

비정형 텍스트 데이터를 활용한 융합연구개발의 정책 동향 분석

Analysis of Policy Trends in Convergence Research and Development Using Unstructured Text Data

이지예 (Jiye Rhee)

경제 · 인문사회연구회¹⁾

신재은 (JaeEun Shin)

소상공인시장진흥공단²⁾

< 국문초록 >

현대사회의 다양하고 복잡한 사회적 문제를 해결하기 위해서는 융합적 시각의 문제 분석과 해결책 모색이 필요하다. 이러한 의미에서 융합연구의 중요성이 더욱 강조되고 있으며 효율성과 효과성 제고를 위한 전략적 접근이 요구되고 있다. 본 연구에서는 융합연구개발 활성화 기본계획의 텍스트 분석을 통하여 융합연구 정책의 시기별 특성과 시기별 토픽의 연관성과 변화추이를 확인하고 이를 통해 기본계획이 제시하는 정책적 변화를 이해하고자 하였다. TF-IDF 분석, 토픽 모델링(LDA), 그리고 네트워크(CONCOR) 분석 등 다양한 분석 방법을 활용하여 각 시기별 핵심 토픽을 도출하여 정책 변화 추이를 파악하였다. 분석결과, 시기별 토픽의 군집화와 토픽의 변화를 확인하였고 융합연구 생태계 조성 및 당면한 과제 해결을 위한 융합연구의 방향성을 제시하고 있다는 것을 파악할 수 있었다. 본 연구 결과는 향후 거시적 관점에서 융합연구개발 활성화 정책 수립 시에 정책의 변화에 대한 새로운 통찰을 제공함으로써 정부, 기업, 학계, 연구기관 등 다양한 이해관계자들에게 중요한 시사점을 제공할 수 있을 것으로 기대한다.

주제어: 융합연구개발 활성화 기본계획, 데이터 분석, 토픽모델링(LDA), 네트워크 분석, 융합연구

1) 제1저자, fraurhee99@gmail.com

2) 교신저자, topaz2033@gmail.com

1. 서론

현대사회의 문제들은 점점 더 복잡하고 다차원적인 특성을 보인다. 예를 들어, 기후변화, 공중보건, 도시화 문제 등은 단일 학문적 관점으로 해결하기 어려운 복잡성을 지닌다. 여러 분야의 지식을 통합하여 문제를 다각적으로 분석하고 이해하는 통합적 이해 과정은 더 종합적이고 효과적인 해결책을 마련하는 데에 도움이 된다. 또한 서로 다른 학문적 배경과 접근 방식이 결합되면 기존의 방법으로는 도달 할 수 없는 새로운 아이디어와 혁신적인 솔루션이 도출될 수 있다. 문제의 복잡성, 통합적 이해의 요구, 혁신 촉진을 통한 문제 해결이라는 측면에서 다양한 분야의 결합을 통한 융합연구는 필수적 접근법이라 할 수 있다. 최근, 정부, 기업, 학계, 연구기관 등의 다양한 이해관계자들은 융합연구를 적극적으로 지원하고 육성하는데 더욱 큰 노력이 필요하다는 점에 공감하고 있다(오지현 등, 2021). 특히 지속가능한 발전을 위한 다학제 간 융합연구의 중요성을 인식하고 모든 학문간 융합을 강조하며 지속가능한 지원을 통해 융합연구가 활성화 될 수 있도록 국가 차원에서 융합연구개발에 대한 투자를 확대하고 정책적 지원을 마련하는 등 융합연구 활성화를 위한 적극적인 전략과 투자를 추진하고 있다. 재정적 측면에서 미국은 2022년 회계 예산안 기준으로 63억 달러(약 7조원)의 예산을 투입하고 있으며 특히, 미국 국립과학재단(US National Science Foundation, NSF)의 Convergence Accelerator 프로그램의 경우 2024년에는 약 113억 규모로 확대하고 있는 추세이다(NSF, 2023). 2021년 제6기 과학기술 혁신 기본계획(2021-2025)을 발표하며 과학기술 뿐 아니라 인문과 학과 융합을 강조하며 5년간 정부연구개발 투자를 30조 엔으로 확대한 일본은 정부와 민간투자금을 합친 총연구개발비를 120조엔으로 설정하여 역시 융합연구 투자를 확대하는 전략을 수립하였다. 중국도 중공중앙의 국민경

제 및 사회발전 제14차 5개년 계획 및 2035년 장기목표에 대한 건의를 발표하며 기술 자립과 내수확대를 목표로 연구개발 투자를 매년 7% 이상 확대하는 전략을 발표하였다. 이렇듯 주요국 모두 정부 주도의 융합연구 활성화를 위한 예산 투자 규모를 확대하고 있다. 융합연구의 내용적 측면에서 살펴보면 미국은 국가 차원의 사회적 과제 해결을 위해 다중기관 융합연구를 최우선으로 지원하고 있으며 융합연구 가속화를 위해 Convergence Accelerator 프로그램을 운영하는 등 융합연구 생태계 확대를 목표로 하고 있다. 일본의 경우 과학기술기본법을 과학기술 혁신 기본법으로 개정하여 인문사회 분야를 포함한 분야 통합의 융합연구를 추진하고 있다. 중국의 경우도 학제간 융합을 변혁의 핵심 구동력이라고 강조하며 기초연구 10년 행동 방안을 마련하고 추진하고 있다(제4차 융합연구개발 활성화 기본계획, 2023). 우리나라도 2008년 국가융합기술 발전 기본계획 수립을 시작으로 융합연구를 활성화 하고 그 생태계를 확장하기 위한 노력을 하고 있다. 이렇듯 융합연구의 핵심은 지속적 지원과 협력을 통한 문제 해결이라고 할 수 있다. 이에 융합연구에 대한 미래전망 및 비전을 수립하기 위해서는 정부가 주도하고 국민의 의견이 반영되어 발표된 융합연구개발 활성화 기본계획을 분석하여 살펴 볼 필요가 있다. 이를 통해 정책적 변화와 그 변화의 흐름을 파악하는 것은 융합연구 분야의 기술적 간극, 공백 분야 등을 파악하여 지속가능한 융합연구 발전 전략, 정책을 수립하는데 통찰력을 확보하는 방법이라 할 수 있다. 융합연구에 대한 중요성과 관심을 반영하듯 최근 다양한 분야에서 융합연구의 현황 분석을 위한 연구가 수행되었다. 융합연구와 관련된 국내 선행연구에서는 융합이 효율적이고 경제성이 있는지 분석한 성과 분석 연구(노영희 등, 2018; 노영희, 신영자, 2023), 융합연구의 동향을 분석한 연구(노영희, 박종희, 2021), 융합 연구자의 학제간 관계를 분석한 연구(송인한, 김혜진, 2019) 등 지난 융합연구의 성과를 돌

아보는 결과론적 측면에서의 연구가 중점적으로 이루어졌다. 본 연구에서는 지속가능한 지원과 투자전략 수립을 위하여, 우리나라의 과거 15년간 융합연구 관련 계획의 변화를 분석하여 시사점을 제시하고자 한다. 제4차 융합연구개발 활성화 기본계획이 수립되었고 이를 기반한 국가연구개발 사업의 분야별 기획이 추진되고 있는 현재 시점에서 정책적 변화를 분석하고 방향성을 제시하는 것은 거시적 관점에서 융합연구 활성화 전략 수립을 위한 의미 있는 일이라 할 수 있다.

이에 본 연구에서는 비정형 텍스트 데이터를 활용하여 융합연구개발 활성화 기본계획을 분석하였고 융합연구개발 기본계획 수립을 통한 정책의 변화를 파악하여 향후 방향성을 제시하고자 다음의 연구 질문을 설정하고 연구를 추진하였다.

- 연구문제 1. 융합연구개발 활성화 기본계획의 시기별 주요 내용은 무엇이며, 시기별 차이점은 무엇인가?
 연구문제 2. 융합연구개발 활성화 기본계획을 통해 시기별 강조된 융합기술과 주요 쟁점은 무엇인가?

이와 같이 본 연구에서는 비정형 텍스트 데이터 분석을 통해 융합연구 분야의 통찰력을 확보하고, 향후 융합연구에 대한 학술적 및 실무적 시사점을 제공하고자 한다. 시기별 주요 융합연구의 변화를 파악함으로써 융합연구개발 활성화 전략의 간극을 식별하고 이를 바탕으로 지속가능한 융합연구 발전 전략을 수립하는데 기여할 수 있는 자료가 되기를 기대한다.

2. 선행연구 고찰

2.1. 융합연구

융합연구(Convergence Research)는 여러 학문 분야나 기술 영역의 지식과 기술을 융합하여 새로운 지식과 기

술을 창출하는 연구 방법론이다. 이는 문제를 효과적으로 해결하기 위하여 단일 학문 분야에 국한하지 않고 다양한 전문 분야를 통합하여 새로운 관점과 해결책을 찾는 것을 의미한다. 융합연구는 학문 분야 간의 경계를 허물고 상호 협력과 상호 작용을 통해 새로운 아이디어와 혁신을 끌어내는 것을 목표로 한다. 이를 통해 복잡한 현실 문제에 대한 창의적이며 효과적인 해결책을 제시하고 사회적, 경제적 가치를 창출한다. 융합연구는 이제 하나의 연구 분야로 자리매김하면서 국내외로 활발한 연구가 이루어지고 있다. 이는 다양한 학문 분야나 기술 영역을 결합하여 새로운 지식, 기술, 혹은 제품을 개발하는 것에서부터 사회문제 해결에 이르기까지 다양한 분야가 결합된 연구가 활발히 추진되고 있다는 것을 의미한다. 따라서 기존 융합연구 결과를 요약하고 선행연구의 한계를 파악하는 것은 매우 중요하다. 성과 분석과 함께 정책적 지향점을 설정하는 것은 융합연구의 발전과 적극적인 지원을 위해 필수적이다.

2.2. 융합연구 연구 현황

융합연구는 크게 서로 다른 학문 분야인 인문사회, 인공지능, 예술 등 분야를 융합하여 새로운 지식 발견, 복잡한 의사결정, 사회적 가치를 창출하는 학제간 융합연구와 이러한 융합연구에 대한 연구 동향을 분석한 연구로 구분된다. 후자의 연구는 연구 목적에 따라 융합연구의 효율성 분석을 통한 성과 분석 연구, 동향 분석, 융합연구 학자들 간의 네트워크 분석 연구로 구분할 수 있다.

첫 번째 성과 분석 연구는 융합연구 기관 단위로 분석한 연구와 사업을 분석한 연구가 있다. 성과 분석 연구는 융합연구 프로젝트의 성과를 다각도로 분석하고, 기존의 성과 평가 방법론을 보완하며 새로운 평가 프레임워크를 제안하는 연구이다. 조용래 등 (2017)은 정부출연연구소의 융합연구 성과를 분석하였다. 국가과학기술연구회 및

국가연구개발사업의 과제에 참여한 연구책임자를 대상으로 설문조사를 실시하고, 협업과 성과 간의 상관관계를 회귀분석으로 분석했다. 연구결과 협업의 수단, 대상, 구조, 내용에 따라 성과의 유형이 다양화되고 협업의 구체적 수준에 따라 세부 성과도 다르게 나타났다. 따라서 정책적으로는 융합연구의 특성을 고려한 사업구조 최적화와 단계별 융합연구투자 전략 마련이 필요하며 융합연구 대상 탐색과 주제 탐색을 위한 지원을 강화하여 실질적 성과를 창출하는 데 기여해야 한다는 시사점을 도출하였다. 노영희, 신영자(2023) 연구에서는 인문사회 기반 융합연구를 통한 학술 및 과학기술 성과를 분석하였는데 학술 및 과학기술 성과는 전문학술지 논문게재와 국내 국제 학술대회 논문발표가 대다수를 차지하였고, 과제 수 대비 논문게재 실적과 국제학술대회 초청강연 건수, 국내외 학술대회 논문 발표 성과, 보고서 발간은 감소하고 있는 것으로 나타났다. 주요 성과는 국제학술대회 초청강연과 전문학술지 게재가 주를 이루었으며 학제간 융합 연구지원사업의 성과가 주를 이루는 것으로 나타났다. 성과 분석 연구는 융합연구 기관 또는 사업의 성과를 분석하는 데 초점을 맞추었다. 그러나 융합연구는 다양한 기관과 사업을 통해 이루어지기 때문에, 특정 기관 또는 사업만을 대상으로 분석하는 것은 융합연구의 전체적인 성과를 파악하기 어려우며, 융합연구 성과에 영향을 미치는 다양한 요인을 고려하지 못했다는 한계점이 있다.

두 번째 동향 분석 연구는 관련 논문의 연구 주제, 발행 기관, 발행 연도, 연구 방법, 연구 대상 등을 통해 학계의 연구동향 파악을 위한 연구이다(권난주, 안재홍, 2012; 심영대, 조근태, 2020). 과학기술 환경의 융합연구의 동향을 분석하기 위하여 1996년부터 2015년까지 SCOPUS 데이터베이스에 구축된 다학제(multidisciplinary) 분야의 논문 정보를 수집하고, 그 가운데 상위 1%의 피인용 논문(HCP) 2,214편의 서지정보를 이용하여 네트워크 분석을

하였다(심영대, 조근태, 2020). 그 결과, 융합연구는 2000년대 초반부터 급격하게 증가하는 추세이며, 개인-기관-국가 간의 융합이 강화되고 있는 것으로 나타났다. 대학을 중심으로 진행되던 융합연구의 형태가 전문연구기관(Max Planck, RIKEN) 중심으로 전환되며, 융합연구를 목적으로 하는 별도의 기관(Broad Institute) 설립이 증가하는 것으로 분석되었다. 동향 연구는 융합연구의 전반적인 모습을 파악하는 데 유용한 방법이다. 그러나 특정 기간, 특정 데이터베이스, 특정 연구 분야를 기반으로 분석하기 때문에 연구 결과를 다른 기간, 데이터베이스, 연구 분야에 일반화하기 어렵다는 문제가 있다.

세 번째, 융합연구자의 학제 간 연결성의 현황과 특성을 이해하기 위한 연구자의 네트워크 분석 연구를 들 수 있다. 네트워크 분석은 융합 연구자들 간의 학제 간 관계를 네트워크로 모델링하여 분석하고, 연구 협력 네트워크의 구조와 특성을 파악하는 연구이다. 송인한, 김혜진(2019)은 국내 한 대학의 융합연구그룹 사례 분석하여 융합연구자 네트워크는 소수의 허브(공과대학, 의과대학) 중심으로 군집 형태로 이루어져 있고, 같은 학제 간 연결성이 높고, 인문 분야는 다른 학제와 연결 참여, 같은 캠퍼스 소속 단과대학(원)간 연결성이 높은 것으로 나타났다. 학제간 네트워크 분석연구는 융합연구자의 학제 간 연결성의 현황과 특성을 파악할 수 있으나 융합연구 정책 수립 및 지원 방안 마련에 직접적인 적용 어려움이 있다. 선행연구들은 주로 특정 학문 및 사업 등에 초점을 맞추어 진행되었다. 융합 선행연구는 연구 대상 및 방법에서 특정 대학 융합연구센터 사례 분석으로 연구 결과 일반화에 어려움이 있으며, 연구 결과 해석에 있어서 융합연구 성과에 영향을 미치는 다른 요인 고려가 부족하다는 점이 있다. 또한 융합연구 정책 수립 및 융합연구자 협력체계 구축 방안 제시에 구체성 부족하다는 점이 있다. 다양한 학문 분야 및 산업 분야 간의 연계를 강화하고, 미래 사회 문제에 대한 융합적 접근을 통해 국내의

〈표 1〉 융합연구개발 활성화 기본계획 개요

시기	구분	참여부처	비고
2008.11 (1시기)	제1차 융합연구개발 활성화 기본계획	교육과학기술부, 문화체육관광부, 농림수산식품부, 지식경제부, 보건복지가족부, 국토해양부, 환경부(7부)	국가융합기술 발전 기본계획('09~'13)
2014.2 (2시기)	제2차 융합연구개발 활성화 기본계획	기획재정부, 미래창조과학부, 교육부, 문화체육관광부, 농림축산식품부, 산업통상자원부, 보건복지부, 환경부, 해양수산부, 중소기업청, 특허청, 농촌진흥청(9부,3청)	창조경제 실현을 위한 융합기술 발전전략
2018.6 (3시기)	제3차 융합연구개발 활성화 기본계획	기획재정부, 교육부, 미래창조과학부, 문화체육관광부, 농림축산식품부, 산업통상자원부, 보건복지부, 환경부, 국토교통부, 해양수산부, 방위사업청, 문화재청, 농촌진흥청, 중소기업청, 기상청(10부, 5청)	-
2023.12 (4시기)	제4차 융합연구개발 활성화 기본계획	기획재정부, 교육부, 과학기술정보통신부, 행정안전부, 문화체육관광부, 농림축산식품부, 산업통상자원부, 보건복지부, 환경부, 국토교통부, 해양수산부, 중소벤처기업부, 식품의약품안전처, 방위사업청, 조달청, 농촌진흥청, 특허청, 기상청(12부, 1처, 5청)	-

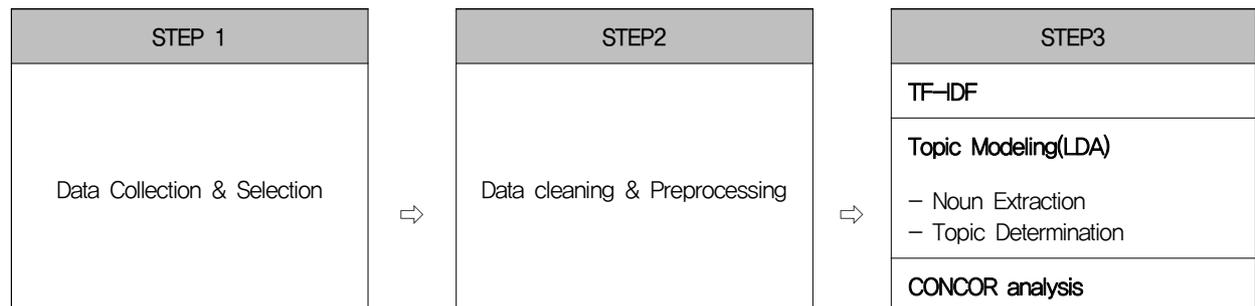
융합연구를 세계적 수준으로 발전시킬 방안을 모색해야 한다.

본 연구는 융합연구개발 활성화 기본계획을 바탕으로 융합연구개발 사업 관련 정책 변화를 분석하고, 거시적인 관점에서 향후 융합연구의 방향성을 제안하는 것을 목표로 하였다. 이를 위하여 정책 동향 변화를 분석함으로써 융합연구의 발전에 필요한 정책적 지향점과 거시적 방향성 설정을 제안하고자 한다. 연구 분석 방법은 텍스트 분석을 통해 키워드 분석, 토픽모델링, 네트워크 분석 기법 활용하여 융합연구개발 정책의 변화와 특성을 심층적으로 분석하였고, 제4차에 걸친 융합연구개발 활성화 기본계획의 정책 변화 추이를 파악하여 향후 정책 발전 방향을 제시함과 동시에 융합연구 정책 발전 및 효과성

제고를 도모하고자 한다.

3. 연구방법

본 장에서는 융합연구개발 정책의 변화와 미래 전망을 분석하기 위해 각 기본계획의 주요 내용, 추진 전략, 예산 규모 등을 분석하여 융합연구개발 정책의 변화 추이를 파악하고, 다양한 텍스트 분석 방법을 통하여 문서내의 패턴, 관계, 규칙 등을 탐색하는 텍스트 마이닝의 수행은 Python 3.11.7을 활용하였다.



〈그림 1〉 연구의 절차

3.1. 연구 절차 및 분석 방법

본 연구의 절차는 연구는 데이터 수집(시기별), 데이터 전처리(시기별), 데이터 분석(통합, 시기별) 순으로 진행하였다(<그림 1> 참조). 분석 대상은 국가융합기술 발전 기본계획(제1차 융합연구개발 활성화 기본계획)부터 총 4차의 융합연구개발 활성화 기본계획으로 하였다(<표 1> 참조).

<그림 1>의 STEP 2에 제시한바와 같이 분석에 활용한 데이터는 2008년 11월 18일에 발표된 1차 국가융합기술 발전 기본계획 부터 2023년 12월 20일 발표된 제 4차 융합연구개발 활성화 기본계획(안) 4개의 PDF 형태로 수집된 데이터를 먼저 텍스트 데이터로 변환하는 작업을 수행하였다. 텍스트 데이터로 변환할 때 활용한 라이브러리는 PDFResourceManager, PDFPageInterpreter이며 이를 통해 각 PDF의 텍스트 데이터를 추출 후 문장 단위로 변환시키기 위해 ‘.’을 기준으로 분리하였다.

데이터 분석에 활용된 문서(단어)의 정제 방법은 먼저 re 패키지를 활용하여 한글, 영문을 제외한 단어들을 제거하고 해당 내용들을 맞춤법 검사기(부산대)를 통해 1차 정제 수행했다. 이후, 아래 분석기법을 적용하기 위해 한 글자 단어를 제거하였고 konlpy의 Okt를 사용하여 의미를 가지고 있는 명사를 추출하였다. 마지막으로 상위 200 위 이내 단어빈도를 Python collections 모듈의 Counter()

함수로 구하여 총 3회에 걸쳐 수정 및 불용어 처리로 2차 정제를 수행하였다(<표 2> 참조).

<그림 1>의 STEP 3에 제시한 바와 같이 융합연구개발 활성화 기본계획의 시기별 주요 내용과 차이점을 분석하기 위하여 TF-IDF로 단어의 가중치 측정하는 방법을 활용하였다. TF-IDF는 문서 내에서 특정 단어의 빈도를 전체 문서군의 단어 출현빈도로 나눈값이다. TF-IDF 값을 통해 특정 단어가 문서 내에서 얼마나 중요한지 파악이 가능하다(Park et al., 2016).

다음으로 융합연구개발 활성화 기본계획의 시기별 강조되는 융합기술과 주요 쟁점을 파악하고자 토픽모델링과 CONCOR 분석을 수행하였다. 단어나 문서들의 집합에 대해 숨겨진 주제를 찾아내어 문서나 키워드별로 주제끼리 묶어주는 토픽모델링으로 Python의 gensim 라이브러리를 활용하여 토픽별 주요단어와 문서들이 분류된 토픽비율을 구현하였다(Blei et al., 2003). CONCOR (CONvergence of iterated CORrelations(김해원, 전채남, 2014) 분석은 단어 간 연관성이 높게 군집된 정도를 파악할 수 있는 방법으로 CONCOR분석을 수행하기 위하여 수행에 필요한 상위 100위 이내 단어 X 단어 매트릭스를 Python을 통해 각 문단 단어 동시출현 빈도로 산출하였고 생성된 매트릭스를 Ucinet6.0을 통해 시각화하였다

<표 2> 데이터 분석에 활용된 문서(단어) 전처리 전·후 비교(단어수)

시기	구분	전처리 전	1차 전처리	전처리 후
2008.11 (1시기)	제1차 융합연구개발 활성화 기본계획(76페이지)	48,622	10,655	10,652
2014.2 (2시기)	제2차 융합연구개발 활성화 기본계획(43페이지)	32,756	7,006	7,005
2018.6 (3시기)	제3차 융합연구개발 활성화 기본계획(78페이지)	64,171	11,687	11,644
2023.12 (4시기)	제4차 융합연구개발 활성화 기본계획(75페이지)	80,440	17,882	17,875

4. 분석 및 결과

융합연구개발 활성화 기본계획의 내용을 네트워크 분석 기법을 활용하여 분석하고, 시기별 변화를 파악하여 향후 융합연구개발 정책 방향성을 제시하고자 하였다. 분석 자료는 텍스트 전처리를 통해 불필요한 단어를 제거하고 분석에 필요한 단어를 선별하였다. 다음으로, 선별된 단어들의 연결성을 분석하여 네트워크를 구축하였다. 연구의 분석 순서는 첫 번째 TF-IDF 분석을 통해 시기별 주요 강조 사항 도출하였다. 두 번째로 토픽모델링(LDA) 분석을 하여 융합정책의 주요 토픽을 파악하고 토픽의 변화를 분석하였다. 마지막으로 네트워크 분석을 실행하여 시기별 융합연구개발 정책의 특징과 연관성을 분석하였다. 본 연구에서는 전체적인 분석 후 TF-IDF와 분석결과와 토픽모델링 분석 결과, 네트워크(CONCOR) 분석결과를 중심으로 연구결과에 대해 논의하고자 한다.

4.1. TF-IDF 분석 결과

TF-IDF 분석 결과 융합연구개발 활성화 기본계획의

1시기에서는 추진에 대한 강조, 2시기에서는 전략에 대한 강조, 3시기에서는 융합에 대한 강조, 4시기에서는 연구에 대해 강조하는 것으로 도출되었다(<표 3> 참조). 1시기에서는 융합기술 개발, 2시기에서는 조원이라는 단어가 등장하여 예산이 많이 투입됨을 확인하였다. 3시기에서는 기술에 대한 투자, 4시기에서는 여러 분야별 연구를 접목하여 미래 지향적 연구를 추진하는 방향을 강조하는 것으로 도출되었다.

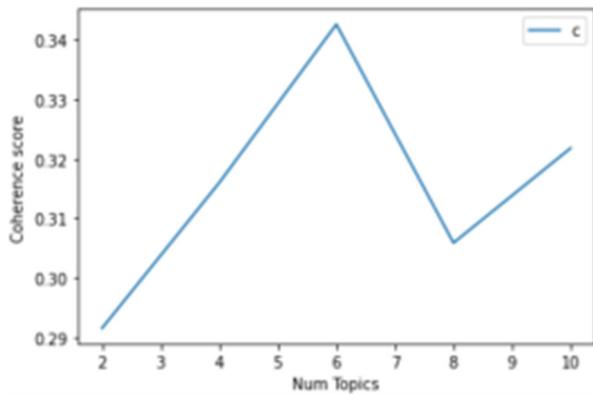
4.2. 토픽모델링(LDA) 분석결과

LDA 토픽모델링 분석에 앞서 최적의 토픽 개수를 추출하기 위한 과정을 진행하였다. LDA 적정 토픽수를 결정하는데 자주 활용되는 응집도(coherence)를 기준으로 도출하였다. 최적의 모델을 선정하기 위해 토픽 K를 가정하고 응집도 점수를 활용하여 토픽수를 산정하였다(<그림 2> 참조). LDA 토픽모델링에서 K가 6일 때 응집도 지수가 가장 최대치로 나타나 토픽수를 6개로 산정하였다.

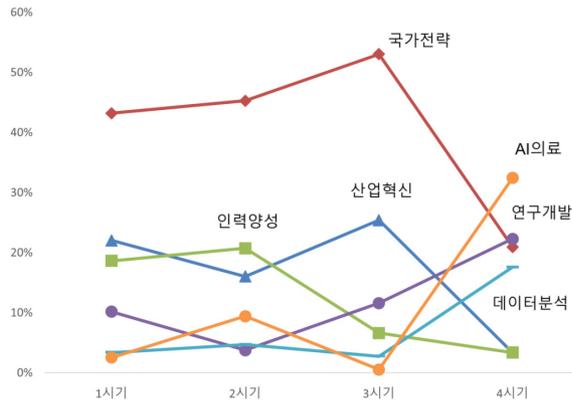
〈표 3〉 TF-IDF 분석결과표(상위50위)

순위	1시기		2시기		3시기		4시기	
	단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
1	기술	17,368	기술	14,262	역원	29,313	융합	15,858
2	융합	15,003	융합	13,372	사업	19,418	기술	13,837
3	추진	8,107	전략	7,893	융합	12,533	연구	12,121
4	개발	6,299	조원	7,638	기술	12,112	미래	8,183
5	산업	6,222	확대	6,799	투자	10,330	추진	7,093
6	계획	5,911	국가	5,691	개발	9,702	분야	6,581
7	연구	5,729	개발	5,079	과제	9,313	전략	5,669
8	사업	5,550	연구	4,524	계획	8,381	평가	5,003
9	지원	4,890	평균	4,098	연구	8,248	사회	4,775
10	분야	4,784	계획	3,924	증감	8,135	개척	4,311
11	부처	4,640	추진	3,918	단위	8,043	달성	4,281

순위	1시기		2시기		3시기		4시기	
	단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
12	국가	4.453	부처	3.886	백만원	7.965	목표	4.175
13	기본	3.998	인력	3.745	추진	7.381	계획	4.161
14	인력	3.959	사업	3.730	증가	7.050	체계	4.125
15	전략	3.834	선정	3.455	중점	6.667	기본	3.782
16	수립	3.796	투자	3.047	연구개발	6.647	사업	3.704
17	위원회	3.561	발전	2.960	부처	6.415	연구개발	3.529
18	강화	3.464	사회	2.830	지원	5.873	개발	3.391
19	발전	3.164	기본	2.825	실적	5.853	성과	3.373
20	목표	3.111	증가	2.665	산업	4.955	수립	3.347
21	현황	3.083	산업	2.637	년도	4.579	국가	3.334
22	양성	2.971	창조경제	2.575	기반	4.330	과제	3.116
23	공동	2.794	양성	2.436	성과	4.279	정책	3.025
24	기반	2.790	지원	2.384	구축	4.265	도전	3.019
25	육성	2.628	미래부	2.355	과학기술	4.049	연구자	3.005
26	발굴	2.608	실현	2.217	혁신	4.022	인공	2.990
27	과학기술	2.567	원천	2.155	감소	4.005	과학기술	2.907
28	바이오	2.563	창의	2.114	기본	3.911	수준	2.885
29	예산	2.542	미래	2.100	국가	3.894	활성화	2.853
30	정의	2.524	목표	2.096	활성화	3.876	기반	2.821
31	원천	2.507	과학기술	2.085	등록	3.573	지원	2.622
32	전문	2.504	예측	2.053	수립	3.484	시스템	2.607
33	투자	2.495	인재	2.044	출원	3.286	환경	2.529
34	규모	2.486	분야	2.024	활용	3.181	지능	2.512
35	구축	2.444	공청회	2.000	정책	2.919	방향	2.484
36	나노	2.383	수요	1.929	개요	2.908	협력	2.449
37	체계	2.348	관리	1.872	동향	2.837	혁신	2.422
38	역원	2.277	평가	1.870	분야	2.776	이상	2.148
39	유형	2.273	참고	1.870	계재	2.593	플라스틱	2.132
40	환경	2.196	방안	1.806	도전	2.476	분석	2.101
41	서비스	2.162	연구개발	1.767	미래	2.463	정통부	2.085
42	협력	2.122	정보	1.687	목표	2.454	가치	2.049
43	명단	2.104	관계	1.685	특허	2.446	우주	2.044
44	정부	2.102	상정	1.645	관리	2.403	강화	2.041
45	한국	2.081	향후	1.629	협력	2.379	확대	2.017
46	실무	2.036	산업부	1.627	정보통신부	2.341	폐기물	1.950
47	창조	2.028	인프라	1.556	플랫폼	2.296	감소	1.948
48	국내외	1.983	육성	1.549	국내	2.274	의료	1.926
49	사회	1.976	미국	1.528	데이터	2.240	경계	1.918
50	확보	1.935	나노기술	1.500	통상	2.225	생태계	1.913



〈그림 2〉 Coherence 그래프



〈그림 3〉 시기별 토픽 비율 변화

LDA 토픽모델링 분석결과를 살펴보면 산업혁신, 국가전략, 인력양성, 연구개발, 데이터 분석, AI의료 등 6개의 토픽에 따른 키워드를 도출하였으며, 도출된 키워드를 중심으로 6개의 토픽을 정하였다. 산업혁신 토픽에서는 해당 단어와 관련된 원천, 산업, 나노, 바이오, 소재,

환경 등의 단어가 도출되었고, 국가전략 토픽에서는 국가전략과 관련된 추진, 계획, 국가, 전략, 수립, 목표, 과제 등의 단어가 나타났다. 이와 마찬가지로 인력양성과 관련된 인력, 양성, 육성, 대학, 교육, 전문 등이, 연구개발 토픽에서는 관련된 연구, 지원, 과제, 성과, 연구자, 연구개발 등이, 데이터분석과 관련된 평가, 문제, 분석, 정보, 플랫폼, 데이터 등이 도출되었으며, 마지막으로 AI 의료 서비스와 관련된 지능, 인공, 예측, 첨단, 치료, 건강 등이 주요 토픽단어로 추출되어 4차에 걸친 융합연구개발에 대한 트렌드를 키워드를 통해 전체적으로 살펴볼 수 있었다(<표 4> 참조).

토픽모델링 결과를 각 시기별 쟁점을 파악하고자 분석을 수행하였다. 각 토픽별 비율을 보게 되면 기술융합과 관련된 사업이 정부 주도하에 진행이 되므로 관련된 토픽들의 비율이 다른 토픽들에 비해 매우 높게(40% 이상) 나타나고 있음을 확인하였다. 또한, 타 토픽의 경우 11~17% 이내 거의 유사하게 도출되어 국가전략 이외의 토픽이 고르게 분포되고 있음을 파악하였다. <그림 3>과 같이 시기별 토픽변화를 살펴보면 1, 2, 3시기에 국가전략의 토픽 비율이 높게 나타났고, 4시기에는 AI 의료 비율이 높게 나타났다. 이와 마찬가지로, 연구개발과 데이터 분석이 다른 1, 2, 3시기와 비교하면 4시기 비율이 증가하는 것으로 나타났다. AI 의료, 데이터 분석, 연구개발 등이 융합연구에서 중요성이 커졌음을 알 수 있다. 이와 반대로 산업혁신과 인력양성의 경우, 최근 감소하는 모습을 보여 영향력이 감소한 것으로 나타났다.

< 4 >

토픽	키워드
산업혁신	기술, 개발, 사업, 지원, 구축, 원천, 실적, 계획, 서비스, 산업, 나노, 바이오, 소재, 핵심, 환경
국가전략	추진, 계획, 국가, 전략, 투자, 발전, 수립, 연구개발, 정책, 과학기술, 정부, 중점, 증가, 목표, 과제
인력양성	기술, 산업, 사업, 인력, 양성, 육성, 대학, 교육, 발굴, 수요, 전문, 산업부, 문화, 강화, 과학
연구개발	연구, 지원, 과제, 성과, 협력, 체계, 혁신, 구축, 연구자, 연구개발, 글로벌, 인재, 활용, 과학기술, 정통부
데이터분석	미래, 평가, 연구, 대응, 문제, 전략, 해결, 국가, 분석, 정보, 플랫폼, 확보, 개척, 국민, 데이터
AI의료	에너지, 지능, 인공, 예측, 첨단, 실현, 치료, 구현, 차세대, 목표, 인간, 통신, 건강, 관리, 로봇

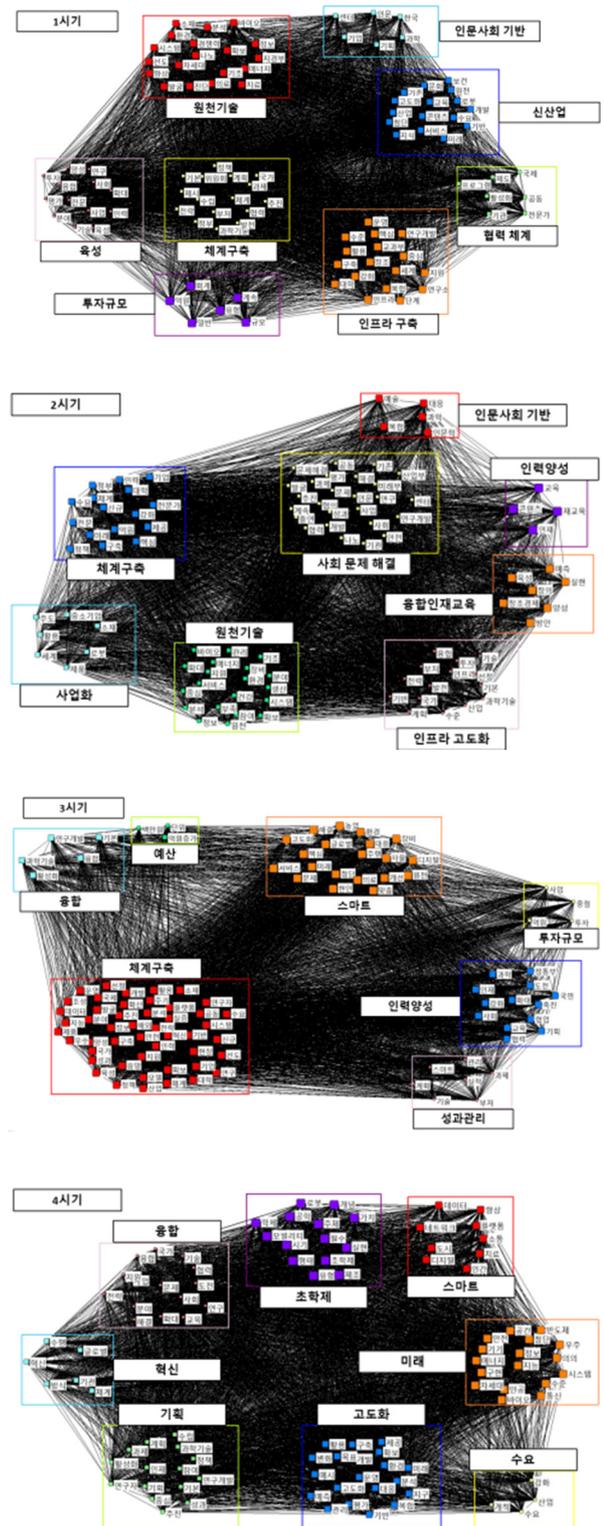
4.3. 네트워크(CONCOR) 분석결과

네트워크(concor) 분석 시각화 결과는 다음과 같다 (<그림 4> 참조). 1시기에서는 ‘원천기술(바이오, 나노, 소재, 에너지 등)’에 대한 군집이 가장 높았으며, ‘인프라 구축(연구소, 교과부, 복합, 연구개발 등)’의 군집이 높게 도출되었다. 다음으로 ‘신산업(콘텐츠, 로봇, 서비스, 첨단, 산업 등)’이 높게 도출되었고, ‘체계구축(정부, 국가, 위원회, 부처 등)’이 ‘육성(기술, 사업, 연구, 분야, 확대 등)’이 도출되었다. ‘협력체계(국제, 공동, 전문가 등)’이 ‘투자규모(역원, 유형, 회계 등)’이 마지막으로 ‘인문사회 기반(인문, 과학, 센터 등)’순으로 도출되었다.

2시기에서는 ‘사회문제 해결(기관, 미래부, 산업부, 공동, 문제해결 등)’이 가장 높게 군집되었으며, 다음으로 ‘체계구축(정부, 인력, 전문, 정책 등)’이 높게 군집되었다. ‘원천기술(바이오, 에너지, 서비스, 환경 등)’, ‘인프라 고도화(융합, 투자, 과학기술, 인프라 등)’, ‘융합인재교육(창조경제, 양성, 방안 등)’, ‘사업화(중소기업, 소재, 로봇 등)’, ‘인문사회 기반(예술, 인문학, 과학 등)’, 그리고 마지막으로 ‘인력양성(교육, 재교육 등)’순으로 도출되었다.

3시기에서는 ‘체계구축(양성, 조성, 연구자, 기업, 선도, 산업 등)’이 가장 크게 군집되었고, ‘스마트(고도화, 장비, 디지털, 의료, 맞춤형 등)’가 그 다음으로 크게 군집되었다. ‘인력양성(과학, 인재, 교육, 협력 등)’이 ‘융합(연구개발, 과학기술, 활성화 등)’, ‘성과관리(실적, 과제, 계획, 관리 등)’, ‘투자규모(역원, 중점, 투자 등)’, 마지막으로 ‘예산(백만원, 역원증가 등)’순으로 군집되었다.

4시기에서는 ‘고도화(환경, 예측, 평가, 고도화 등)’가 가장 높게 군집되었으며, 다음으로는 ‘미래(반도체, 에너지, 시스템 등)’가 높게 도출되었다. ‘초학제(모빌리티, 학제, 초학제 등)’, ‘기획(인재, 연구자, 성과, 수립 등)’, ‘융합(지원, 사업, 분야, 협력 등)’, ‘스마트(플랫폼, 디지



<그림 4> 네트워크(CONCOR) 분석 시각화

털, 치료 등)', '혁신(글로벌, 기관, 수행 등)'이 도출되었고, 마지막으로 '수요(산업, 개척, 강화 등)'가 도출되었다.

5. 결론

5.1. 결과 토의

본 연구는 2008년부터 다부처 간 협력에 따라 추진되어 온 융합연구개발 활성화 전략의 변화에 대해 4-6년의 주기를 두고 기획, 추진된 기본계획의 주요 내용을 분석한 것이다. 우선 국가과학기술위원회에서 확정된 '국가융합기술발전 기본방침'의 후속조치로 제1차 융합연구개발 활성화 기본계획을 수립하고 이에 따른 시행계획을 수립하여 국가연구개발사업에 반영하였다. 1차부터 부처 간 연계 및 협력의 고도화에 기여하고자 수립된 정책으로 차수가 지날수록 융합연구 생태계를 조성하고 확립하기 위한 방향성과 추진체계 등 다양한 전략을 제시하고 있다. 본 연구에서는 융합연구개발 활성화 기본계획의 시기별 특성과 쟁점사항을 분석하고 종합적 시각에서 그 결과를 해석하고자 하였다. 연구 결과를 요약하면 다음과 같다.

첫째, TF-IDF 분석 결과를 통해 각 시기별 주요 키워드를 확인할 수 있었다. 각 시기별로 창조적 융합기술 선점, 융합기술 확보와 고도화, 다양한 융합 시도와 노력, 기존 과학기술로만 해결할 수 없는 복합문제에 관한 연구 등의 주요 키워드가 도출되었다.

둘째, 토픽모델링(LDA) 분석 결과를 통해 6가지 주요 토픽이 도출되었고, 국가 전략과 관련된 토픽이 다른 토픽들에 비해 매우 높은 비율로 나타났다. 이를 통해 국가가 기술융합과 관련된 사업을 주도하고 있음을 확인할 수 있었다.

셋째, 네트워크(CONCOR) 분석 결과를 통해 시기별로 원천기술, 인프라 구축, 신산업, 체계구축, 육성, 협력 체계 등의 군집이 도출되었고, 각 시기별로 강조되는 요소들이 어떻게 변화하는지를 확인할 수 있었다.

이러한 결과들을 종합적으로 고려할 때, 융합연구개발 활성화 기본계획은 시기별로 다양한 변화와 발전을 거듭하고 있으며, 이는 미래 지식 기반 사회를 이끌어 가는 중요한 역할을 한다는 점을 재확인할 수 있었다. 따라서 이러한 분석 결과를 토대로 융합연구개발 활성화 기본계획의 효율성을 높이고 미래 지식 기반 사회의 발전에 기여할 수 있는 방안을 모색해 나가는 것이 중요하다. 특히 본 연구에서는 통합적이며 거시적 관점의 우리나라 융합연구의 정책적 방향을 살펴보았다는 것에 그 의의가 있다.

5.2. 학술적 및 실무적 시사점

본 연구의 학술적이며 실무적 시사점은 다음과 같다. 먼저 본 연구에서의 학술적 시사점은 그간 축적되고 누적된 융합연구 관련 거시적 관점의 융합연구개발 활성화 기본계획이 제시하고 있는 정책적 변화를 주관적인 내용 분석이 아닌 텍스트 분석을 적용하였다는 점이다. 방대한 내용의 융합관련 기본계획을 군집화 하여 시기별 정책 변화를 한눈에 볼 수 있도록 정리하여 정책을 연구하는 연구자에게 융합연구 정책 변화를 통한 향후 연구방향을 제시했다는 점에서 그 의의가 있다. 실무적 시사점으로는 융합연구 관련 정책을 수립하고 국가연구개발사업을 기획하는 정책 입안자 등에게 융합연구의 정책적 변화 추이와 통합적 시각이 분석결과를 제공함으로써 융합연구개발 활성화 기본계획의 구조적 이해 제고를 도모하였다.

5.3. 연구의 한계점 및 제언

본 연구는 텍스트 분석을 통해 융합연구개발 활성화 기본계획의 변화를 비교 분석하여 시사점을 도출하고자 하였으나 다음과 같은 한계점이 있다. 제도화되지 않은 융합연구개발 활성화 기본계획만을 활용한 분석으로 향후 각 시기별 시행계획에 따른 융합연구개발 사업의 투자 분석과 추진 과제별 성과 분석이 필요할 것이다. 향후 연구에서는 각 시행계획에 따른 융합연구개발 사업의 투자 분석과 연구개발 과제의 성과 평가 그리고 시사점을 도출하여 향후 융합연구개발 활성화를 위한 보다 거시적인 방향의 기본계획 수립에 이바지할 수 있기를 기대한다.

〈참고문헌〉

[국내 문헌]

1. 오지현, 신현우, 전은진, 이천환, 정현덕 (2021). 문제해결형 융합 연구를 위한 디지털 플랫폼 구축 동향과 시사점. **GTC BRIEF**, 2(3), 1-16.
2. 도계훈 (2021). 일본의 제6기 「과학기술·혁신기본계획」 주요내용과 시사점. **KISTEP 정책 브리프**, 4, 1-3.
3. 권난주, 안재홍 (2012). 융합 및 통합 과학교육 관련 국내 연구 동향 분석. **한국과학교육학회지**, 32(2), 265-278.
4. 김근현, 정성훈, 양재동, 위주연 (2023). 빅데이터 분석기법을 활용한 10년간의 공공데이터 관련 정책 연구: 정관별 비교분석을 중심으로. **국가정책연구**, 37(4), 45-67.
5. 노영희, 이광희, 정대근 (2018). 인문사회기반 융합연구 성과 영향 요인 연구. **한국콘텐츠학회논문지**, 18(6), 667-678.
6. 노영희, 박종희 (2021). 학제간 융합연구 트렌드 분석에 관한 연구. **인문사회** 21, 12(1), 3359-3374.
7. 노영희, 신영지 (2023). 국내 인문사회기반 융합연구의 학술 및 과학기술 성과 분석에 대한 연구: 한국연구재단 학제간융합연구지원사업과 일반공동연구지원사업(융복합연구)을 중심으로. **문화외용합**, 45(4), 989-999.
8. 송인한, 김혜진 (2019). 융합연구자 학제 간 연결성에 관한 네트워크 분석: 대학 내 융합연구그룹의 사례. **한국융합학회논문지**, 10(4), 153-163.
9. 심영대, 조근태 (2020). 네트워크분석을 활용한 다학제 분야의 융합연구 동향 분석. **기술혁신학회지**, 23(3), 505-528.
10. 조용래, 우청원, 최중화 (2017). 국가 출연연구소의 협업적 융합연구 성과 분석. **한국기술혁신학회 학술대회**, 20(4), 1089-1121.
11. 김해원, 전채남 (2014). 빅데이터를 활용한 콘텐츠 제작방안에 관한 탐색적 연구: TV홈쇼핑을 중심으로. **사이버커뮤니케이션학보**, 31(3), 5-51.

[국외 문헌]

12. Cavnar, W. B., & Trenkle, J. M. (1994). N-gram-based text categorization, *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval* (Vol. 161175, pp. 161-175).
13. Park, H., Lee, M., Hwang, S., & Oh, S. (2016). TF-IDF

based association rule analysis system for medical data, *KIPS Transactions on Software and Data Engineering*, 5(3), 145-154.

14. Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation, *Journal of Machine Learning Research*, 3, 993-1022.

[URL]

15. 한국과학기술기획평가원 (2021). 중구 <14.5 구획> 과학기술 정책방향과 시사점. https://www.kistep.re.kr/gpsIssueView.es?mid=a30101000000&list_no=48678&nPage=8186
16. National Science Foundation (2023). FY 2024 Budget Request to Congress. <https://new.nsf.gov/about/budget/fy2024>
17. Executive Office of the President Office of Management and Budget (2021). 2022 Discretionary request. <https://www.whitehouse.gov/wp-content/uploads/2021/04/FY2022-Discretionary-Request.pdf>
18. 한국과학기술연구원 융합연구정책센터 (2008). 국가융합기술 발전 기본계획(제1차 융합연구개발 활성화 기본계획). <https://crpc.kist.re.kr/user/nd49151.do>
19. 한국과학기술연구원 융합연구정책센터 (2014). 창조경제 실현을 위한 융합기술 발전전략(제2차 융합연구개발 활성화 기본계획). <https://crpc.kist.re.kr/user/nd49151.do>
20. 한국과학기술연구원 융합연구정책센터 (2018). 제3차 융합연구개발 활성화 기본계획. <https://crpc.kist.re.kr/user/nd49151.do>
21. 한국과학기술연구원 융합연구정책센터 (2023). 제4차 융합연구개발 활성화 기본계획. <https://crpc.kist.re.kr/user/nd49151.do>

● 저 자 소 개 ●



이 지 예 (Jiye Rhee)

현재 경제·인문사회연구회 전문위원으로 재직하고 있다. 전남대학교에서 식물생명공학 전공으로 박사학위를 받은 후 Leibniz-Institute for Plant genetic and Crop plant research에 박사후 연구원으로 재직하였다. *Physiologia plantarum*, *Plant Molecular Biology*, 한국콘텐츠학회논문지 등에 논문을 게재한 바 있으며, 주요 관심분야는 학제간 융합연구 기획 및 평가, 성과관리, 융합인재 양성 등이다.



신 재 은 (Jaeun Shin)

현재 소상공인시장진흥공단 정책연구센터 부연구위원으로 재직 중이다. 충북대학교에서 경제학 박사 학위를 취득하였다. 주요 관심분야는 건강 영향 평가, 환경 정책, 융합 연구 등이다. 지금까지 환경정책, 보건경제와 정책연구, 정책개발연구, 입법과 정책 등 주요 학술지에 논문을 발표하였다.

〈 Abstract 〉

Analysis of Policy Trends in Convergence Research and Development Using Unstructured Text Data

Jiye Rhee^{*}, JaeEun Shin^{**}

This study aims to analyze policy changes over time by conducting a textual analysis of the basic plan for activating convergence research and development. By examining the basic plan for convergence research development, this study looks into changes in convergence research policies and suggests future directions, thereby exploring strategic approaches that can contribute to the advancement of science and technology and societal development in our country. In particular, it sought to understand the policy changes proposed by the basic plan by identifying the relevance and trends of topics over time. Various analytical methods such as TF-IDF analysis, topic modeling (LDA), and network (CONCOR) analysis were used to identify the key topics of each period and grasp the trends in policy changes. The analysis revealed clustering of topics by period and changes in topics, providing directions for the convergence research ecosystem and addressing pressing issues. The results of this study are expected to provide important insights to various stakeholders such as governments, businesses, academia, and research institutions, offering new insights into the changes in policies proposed by previous basic plans from a macroscopic perspective.

Key words: Convergence Research and Development(R&D) Activation Plan, Data Analysis, Topic Modeling(LDA), Network Analysis, Convergence Research

^{*} National Research Council for Humanities, Economics and Social Science
^{**} Korea SMEs and Startups Agency