나타났다. 이를 바탕으로 여덟 번째 토픽은 “의류판매”로 명명하였다.

아홉 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 디자인(0.0575)이 가장 높고, 플랫폼(0.0167), 패션(0.0167), 인터넷(0.0133), 생활용품(0.0133) 순으로 나타났다. 이를 바탕으로 아홉 번째 토픽은 “디자인 개발”로 명명하였다.

4.3.2 교원 창업 아이템 LDA 분석

Table 5는 LDA 분석을 통해 분석된 교원 창업의 9개 토픽의 주요 키워드를 나타내고 있다.

첫 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 치료제(0.0837)가 가장 높고, 프로그램(0.0201), 콘텐츠(0.0103), 의류(0.0103), 세포치료제(0.0103) 순으로 나타났다. 이를 바탕으로 첫 번째 토픽은 “치료제 및 프로그램”으로 명명하였다.

두 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 서비스(0.0398)가 가장 높고, 의료기기(0.0311), 시스템(0.0311), 반도체(0.0267), 에너지(0.0223) 순으로 나타났다. 이를 바탕으로 두 번째 토픽은 “의료서비스 시스템”으로 명명하였다.

세 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 빅데이터(0.0844)가 가장 높고, 플랫폼(0.0424), 유전체(0.0145), 미디어(0.0145), 마스크(0.0145) 순으로 나타났다. 이를 바탕으로 세 번째 토픽은 “빅데이터 기반 플랫폼”으로 명명하였다.

Table 5. LDA Topic Model of Startup Items of Faculties

Topics	Keywords
Treatments and Programs	Therapy, programs, content, clothing, cell therapy
Medical Service System	Services, medical devices, systems, semiconductors, Energy
Big Data-Based Platform	Big data, platforms, genomes, media, masks
Medical AI Software	Software, therapy, artificial intelligence, video, vaccine
AI-based medical system	Systems, artificial intelligence, therapy, content, medical devices
Medical proteins and stem cells	Systems, proteins, stem cells, cells, Medical Devices
Development of Cosmetic Materials	Cosmetics, electronics, health, genes, deep learning
Medication and medical device platform	Platforms, medicines, medical devices, content, battery
Development of new drugs	New drugs, food, platforms, filters, Microsoft

네 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 소프트웨어(0.0939)가 가장 높고, 치료제

(0.0742), 인공지능(0.0251), 영상(0.0152), 백신(0.0152) 순으로 나타났다. 이를 바탕으로 네 번째 토픽은 “의료용 AI 소프트웨어”로 명명하였다.

다섯 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 시스템(0.0765)이 가장 높고, 인공지능(0.0528), 치료제(0.0480), 콘텐츠(0.0337), 의료기기(0.0147) 순으로 나타났다. 이를 바탕으로 다섯 번째 토픽은 “AI 기반 의료 시스템”으로 명명하였다.

여섯 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 시스템(0.0462)이 가장 높고, 단백질(0.0279), 줄기세포(0.0234), 세포(0.0188), 의료기기(0.0142) 순으로 나타났다. 이를 바탕으로 여섯 번째 토픽은 “의료용 단백질과 줄기세포”로 명명하였다.

일곱 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 화장품(0.0462)과 전기제품(0.0462)이 가장 높고, 헬스(0.0279), 유전자(0.0188), 딥러닝(0.0188) 순으로 나타났다. 이를 바탕으로 일곱 번째 토픽은 “화장품 원료 개발”로 명명하였다.

여덟 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 플랫폼(0.0315)이 가장 높고, 의약품(0.0264), 의료기기(0.0264), 콘텐츠(0.0212), 배터리(0.0160) 순으로 나타났다. 이를 바탕으로 여덟 번째 토픽은 “의약품·의료기기 플랫폼”으로 명명하였다.

아홉 번째 토픽을 추출한 후 베타값이 높은 주요 키워드를 살펴보면 신약(0.0638)이 가장 높고, 식품(0.0346), 플랫폼(0.0200), 필터(0.0151), 마이크로소프트(0.0151) 순으로 나타났다. 이를 바탕으로 아홉 번째 토픽은 “신약 개발”로 명명하였다.

5. 결론

기업가적 대학의 역할이 강조되면서 대학생과 교원의 창업에 관한 논의는 활발하였으나, 대학 창업의 아이템에 대한 실증적 연구는 부재한 상황이다. 최근 COVID-19 이후 판매 아이템이 창업 기회와 많은 관련이 있다는 점에서 본 연구는 2020년부터 2021년 기간 동안 대학생과 교원의 창업 주요 아이템을 각각 분석하였다.

데이터 마이닝 기법을 활용하여 분석한 결과, 대학생 창업 아이템은 주로 의류, 컴퓨터와 앱을 활용한 온라인, 전자상거래, 쇼핑, 영상 등 플랫폼 기반 창업이 활발하게 이루어지고 있는 것으로 나타났다. 이와 달리 교원 창업 아이템은 치료제, 세포, 신약, AI 기반 의료 시스템 등 의료 제품·서비스 및 기술 중심의 창업이 이루어지고 있는 것으로 분석되었다. 즉, 교원 창업 아이템이 학생 창업 아이템에 비하여 높은 기술력과 전문성을 바탕으로 한 아이템이 많으며 더 많은 부가가치를 창출할 수 있음을 예상할 수 있다.

분석 결과를 토대로 제시할 수 있는 정책적 시사점은 다음과 같다. 첫째, 향후 창업을 준비하는 대학생 예비창업자를 위해 대학과 정부 차원에서의 컨설팅이 필요하다. COVID-19로 인해 서비스와 상품 중개 플랫폼이 급격하게 성장했지만, 플랫폼 사업이 대기업에게 유리하기 때문에 대학생 창업기업은 경쟁력이 부족할 가능성이 높다. 따라서 창업 준비 단계에서부터 차별화된 기술력과 아이디어를 겸비할 수 있도록 대학생 예비창업자들에 대한 컨설팅 지원이 필요하다.

둘째, 교원 창업의 경우 신약, 줄기세포, 미생물, 세포치료 등과 같은 바이오 분야의 기술창업이 활발하게 이루어지고 있으나, 연구 장비 등 창업을 위한 초기 비용이 많이 들 뿐만 아니라 시장 진출의 어려움이 존재한다. 따라서 바이오 창업을 고려하는 교원들을 대상으로 투자 유치, 인허가 규제, 절차 등에 대한 전문적인 지식과 경험이 있는 교육을 제공하는 것이 바람직하다.

마지막으로, 고부가가치를 창출할 수 있는 창업 교원에 대한 창업 친화적 인사제도를 확대하는 것이 필요하다. 교원 창업 아이템은 대부분 고기술·고부가가치 분야인 경우가 많기 때문에 유망한 창업 분야일 가능성이 높다. 그러나 앞서 살펴본 바와 같이 창업으로 인한 휴직제를 사용하는 교원은 많지 않다. 따라서 기술 기반의 창업 성공성을 확대하기 위해서는 휴직 또는 겸직 제도를 이용한 교원에 대해서 연구자 및 교육자로서의 의무를 상대적으로 유예할 수 있는 제도를 도입하거나, 창업 휴직으로 인한 업무상 공백을 해결할 수 있는 제도적 논의가 필요하다.

연구의 의의와 정책적 시사점에도 불구하고, 본 연구는 다음과 같은 한계점이 있다. 첫째, 연구의 시간적 범위가 2020년에서 2021년에 한정되어, COVID-19 상황의 특수성이 큰 편이라고 볼 수 있다. 둘째, 본 연구는 대학생과 교원의 창업에 있어서 일반 대학과 전문 대학을 구분하지 않았으나, 추후 대학의 설립 유형을 고려하여 분석하는 것이 바람직하다. 마지막으로 창업 중심대학, 창업 교육거점대학 등 창업을 위한 정책적 지원을 받은 대학과 그렇지 않은 대학의 창업 아이템을 비교하는 것이 필요하다.

References

[1] S. M. Nah, C. O. Kim, H. S. Lee, "A comparative study of the effect of university competence on technology transfer & commercialization and startups", *Journal of the Korean Society of Management Sciences,* pp.462-476, 2014.
DOI: https://doi.org/10.7232/JKIIE.2014.40.5.462

[2] Y. J. Lee, J. H. Kim, "Competitiveness Evaluation of Entrepreneurial Research Universities: A Comparative Analysis of Korean and Foreign Universities", *Innovation studies*, Vol.16, No.2, pp.25-59, 2021.

[3] J. W. Kim, "An Analysis on the Effects of University Capacity and Resources on the Professor Startups' Performance", *Journal of Korea Technology Innovation Society,* Vol.20, No.3, pp.642-663, 2017.

[4] M. S. Kim, Y. G. Kim, "Effects of Individual Trait, Educational, Organizational Factors on Entrepreneurial Intentions in Tourism College Students: the Mediating Effects of Entrepreneurial Self-Efficacy", *Journal of Foodservice Management,* Vol.15, No.5, pp.61-84, 2012.

[5] Report on the Results of the Survey on the Status of Closed Small Businesses, Korea Federation of Small and Medium Enterprises, pp.1-20.

[6] J. Y. Heo, From Learning to Industry...The Enhancing Trend of Professional Start-ups to Boost National Competitiveness, UNN, c2020 [cited 2020 November 09], Available From: http://news.unn.net/news/articleView.html?idxno=236660 (accessed July, 2022)

[7] J. M. Lee, G. J. Lee, "A Study on the Factors Associated with College Student Startups", *The Journal of Educational Research,* Vol.18, No.4, pp.173-190, 2020.
DOI: http://doi.org/10.31352/JER.18.4.173

[8] Y. J. Yoon, D. S. Park, "The Impact of the University's Capacity for the Industry-Academia Collaboration on the Performance of TechnologyCommercialization", *Journal of Social Science*, Vol.26, No.3, pp.157-177, 2015.

[9] H. Kim, "The Entrepreneurial Factors and Entrepreneurial Intention on Student", *Asia-Pacific Journal of Business Venturing and Entrepreneurship*, Vol.7, No.1, pp.263-271, 2012.
DOI: https://doi.org/10.16972/apjbve.7.1.201203.263

[10] D. W. Hong, I. H. Lee, J. S. Son, "A Study on the Effect of Support Services for Start-up Child Care Centers in Gyeonggi-do Province on the Performance of Resident Companies: Focusing on the Business Feasibility of Start-up Items", *Venture start-up research*, Vol.12, No.4, pp.47-60, 2017.

[11] H. Etzkowitz, C. Zhou, The Triple Helix: University Industry-Government Innovation and Entrepreneurship, Routledge, London, 2018.

[12] P. O'Shea, J. Allen, P. Morse, C. O'Gorman, F. Roche, "Delineating the anatomy of an entrepreneurial university: the Massachusetts institute of technology experience," *R&D Management*, Vol.37, No.1, pp.1-16, 2007.
DOI: https://doi.org/10.1111/j.1467-9310.2007.00454.x

[13] H. Etzkowitz, "Innovation Lodestar: The entrepreneurial university in a stellar knowledge firmament", Technological Forecasting and Social Change, Vol.123, pp.122-129, 2017.
DOI: https://doi.org/10.1016/j.techfore.2016.04.026.

[14] A. Y. Kim, University Startup Management Guide, Ministry of Education, Korea, pp.1-8.

[15] N. U. Park, A Study on the Selection of KISTEP Future Prospect Technology in 2021, A Study on Future Issues and Growth Strategy Planning, Korea, pp.1-170.

[16] H. J. Park, K. W. Lee, S. H. Jin, Coronavirus and Digital Transformation Accelerated NRF Issue Report 2020-13, Korea Research Foundation, Korea, pp.1-20.

[17] Mansoori Y, "Enacting the lean startup methodology: The role of vicarious and experiential learning processes", *International Journal of Entrepreneurial Behavior & Research*, Vol.23, No.5, pp.812-838, 2017.
DOI: https://doi.org/10.1108/IJEBR-06-2016-0195

[18] Ries. E, Lean Startup: How today's entrepreneurs use continuous innovation to create a fundamentally successful business, p.320, Crown Business Publisher, 2011, pp.1-320.

[19] J. L. Lee, College students are turning to start-ups in COVID-19 as they talk about youth start-ups in their 20s and 30s, Magazine Hankyung, c2021 [cited 2021 May 07], Available From: https://magazine.hankyung.com/job-joy/article/202105062713d (accessed July, 2022)

[20] J. Y. Kim, Research Report on Industry-Academic Cooperation Activities, Korea Research Foundation, Korea, pp.69-70.

[21] Start-up policy researcher, A Survey on Start-up Companies, Start-up Promotion Agency, Korea, pp.1-249.

[22] National Statistical Portal, Corporate life and death administration statistics, c2018 [cited 2018 July], Available From: http://kosis.kr (accessed July, 2022)

[23] M. J. Lee, Y. J. Joo, "The effect of corporate activities of technology-based start-ups on corporate growth", *Venture Business Research,* Vol.14, No.6, pp.59-76.
DOI: https://doi.org/10.16972/apjbve.14.6.201912.59

[24] S. Shane, Academic Entrepreneurship: University Spinoffs and Wealth Creation, p.334, Edward Elgar Publishers, 2004, pp.1-334.

[25] B. J. Koo, S. W. Kim, H. B. Kim, "Research on how to revitalize university teacher start-ups: Focusing on

conflict of interest regulations," *Technology Innovation Research*, Vol.26, No.1, pp.59-83, 2018.

[26] N. J. Jo, K. S. Han, Y. S. Lee, J. S. Park, H. J. Jung, A study on performance analysis and development direction for university start-up education and revitalization of start-up culture, Korea Research Foundation, Korea.

[27] S. C. Lee, "Topic Modeling of Newspaper Articles on Government 'Senior job program' via Latent Dirichlet Allocation", *Journal of Digital Convergence,* Vol.18, No.10, pp.537-546, 2020.
DOI: https://doi.org/10.14400/JDC.2020.18.10.537

[28] Y. W. Kim, (Do It!) Easy-to-Learn R Text Mining, p.344, Aegis Publishing, 2021, pp.344.

[29] J. Y. Kang, S. G. Kim, S. K. Roh, "A Topic Modeling Analysis for Online News Article Comments on Nurses Workplace Bullying", *Journal of Korean Academy of Nursing*, Vol.49, No.6, pp.736-747, 2019.
DOI: https://doi.org/10.4040/jkan.2019.49.6.736

[30] N. R. Kim, N. J. Lee, "An Analysis of Changes in Social Issues Related to Patient Safety Using Topic Modeling and Word Co-occurrence Analysis", *Journal of the Korean Society of Contents*, Vol.21, No.1, pp.92-104, 2021.
DOI: https://doi.org/10.5392/JKCA.2021.21.01.092

[31] D. M. Blei, "Probabilistic topic models", *Communications of the ACM*, Vol.55, No.4, pp.77-84, 2012.
DOI: https://doi.org/10.1145/2133806.2133826

[32] M. R. Jin, H. K. Ko, "Analysis of trends in mathematics education research using text mining", *Communications of Mathematical Education*, Vol.33, No.3, pp.275-294, 2019.
DOI: https://doi.org/10.7468/jksmee.2019.33.3.275

[33] Y. H. Yang, "Analysis on Types and Trends of Public Conflicts using Topic Modeling", *Local administrative research*, Vol.35, No.2, pp.159-188, 2021.
DOI: https://doi.org/10.22783/krila.2021.35.2.159

[34] S. J. An, G. B. Kim, J. S. Lee, S. C. Yeom, "A Topic Analysis on the Patent Developed by National Climate Technology R&D Project", *The Journal of Intellectual Property,* Vol.15, No.3, pp.293-332, 2020.
DOI: https://doi.org/10.34122/jip.2020.15.3.293

[35] S. J. Park, "A Topic Analysis of SW Education Textdata Using R", *Journal of The Korean Association of Information Education*, Vol.19, No.4, pp.517-524, 2015.
DOI: https://doi.org/10.14352/jkaie.2015.19.4.517

김 재 훈(Jae-Hoon Kim) [준회원]

- 2020년 8월 : 동아대학교 행정학과 (행정학사)
- 2021년 3월 ~ 현재 : 부산대학교 일반대학원 행정학과 석사과정

〈관심분야〉
공공정책, 산업기술

정 혜 진(Hyejin Jung) [정회원]

- 2007년 8월 : 성균관대학교 국정관리대학원 (행정학 석사)
- 2015년 12월 : Cleveland State University (행정학 박사)
- 2019년 9월 ~ 현재 : 부산대학교 행정학과 교수

〈관심분야〉
지역 경제, 창업, 기술혁신