

기술기회 발굴 지원을 위한 텍스트 분석 기반의 기술-디자인 트리 구축

윤주호 · 권세훈 · 김병훈[†]

한양대학교 산업경영공학과

Text Analytics-Based Construction of Technology-Design Trees for Technology Opportunity Discovery

Juho Yoon · Sae Hoon Kwon · Byunghoon Kim

Department of Industrial and Management Engineering, Hanyang University ERICA

This paper introduces a novel methodology for constructing technology-design trees based on text analysis to support the discovery of technology opportunities. The methodology employs KeyBERT(Keyword extraction with Bidirectional Encoder Representations from Transformers) model to extract meaningful keywords from patents and design rights documents, thus facilitating the analysis of semantic similarities and categorization into similar technological and design groups. Further, it utilizes Cooperative Patent Classification (CPC) and Locarno Classification (LOC) to build technology trees and design trees, respectively, through the analysis of technical and product similarities. Additionally, the methodology employs structural similarity index mapping (SSIM) on patent and design right drawings to validate the constructed trees. The approach is distinguished by its minimal reliance on expert intervention, enhancing the efficiency and scalability of technology opportunity discovery processes.

Keywords: Technology Opportunities Discovery, Patent-Design Right Linkage Data, Technology-Design Tree, Keyword Extraction, Similar Groups

1. 서론

기술기회는 잠재적으로 새롭고 유망한 기술 동향을 식별하는 것으로, 이러한 발견은 기술 개발을 계획하는 조직의 의사결정을 지원한다(Rodriguez *et al.*, 2016). 특히 특허와 같은 기술 문서의 분석을 통해 기술기회를 발굴하기 위한 다양한 연구들이 발전적으로 이루어져 왔다. Cozzens *et al.*(2010)와 Trappey *et al.*(2011)은 특정 기술 분야에 관한 기술 트렌드 파악을 시도하였고, Lee *et al.*(2009)과 Kim *et al.*(2011)은 네트워크 분석모델을 활용하여 기술지식의 파급효과를 파악하고 기회를 발굴하였다. Lee *et al.*(2009)과 Yoon and Kim(2012)은 다변량 자료

분석모델을 활용하여 특허 맵을 구축하고 공백 영역, 이상치, 군집 등을 파악함으로써 공백 기술과 특이 기술의 기회를 탐색한 바 있다. 하지만 이들 연구는 주로 기술분류코드나 특허 간 인용관계 등 정형화된 서지정보만을 활용하고 있어, 이를 통해 도출해 낸 기술기회의 의미가 구체적이지 못하고 포괄적이라는 한계가 있다. 이에 Choi *et al.*(2011), Feng *et al.*(2019) 그리고 Lee *et al.*(2021)은 개별 특허가 내포하고 있는 기술적 의미를 반영할 수 있도록 텍스트마이닝을 활용한 연구를 수행하였다. 그러나 이들 연구도 특허가 내포한 텍스트 정보를 이용해서 그룹화하고 있어, 세부 요소기술의 계층적 구조는 반영하지 못하는 한계가 있다(Seo, 2022). 또한 유망한 기술 아이

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2022R1F1A1063273).

[†] 연락저자 : 김병훈 교수, 15588 경기도 안산시 상록구 한양대하로 55 5공학관 510호 산업경영공학과, Tel: 031-400-5269,

E-mail: byungkim@hanyang.ac.kr

2024년 3월 7일 접수; 2024년 3월 21일 수정본 접수; 2024년 4월 12일 게재 확정.

템을 발굴해 내더라도 연구개발 주체의 역량과 적합성이 결여된 경우, 발굴된 기술기회의 적용 가능성은 매우 낮아질 수밖에 없다(Choi *et al.*, 2019). 이에 Kim and Shin(2013)은 내부적으로 보유한 특허를 기술역량으로 일반화하고 기술기회와 매칭하는 기준으로 활용하였고, Yoon *et al.*(2015)은 제품과 기술 간 고정된 형태의 시나리오를 가정한 연구를 수행한 바 있다. 그러나 이들 연구도 실질적인 보유역량을 활용한 기술기회 탐색 과정은 지원하지 못한다. 다시 말해서, 특허 기반의 분석만으로는 실제 비즈니스가 이루어지는 제품과 서비스 수준(product and service-levels)은 고려하지 못하는 한계가 있다(Yoon and Kim, 2024).

한편, 최근에는 기술 개발의 고도화로 기업들의 기술 수준이 평균화되면서 기술 중심 개발보다 먼저 디자인 중심 개발을 선행하는 경향이 있다(Chen and Chen, 2007). 즉, 디자인권 분석을 통해 디자인 중심의 기술동향을 살펴볼 수 있다(Chen, 2009). 더욱이 디자인권에 대한 소송 규모가 점차 확대되고 있고, 이는 기업의 수익과 시장 진출 상황과 직결된다는 점에서 그 가치가 인정되는 바(Jeong *et al.*, 2019). 이처럼 기술적이면서도 합법적인 산업재산권(industrial property)의 분석을 통해 경쟁자의 연구개발을 이해하고, 기업의 시장 기회를 파악할 수 있다(Huang *et al.*, 2008). 그러나 특허와 디자인권을 동시에 분석한 연구가 거의 없고, 일부 디자인권을 활용한 연구는 디자인 출원 동향과 같은 거시적인 관점에서의 전략 지원에 국한되어 있다(Song *et al.*, 2023).

이에 본 연구에서는 연구개발 주체의 보유역량(기술/제품)에 기반한 기술기회 발굴을 지원하기 위해, 특허와 디자인권의 텍스트 분석을 기반으로 기술-디자인 트리를 구축하는 방법을 제안한다. 이를 위해서 먼저, 특허와 디자인권으로부터 KeyBERT(Keyword Extraction with Bidirectional Encoder Representations from Transformers) 모델을 이용하여 문맥을 고려한 주요 키워드를 추출하여 의미론적 유사도(semantic similarity)를 분석한다. 이를 통해 각 특허와 디자인권을 유사한 기술군과 디자인군으로 유형화한다. 다음으로, 각 기술군에 대하여서 선진특허분류(Cooperative Patent Classification, 이하 CPC)를 활용한 기술적 유사도(technical similarity) 분석을 통해 기술 트리(technology tree)를 구축하고, 각 디자인군의 로카르노 분류(Locarno Classification, 이하 LOC)를 이용하여 제품적 유사도(product similarity) 분석을 통하여 디자인 트리(design tree)를 구축한다. 즉, 기술 트리란 수집된 각 특허를 기술적 의미와 기술 분류(technology class)를 모두 일치시킨 후에 특허가 내포한 요구기능(demand function)을 계층적으로 구조화한 것이다. 마찬가지로 디자인 트리는 수집된 각 디자인권에 대하여서 디자인 컨셉과 제품 분류(product class)가 일치하도록 한 후에 디자인권이 내포한 제품 디자인 니즈(needs)를 계층적으로 구조화한 것이다. 마지막으로, 각 특허와 디자인권의 도면 정보를 활용하여 구축된 트리를 검증한다. 이러한 기술-디자인 트리의 구축은 거시적으로 특허와 디자인권의 동

향을 비교하고, 미시적으로는 연구개발 주체가 기술-디자인 관점을 모두 고려하여 개발전략 수립을 지원하는 데 목적이 있다(Song *et al.*, 2023).

본 연구는 학술적 관점에서 특허-디자인권 융합 분석을 통한 기술기회 발굴 초기 연구라는 점에 의의가 있다. 특허 데이터의 요구기능과 디자인권 데이터에 있는 제품 디자인 니즈에 관한 정보를 계층적으로 구조화하고, 이들 간의 관계를 분석함으로써 기술-시장의 양면적 통찰력을 제공할 수 있다. 이는 다양한 산업재산권을 연계한 후속 연구의 모범 사례가 될 것으로 기대한다. 또한 실무적 관점에서는 인적, 물적 자원이 부족한 기술기반 기업과 중소벤처기업들의 콜드 스타트를 적시에 효율적으로 지원할 것으로 사료된다.

본 논문은 다음과 같이 구성된다. 제2장에서는 기술-디자인 트리에 기반한 잠재적 기술기회 발굴 연구를 고찰하고, 제3장에서는 제안 방법론을 설명한다. 제4장에서는 제안 방법론의 실시 예를 보이고, 제5장에서는 본 연구를 정리한다.

2. 관련 연구

본 장에서는 이론적 관점에서 기술-디자인 트리를 기반으로 한 기술기회 발굴 연구를 대관하고, 방법론적 관점에서 키워드 추출을 활용한 특허분석 연구를 세찰함으로써 본 논문의 연구범위를 명확히 한다.

2.1 기술-디자인 트리 기반 잠재적 기술기회 발굴

기업의 잠재적 기술기회를 조기에 발굴할 수 있도록 지원하는 것은 해당 기업의 연구개발 기획역량을 강화하고, 연구개발 결과물의 사업화 성과를 제고하는 데 있어 매우 중요하다(Seo, 2021). 더불어 최근에는 정보화 기술이 발전함에 따라 정량적 데이터 분석을 기반으로 하는 기술기회 발굴 연구가 활발하게 진행되고 있다(Lee *et al.*, 2021). 특허는 기술혁신과 연구개발의 동향에 대한 정량적 대리변수로 널리 활용되고 있으며, 이러한 특허 데이터의 분석을 통해 자동화된 기술기회 발굴 방법론에 대한 연구가 다양한 관점에서 수행되고 있다(Park and Yoon, 2017).

Song *et al.*(2023)은 미국의 발명 특허(utility patent)와 디자인 특허(design patent)를 융합 분석하여 기술과 디자인 간의 유사성을 토대로 기술기회 발굴을 지원하기 위한 방법론을 제안하였다. 이를 위해서 먼저, 수집된 디자인 특허로부터 도메인 분야를 나타내는 field keyword와 디자인 목적의 대상을 나타내는 target object keyword를 추출하여 <Table 1>과 같이 디자인 특허 트리를 도출한다. 이때 field keyword와 target object keyword는 사전에 정의된 키워드 사전에 의해 정성적으로 그룹화하여 정의된다. 다음으로 수집된 발명 특허 간의 키워드와 CPC의 코사인 유사도를 통해 군집 분석을 수행하고 <Table 2>

와 같은 발명 특허 트리를 도출한다. 여기서 도출된 각 군집은 대표 키워드와 CPC를 참조하여 정성적으로 군집명을 정의한다. 마지막으로, <Table 3>과 같이 각 트리 간 코사인 유사도 분석을 통해 기술 개발-제품 디자인 영역을 융합하여 개발 전략을 도출한다. 이와 같은 시도는 특허와 디자인권을 동시에 분석하여 연구개발 주체의 보유역량에 기반한 미래의 유망 기술 또는 제품을 식별한다는 점에 의의가 있다. 그러나 분석 프로세스 내 전문가 의존도가 높거나 개입이 빈번하여 분석사례

별로 한정된 기술기회 발굴을 지원하는 한계가 있다. 구체적으로, 디자인 특허 트리를 구축하는 과정에서 속지주의 (territorial principle)를 따르는 산업재산권의 특성상 미국의 경우 디자인권(design right)을 특허권의 일종인 디자인 특허로 파악한다. 따라서 디자인 특허는 미국특허법(35 U.S.C.)에 따라 <Figure 1>의 (a)와 같이 도시된 대로를 청구하는 전형적인 청구항(claim) 하나만 필요하고, 복수청구항은 허용되지 않는다(Kim, 2010). 즉, 간결한 청구항을 중심으로 추출된 주요 키

Table 1. Design Patent Tree Structure(Song *et al.*, 2023)


Field keyword	Target object keyword
Automobile	Display system
	Surface configuration
	Dashboard constructure
	⋮
Commercial & Working vehicle	Surface configuration
	Steering wheel
⋮	⋮

Table 2. Characteristics of Each Utility Patent Group(Song *et al.*, 2023)

Cluster name	CPC	Main keyword (based on TF-IDF)
Dashboard support connection structure	B62D25/145 (Dashboard crossbeam components) B60H1/242 (Front structure of air filter)	air condition system, dashboard support, arrangement structure, layout structure, cross member, duct structure, dash cover, support lip
Display control system	B60K37/06 (Dashboard control knob) G02B27/0101 (HUD: optical feature)	head-up display, vision system, image display, opening cover, display unit, image projection, virtual image, virtual control
⋮	⋮	⋮

Table 3. Result of Similarity Analysis between Utility Patent-tree and Design Patent-tree(Song *et al.*, 2023)

Utility patent tree	Applicant ratio		Similarity		Design patent tree	Applicant ratio	
	OEM	Others	Measure	Standardization value		OEM	Others
Dashboard support connection structure	41%	59%	High	0.7789	Automobile - Center fascia	75%	25%
			Low	0.0000	Commercial & working vehicle Surface configuration	0%	100%
Display control system	50%	50%	High	0.7945	Automobile - Center fascia	75%	25%
			Low	0.0000	Commercial & working vehicle Surface configuration	0%	100%
Airbag module	36%	64%	High	1.0000	Automobile - Steering wheel	60%	40%
			Low	0.0000	Commercial & working vehicle Surface configuration	0%	100%
Device holder	19%	81%	High	0.5882	Automobile - Center fascia	75%	25%
			Low	0.0000	Commercial & working vehicle Surface configuration	0%	100%
Safety devices around dashboard	36%	64%	High	0.8758	Automobile - Center fascia	75%	25%
			Low	0.0000	Commercial & working vehicle Surface configuration	0%	100%



US9737445

(12) **United States Design Patent** (10) Patent No.: **US D978,744 S**
 Kupitza (43) Date of Patent: **Feb. 21, 2023**

(54) **DASHBOARD FOR A VEHICLE** 4,972,913 A 11/090 Ray et al.
 D416,829 S 11/090 Sees et al.
 D422,044 S 4/2009 Heilbake et al.
 D425,826 S 5/2009 Heilbake et al.
 D429,627 B1 5/2009 von Muehlstein et al.
 D438,667 S 1/2007 Fiedler
 D448,219 S 4/2007 Oka
 D507,885 S 4/2009 Amador et al.
 D622,195 S * 8/2010 Lantz
 D683,284 S * 7/2013 Hanson et al.
 (Continued)

(*) Term: **15 Years**

(21) Appl. No.: **29807,566**

(22) Filed: **Sep. 13, 2021**

Related U.S. Application Data
 (62) Division of application No. 35/509,887, filed on Nov. 26, 2019, U.S. filing date under 35 U.S.C. 384, and having an international filing date of Nov. 26, 2019, now Int. No. Des. 933,532.

Foreign Application Priority Data
 Jun. 6, 2019 (DE) 40 2019 100 633.8
 (51) **LOC (14) CL** 12-11
 (52) **U.S. CL** D12/192
 (53) **U.S. CL** D12/192

(58) **Field of Classification Search**
 USPC: D12/906, 192, 415; D10/46, 98, 102, 103, D10/122, 127; D15/17, 28
 CPC: B60K 35/00; B60K 37/04; B60K 37/06; B60R 11/0235; B60R 11/0252
 See application file for complete search history.

(56) **References Cited**
 U.S. PATENT DOCUMENTS
 2,699,223 A * 1/1915 Birmingham
 4,079,087 A * 3/1978 Burgstamer 1608/11,02
 D276,710 S 12/1994 Boungar et al. D12/418

OTHER PUBLICATIONS
 "Simplifying Business," Man Truck&Bus Asia, May 27, 2020 [online], Internet on Oct. 9, 2023. Retrieved from the Internet <URL: http://www.man-truck&bus-asia.com/pdf/Company%20Profile%20MAN.pdf>.*

CLAIM
 (57) The ornamental design for a dashboard for a vehicle, as shown and described.

DESCRIPTION
 FIG. 1 is a top view of a dashboard for a vehicle embodying my new design.
 FIG. 2 is a right side elevation view thereof.
 FIG. 3 is a left side elevation view thereof.
 FIG. 4 is a bottom view thereof.
 FIG. 5 is a front elevation view thereof.
 FIG. 6 is a rear elevation view thereof.
 FIG. 7 is a front and right side perspective view thereof.
 FIG. 8 is a front and left side perspective view thereof.
 FIG. 9 is a top, front, and right perspective view thereof, and.
 FIG. 10 is a top, front, and left side perspective view thereof.

Claim, 10 Drawing Sheets



(a) Structure of a Design Patent

디자인도면
품목명
 제12부

디자인의 대상이 되는 품목
 차량용 대시보드

디자인의 설명

- 제1도는 측위, 합성수지 및 금속임.
- 본원디자인은 자동차의 내부 공간에 설치되는 내장개로서, 주행에 필요한 각종 정보를 나타내는 디스플레이가 장착되는 센터 디스플레이, 운전자 측 전면에는 계기장치를 나타내는 클러스터(cluster)와 정보 안내 디스플레이(CID, Center Information display)가 구비되어 있고, 조수석 측 전면에는 [도면 B1.1]에서 볼 수 있는 바와 같이 노출될 수 있는 숨겨진 디스플레이(hidden display)가 구비되어 있는 것임.
- 도면에 표시된 세선은 평면 내지 곡면을 나타내기 위한 선임.
- 실선으로 표현된 부분이 본원디자인으로서 디자인등록을 받고자 하는 부분이며, 1점쇄선은 등록하고자 하는 부분이 등록번호가 되지 않는 부분과의 경계면을 나타내는 선임.
- 본원디자인은 [도면 A1]부터 [도면 B1]까지 순차적으로 변화하는 것으로, [도면 A1]는 변화 전 상태의 도면이고, [도면 B1]는 변화 후 상태의 도면임.
- [도면 A1.1]은 본원디자인의 조수석 측 디스플레이부가 숨겨진 상태에서의 전체적인 형태를 표현한 것임.
- [도면 A1.2]는 본원디자인의 조수석 측 디스플레이부가 숨겨진 상태에서의 정면 부분을 표현한 것임.
- [도면 A1.3]은 본원디자인의 조수석 측 디스플레이부가 숨겨진 상태에서의 배면 부분을 표현한 것임.
- [도면 A1.4]는 본원디자인의 조수석 측 디스플레이부가 숨겨진 상태에서의 좌측 부분을 표현한 것임.
- [도면 A1.5]는 본원디자인의 조수석 측 디스플레이부가 숨겨진 상태에서의 우측 부분을 표현한 것임.
- [도면 A1.6]은 본원디자인의 조수석 측 디스플레이부가 숨겨진 상태에서의 전면 부분을 표현한 것임.
- [도면 B1.1]은 본원디자인의 조수석 측 디스플레이부가 노출된 상태에서의 전체적인 형태를 표현한 것임.
- [도면 B1.2]는 본원디자인의 조수석 측 디스플레이부가 노출된 상태에서의 정면 부분을 표현한 것임.
- [도면 B1.3]은 본원디자인의 조수석 측 디스플레이부가 노출된 상태에서의 배면 부분을 표현한 것임.
- [도면 B1.4]는 본원디자인의 조수석 측 디스플레이부가 노출된 상태에서의 좌측 부분을 표현한 것임.
- [도면 B1.5]는 본원디자인의 조수석 측 디스플레이부가 노출된 상태에서의 우측 부분을 표현한 것임.
- [도면 B1.6]은 본원디자인의 조수석 측 디스플레이부가 노출된 상태에서의 전면 부분을 표현한 것임.
- [도면 B1.7]은 본원디자인의 조수석 측 디스플레이부가 노출된 상태에서의 전면 부분을 표현한 것임.

디자인 창작 내용의 요점
 본원 차량용 대시보드의 형상과 모양의 결합을 디자인 창작내용의 요점으로 함.

(b) Structure of a Design Right

Figure 1. Example of Design Patents and Design Right

워드가 해당 디자인 특허의 대표성을 나타내기에 충분하다고 볼 수 있다(Song et al., 2023). 그러나 <Figure 1>의 (b)와 같이 한국 디자인권과 같이 물품명칭, 창작내용, 창작요점 등 텍스트 데이터가 복잡한 구조와 형태일 경우에는 디자인권의 특성을 대표하는 주요 키워드를 추출하는 데 한계가 있다. 또한 발명 특허의 CPC와 같은 서지정보를 활용하여 디자인 트리를 도출하는 프로세스 보완도 필요하겠다. 둘째로, 발명 특허 트리의 각 군집을 대표하는 주요 키워드(main keyword)를 추출하는 과정에서 TF-IDF(term frequency-inverse document frequency)를 기반으로 하고 있어, 동음이의어나 이음동지어같이 문장의 맥락에서 변화하는 단어들의 의미를 파악하지 못한다는 한계가 있다(Yu et al., 2012). 마지막으로, 도출된 각 트리를 검증하는 프로세스도 필요하다. 이에 착안하여 본 연구에서는 텍스트 데이터의 구조와 형태가 복잡한 경우에도 전문가의 개입을 최소화하는 텍스트 분석 기반의 기술-디자인 트리 구축 방법을 제안한다.

2.2 지능형 키워드 추출

특허분류코드와 같은 기술군 단위는 상당히 넓은 범위의 기술 속성을 포괄하기 때문에 이를 통해 도출해 낸 기술기회의 의미가 구체적이지 못하고 모호하다는 한계가 존재한다. 디자인분류코드의 경우에도 마찬가지이다. 예를 들어, <Figure 2>의 (a)는 현대 자동차의 인테리어 파트인 스티어링 휠이고, (b)는 익스테리어 파트인 자동차용 휠로 LOC가 12-16으로 동일하다. 즉, 자동차와 관련된 신제품개발 기회를 찾기 위해서 LOC만으로는 디자인 트렌드 영역을 살피기 어렵고, 심지어 인테리어와 익스테리어의 구분

도 어렵다는 것을 보여준다. 이에 본 절에서는 지능형 키워드 추출 방법을 활용하여 전문가의 개입을 최소화하고, 보다 명확한 기술적 의미와 제품 디자인 컨셉을 내포한 기술-디자인 트리를 구축할 수 있도록 키워드 추출 방법을 살펴본다.



(a) Application No: 30-2017-0006509



(b) Application No: 30-2023-0020755

Figure 2. Examples of LOC 12-16 Applied to Both Interior and Exterior Design

Table 4. Comparison of Intelligent keyword Extraction Models

Author	Model	Application data			Task	
		News	Articles	Patents	Keyword extraction	Keyphrase extraction
Witten <i>et al.</i> (1999)	Naive Bayes		✓			✓
Zhang <i>et al.</i> (2008)	CRF		✓		✓	
Li <i>et al.</i> (2007)	TF-IDF	✓			✓	
Mihalcea and Taru(2004)	TextRank		✓		✓	
Rose <i>et al.</i> (2012)	RAKE					✓
Zhichao <i>et al.</i> (2024)	KeyBERT			✓	✓	✓

일반적으로 키워드 추출 방법은 라벨링 된 말뭉치(corpus)의 필요 여부에 따라 지도 및 비지도 방법으로 분류된다. 지도학습 기반의 키워드 추출 방법은 이진 분류(binary classification) 문제로 처리된다. 즉, 머신러닝 기반의 분류기를 통해 문서 내의 각 단어나 구문이 키워드인지를 결정한다. Witten *et al.*(1999)은 Naive Bayes 알고리즘을 기반으로 간단하고 효율적인 핵심 구문 추출 알고리즘 KEA(Keyphrase Extraction Algorithm)를 제안하였다. Zhang *et al.*(2008)은 조건부 무작위성을 기반으로 문서의 특징이 효과적으로 반영된 레이블링 필드로 키워드 추출을 진행하여 SVM, 선형회귀모델 등 다른 기계학습 방법에 비해 키워드 추출 성능을 향상시킨 바 있다. 그러나 지도학습 기반의 키워드 추출 방식은 라벨링 된 말뭉치가 필요하고, 훈련 말뭉치의 품질이 모델의 성능에 직접적인 영향을 주어, 키워드 추출 결과에 영향을 미친다는 한계가 있다. 더욱이 라벨을 태깅하는 과정에서의 오류나 비용, 시간 등의 문제도 있다(Hu *et al.*, 2018). 반면에 비지도학습 기반의 키워드 추출 방법은 사전에 라벨링 되지 않은 말뭉치에서 키워드를 추출한다는 장점이 있다. Li *et al.*(2007)은 뉴스 데이터에서 키워드를 추출하기 위한 TF-IDF 기반의 접근 방식을 제안하여 기존 방법들보다 정확도 측면에서 훨씬 뛰어난 성능을 보였다. Mihalcea and Taru(2004)는 PageRank 알고리즘과 유사한 그래프 기반의 TextRank를 제안하였다. TextRank는 PageRank 알고리즘을 사용하여 각 단어의 중요성을 반복적으로 계산하고, 각 단어별로 중요성을 정렬하여 상위 순위에 있는 단어를 키워드로 선택한다. 그러나 PageRank 값을 계산하는 데 많은 반복이 필요하고, 높은 계산 복잡성으로 인해 대규모 텍스트 키워드 추출 작업에서는 거의 사용되지 않는다. Rose *et al.*(2012)은 개별 문서에서 핵심 문구를 추출하기 위해 RAKE(rapid automatic keyword extraction) 알고리즘을 제안하였고, TextRank에 비해 긴 핵심 구문(keyphrase) 추출에 있어서 더 향상된 성능을 보였다. Zhichao *et al.*(2024)은 KeyBERT 알고리즘을 기반으로 과학과 기술 간의 선행-지연 거리를 식별하여 기술기회를 발굴하는 방법을 제안하였고, 과학논문과 특허에서 KeyBERT를 활용하여 키워드를 추출한 결과 TF-IDF, TextRank, RAKE 알고리즘 보다 우수한

성능과 안정성을 보임을 확인하였다. 본 연구는 이에 착안하여, 특허 데이터에서 특징적인 키워드(discriminative keyword) 추출에 적합하고, 별도의 라벨링된 말뭉치가 필요하지 않은 KeyBERT를 기반으로 특허와 디자인권의 주요 키워드를 추출한다. 이를 통해 <Table 1>과 같이 디자인 트리를 구축하는 과정에서 사전에 정의된 키워드 사전이 없고, 한국의 디자인권과 같이 텍스트 데이터의 구조와 형태가 복잡한 경우에도 전문가의 개입을 최소화 하여 제품 디자인 컨셉을 내포한 키워드를 추출하고자 한다. 또한 <Table 2>와 같은 기술 트리의 구축 과정에서 기존 연구(Song *et al.*, 2023)에서 활용된 TF-IDF 모델을 사용할 경우, 한계점인 동음이의어나 이음동이의와 같은 문제가 발생할 수도 있다. 이는 신규성(novelty)과 진보성(inventive step and non-obviousness)을 만족해야 하는 특허의 요건 상 유사한 기술일지라도 다른 용어로 표현되는 특성을 고려하지 못하므로, <Table 4>와 같이 KeyBERT를 활용하여 특허 데이터의 문맥의 의미를 살피고, 기술적 의미가 담긴 키워드를 추출하는 것이 바람직하다.

3. 제안 방법론

특허와 디자인권의 유사성을 토대로 기술기회 발굴을 지원하기 위한 기술-디자인 트리를 구축하는 과정은 <Figure 3>과 같이 3단계로 진행된다. 첫 번째 단계에서는 특허의 텍스트 데이터로부터 KeyBERT 모델을 활용하여 의미적 유사도를 분석하고, CPC를 활용하여 기술적 유사도를 분석하여 기술 트리를 구축한다. 두 번째 단계에서는 디자인권에 대하여서 발명 특허와 동일한 방식으로 의미적 유사도를 분석하고, LOC를 활용하여 제품적 유사도를 분석하여 디자인 트리를 구축한다. 세 번째 단계에서는 각 특허와 디자인권의 도면 정보를 활용하여 구축된 트리를 검증한다. 이를 통해 기술-제품 디자인 간 유사성이 높은 분야를 도출하거나 연관성이 거의 없는 분야를 도출한다. 기존 연구(Song *et al.*, 2023)와 제안 방법론의 트리 구조를 비교하면 <Table 5>와 같다.

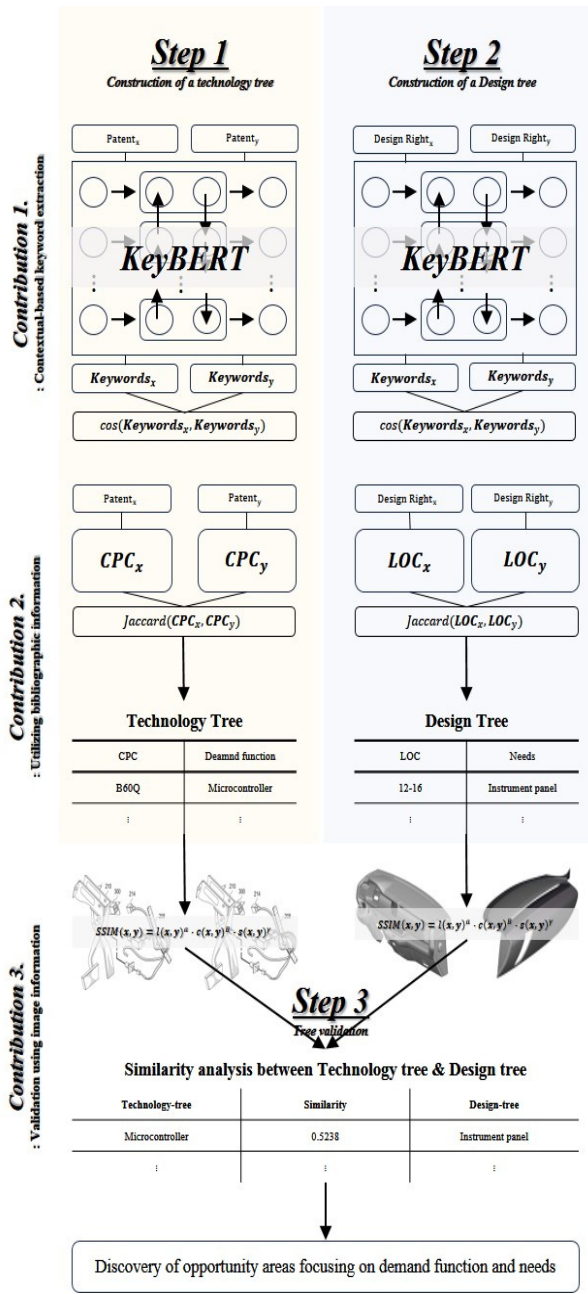


Figure 3. Framework of the Proposed Methodology

3.1 KeyBERT 기반의 의미적 유사도 분석

Grootendorst(2020)에 의해 개발된 KeyBERT는 <Figure 4>와 같이 후보 구문 선택, 임베딩 추출, 코사인 유사도 계산의 세 단계로 진행된다. 먼저, 문서로부터 N-gram 단위의 후보 구문을 선택한다. 다음으로, 사전 훈련된 Sentence-BERT(Reimers and Gurevych, 2019) 모델을 활용하여 문서와 후보 구문의 임베딩 벡터를 추출한다. 여기서 Sentence-BERT는 <Figure 5>와 같이 두 개의 BERT 모델에 문장을 하나씩 입력하여 BERT 간에 서로 같은 weight를 공유하는 siamese network 구조로 문장 임베딩 성능을 극대화한 모델이다(Kang *et al.*, 2023). 이를 통해 어간 추출(stemming)이나 표제어 추출(lemmatization)과 같은 텍스트 정규화나 불용어 처리(stopword) 등의 텍스트 분석을 위한 전처리 과정을 생략하여도 문장 단위(sentence-level)에서 단어의 의미적(semantic), 구문적(syntactic) 정보가 반영된 키워드가 추출된다. 마지막으로, 임베딩된 문서와 각 후보 구문 간의 코사인 유사도를 계산한다. 이때 문서와 가장 유사한 후보 구문을 전체 문서를 가장 잘 설명하는 키워드로 식별할 수 있다고 가정한다(Choi and Lim, 2023). 예를 들어, <Figure 6>과 같이 “본 발명은 나노기술을 사용하여 복합재료를 생산하는 새로운 방법에 관한 것.”에 관한 내용을 담고 있는 문서에서 3-gram 후보 구문을 선택하고, 문서와 후보 구문을 각각 임베딩한 뒤 유사도를 계산한 결과 ‘나노기술’, ‘이용한’, ‘소재’라는 3개의 단어로 구성된 구문이 해당 문서를 가장 잘 설명하는 키워드로 도출된다. 이렇게 도출된 키워드는 문맥을 고려해 도출된 구문의 형태이기 때문에 해석하기 수월한 장점이 있다. 이처럼 본 연구에서는 각 특허와 디자인권에 대하여서 KeyBERT를 이용하여 기술적·제품적 의미를 내포한 키워드를 추출하고, 각 키워드 간의 코사인 유사도 분석을 통해 의미적 유사도 행렬(semantic similarity matrix)을 도출한다. 이때, 유사도 값에 임계값을 주어 일정 수준 이상(δ_1)인 경우 유사한 기술군, 유사한 디자인군으로 유형화할 수 있다. 상기 과정은 Algorithm 1로 정리된다. Algorithm 1의 라인 3부터 라인 12는 KeyBERT의 3단계를 함수로 정의한다. 라인 14부터 라인 16은 N개의 문서에 대하여서, i번째 문서의 키워드와 j번째 문서의 키워드 간 코사인 유사도를 계산을 반복하여 의미적 유사도 행렬 SS가 생성된다. 이때, 라인 17과 라인 18을 통해 의미적 유사도가 임계값(δ_1) 이상인 경우 최

Table 5. Comparison of Tree Structures between Existing Methodology and Proposed Methodology

Tree structure	Existing methodology (Song <i>et al.</i> , 2023)		Proposed methodology	
	Utility patent-tree	Design patent-tree	Technology-tree	Design-tree
Database	(USPTO) Utility patents	(USPTO) Design patents	(KIPRIS) Patents	(KIPRIS) Design rights
Keyword extraction	TF-IDF	Dictionary of keywords predefined by experts	KeyBERT	
Bibliographic information	CPC	-	CPC	LOC
Validation	-	-	Patent drawings	Perspective and hexagonal views

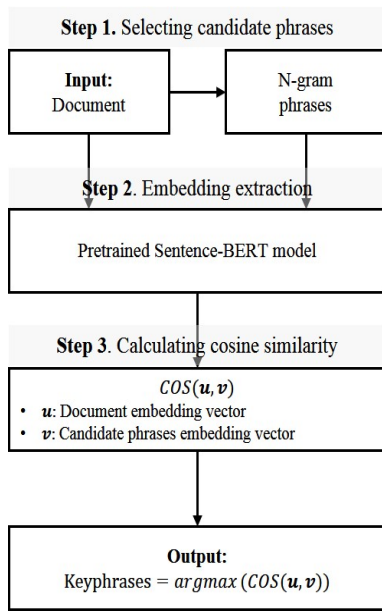


Figure 4. KeyBERT Steps

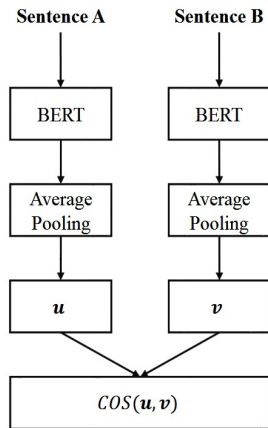


Figure 5. Structure of Sentence-BERT model

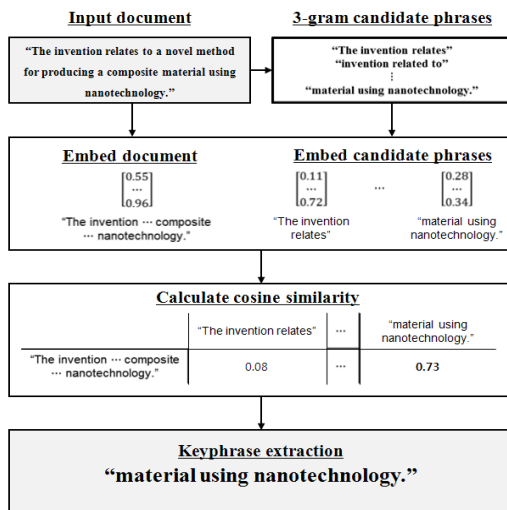


Figure 6. Example of Keyword Extraction

중적으로 활용 가능한 *SS*가 반환된다.

Algorithm 1. Semantic Similarity based on KeyBERT

```

D : Meaningful document
Pi : Candidate phrases
u : Document embedding vector
v : Candidate phrases embedding vector
l : The length of words
N : The number of documents
n : Parameter of N-gram
δ1 : Threshold
SS : Semantic similarity matrix

1 Input: Meaningful documents
2
3 def KeyBERT(D):
4     Pi ← i = 1, ..., l - n + 1 # Step 1
5
6     for i = 1 to l - n + 1 do:
7         vi ← SentenceBERT(Pi) # Step 2
8     u ← SentenceBERT(D)
9
10    for i = 1 to l - n + 1 do:
11        return keyword ← argmax(cos(u, vi)) # Step 3
12    end
13
14    for i = 1 to N do:
15        for j = 1 to N do:
16            SSij ← cos(KeyBERT(Di), KeyBERT(Dj))
17    if SS ≥ δ1
18        return SS
19    end
20
21 Output: Semantic similarity matrix
    
```

3.2 서지정보를 활용한 기술적·제품적 유사도 분석

한편 특허를 출원하기 위해서는 신규성, 진보성 등을 만족해야 하므로 특허 내 표현되는 기술 용어의 경우 유사한 기술에 대한 권리를 주장함에도 서로 다른 용어로 표현되는 특성이 있다(Chae and Gim, 2018). 따라서 의미적 유사도 분석을 통해 유형화한 기술군과 디자인군은 각 산업재산권의 개별속성을 설명하는 키워드가 유사한 정도에 따라 그룹화한 것이므로, 그 자체를 기술 트리나 디자인 트리로 간주하기에는 어려움이 있다. 따라서 본 절에서는 특허와 디자인권의 서지정보(bibliographic information)를 토대로 기술적·제품적 유사성을 고려하여 기술-디자인 트리를 구축하기로 한다.

특허의 CPC는 현존하는 특허분류체계 중 가장 세분화된 코드 수를 가지고 있다. CPC는 <Table 6>과 같이 섹션, 클래스,

서브클래스, 메인그룹, 서브그룹의 계층적 구조로 국제특허분류(International Patent Classification, 이하 IPC)의 구조와 동일하다. 그러나 IPC에는 존재하지 않는 기술 주제를 분류하기 위해 <Table 7>과 같이 Y 섹션이 추가되었고, IPC 대비 메인그룹 기준 3,212개, 서브그룹 기준으로는 195,406개의 분류가 더 많다. 즉, CPC를 통해 IPC보다 세분화되고 구체적인 기술 분류가 가능하다(Park *et al.*, 2022). 또한 발명과 관련된 기술내용이 여러 개일 경우에는 복수 개의 CPC 코드를 부여할 수 있다. 본 연구에서는 자카드 유사도를 활용하여 특허 간 CPC의 공유 여부에 따라 기술적 유사도 행렬(technology similarity matrix)을 도출한다. 이때, 유사도 값에 임계값을 주어 일정 수준 이상(δ_1)인 경우 동일한 기술 분야로 간주한다. 상기 과정은 Algorithm 2로 정리된다. Algorithm 2의 라인 3부터 라인 5은 N 개의 문서에 대하여서, i 번째 문서의 CPC와 j 번째 문서의 CPC 간 자카드 유사도를 계산을 반복하여 기술적 유사도 행렬 TS 가 생성된다. 이때, 라인 6과 라인 7을 통해 기술적 유사도가 임계값(δ_1) 이상인 경우 최종적으로 활용 가능한 TS 가 반환된다. 이로써 3.1절에서 도출된 유사한 기술군은 식 (1)과 같이 기술 분야의 필터링을 통해 기술 트리로 구축된다.

$$\text{Technology Tree} = C_{ij}^{ech} = SSP_{ij} \times TS_{ij} \quad (1)$$

여기서 SSP_{ij} 는 각 특허 간 의미적 유사도 행렬을 의미하고, TS_{ij} 는 기술적 유사도 행렬을 의미한다. C_{ij}^{ech} 는 SSP_{ij} 와 TS_{ij} 의 동일한 위치에 있는 원소들끼리 곱(element-wise product)하여 생성된 유사도 행렬이다. 즉, 각 특허는 기술적 의미와 기술 분야가 모두 유사할 경우 기술 트리로 구축된다.

디자인권의 LOC는 디자인 물품 분류의 국제적 통일화를 위해 채택된 분류체계이다. LOC는 <Table 8>과 같이 원칙상 물품의 목적(purpose)에 따라 분류하고, 가능한 경우 물품이 나타내는 형상(the object)에 따라 보조적으로 분류한다. 로카르노 분류 14판 기준 32개류(Class)와 241개군(Subclass)으로 구성되어 있다(Yoon and Kim, 2023). 본 연구에서는 특허와 동일한 방식으로 디자인권 간 LOC의 공유 여부에 따라 제품적 유사도 행렬(product similarity matrix)을 도출한다. 이때, 유사도 값에 임계값을 주어 일정 수준 이상(δ_1)인 경우 동일한 디자인 분야로 간주한다. 상기 과정은 Algorithm 3으로 정리된다. Algorithm 3의 라인 3부터 라인 5은 N 개의 문서에 대하여서, i 번째 문서의 LOC와 j 번째 문서의 LOC 간 자카드 유사도를 계

Table 6. Example of CPC Structure

Symbol	CPC	Description
Section	B	PERFORMING OPERATIONS; TRANSPORTING
Class	B60	VEHICLES IN GENERAL
Subclass	B60H	ARRANGEMENTS OF HEATING, COOLING, VENTILATING OR OTHER AIR-TREATING DEVICES SPECIALLY ADAPTED FOR PASSENGER OR GOODS SPACES OF VEHICLES
Main group	B60H1/00	Heating, cooling or ventilating {[HVAC]} devices (heating, cooling or ventilating devices providing other air treatment, the other treatment being relevant, B60H3/00; ventilating solely by opening windows, doors, roof parts, or the like B60J; heating or ventilating devices for vehicle seats B60N2/56; vehicle window or windscreen cleaners using air, e.g. defrosters, B60S1/54)
Subgroup	B60H1/00985	Control systems or circuits characterised by display or indicating devices, e.g. voice simulators (characterised by the input of sound using a voice synthesizer B60H1/00757; B60H1/00978 takes precedence)

Table 7. CPC Sections and Number of Hierarchical Codes

Section		Class	Subclass	Main group	Subgroup
A	HUMAN NECESSITIES	16	87	1,461	29,891
B	PERFORMING OPERATIONS, TRANSPORTING	38	172	2,876	57,953
C	CHEMISTRY, METALLURGY	21	89	1,787	38,581
D	TEXTILES, PAPER	10	41	411	5,793
E	FIXED CONSTRUCTIONS	8	32	433	9,227
F	MECHANICAL ENGINEERING, LIGHTING, HEATING, WEAPONS; BLASTING ENGINES OR PUMPS	19	103	1,647	28,065
G	PHYSICS	15	87	985	38,398
H	ELECTRICITY	6	52	808	39,608
Y	NEW TECHNOLOGY DEVELOPMENTS	3	11	349	18,081
Total	136	136	674	10,757	265,597

산을 반복하여 제품적 유사도 행렬 PS 가 생성된다. 이때, 라인 6과 라인 7을 통해 제품적 유사도가 임계값(δ_1) 이상인 경우 최종적으로 활용 가능한 PS 가 반환된다. 이로써 3.1절에서 도출된 유사한 디자인군은 식 (2)과 같이 디자인 분야를 필터링함으로써 디자인 트리로 구축된다.

$$\text{Design Tree} = C_{ij}^{\text{prod}} = SSD_{ij} \times PS_{ij}$$

여기서 SSD_{ij} 는 각 디자인권 간 의미적 유사도 행렬을 의미하고, PS_{ij} 는 제품적 유사도 행렬을 의미한다. C_{ij}^{prod} 는 SSP_{ij} 와 TP_{ij} 의 동일한 위치에 있는 원소들끼리 곱하여 생성된 유사도 행렬이다. 즉, 각 디자인권은 디자인 컨셉과 제품 분야가 모두 유사할 경우 디자인 트리로 구축된다.

Algorithm 2. Technical Similarity based on CPC

CPC : CPC of patents

TS : Technical similarity matrix

N : The number of documents

```

1 Input : CPC of utility patents
2
3 for  $i = 1$  to  $N$  do:
4   for  $j = 1$  to  $N$  do:
5      $TS_{ij} \leftarrow \text{Jaccard}(CPC_i, CPC_j)$ 
6   if  $TS \geq \delta_1$  :
7     return  $TS$ 
8 end
9
10 Output: Technical similarity matrix

```

Table 8. Example of LOC Structure

Structure	Mark	Titel
Class	17	Musical instruments
Subclass	17-01	Keyboard instruments

Algorithm 3. Product Similarity based on LOC

LOC : LOC of design rights

PS : Product similarity matrix

N : The number of documents

```

1 Input : LOC of design rights
2
3 for  $i = 1$  to  $N$  do:
4   for  $j = 1$  to  $N$  do:
5      $PS_{ij} \leftarrow \text{Jaccard}(LOC_i, LOC_j)$ 
6   if  $PS \geq \delta_1$  :
7     return  $PS$ 
8 end
9
10 Output: Product similarity matrix

```

3.3 도면정보의 구조적 유사도 기반 트리 검증

본 절에서는 특허와 디자인권의 도면 정보를 토대로 구조적 유사도(Structural Similarity Index Measure, 이하 SSIM)를 평가하여 구축된 기술-디자인 트리를 검증한다.

SSIM은 사람의 눈이 이미지에서 구조적 정보를 추출하는데 집중한다는 것에 착안하여 휘도(Luminance), 명암비(Contrast), 구조(Structure)로 이들의 유사도를 구하여 서로 곱함으로써 <Figure 7>과 같이 전체적인 이미지의 구조적 유사도를 측정한다(Wang et al., 2004). 따라서 $N \times N$ 크기의 window 안에서 original signal을 $x = x_i \mid i = 1, 2, \dots, N \times N$ distorted signal을 $y = y_i \mid i = 1, 2, \dots, N \times N$ 표현했을 때 식 (3)과 같이 휘도 $l(x, y)$, 명암비 $c(x, y)$, 구조 $s(x, y)$ 의 유사도를 구하고, 이들을 서로 곱함으로써 식 (4)과 같이 SSIM을 구한다.

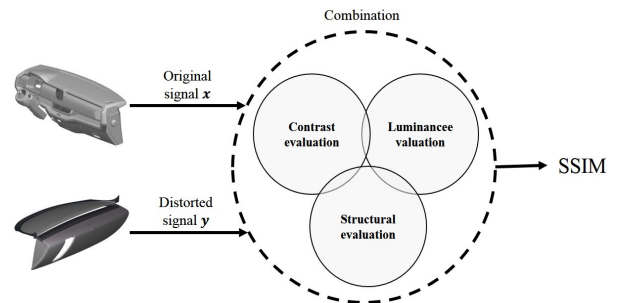


Figure 7. Processing of SSIM Image Assessment

$$l(x, y) = \frac{2\mu_x\mu_y}{\mu_x^2 + \mu_y^2}, c(x, y) = \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2}, s(x, y) = \frac{2\sigma_{xy}}{\sigma_x\sigma_y} \quad (3)$$

$$SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y) \\ = \frac{4\mu_x\mu_y\sigma_{xy}}{(\mu_x^2 + \mu_y^2)(\sigma_x^2 + \sigma_y^2)} \quad (4)$$

여기서 μ_x 와 μ_y 는 각각 signal x 와 y 의 평균값이며 이는 signal의 밝기를 반영한다. σ_x 와 σ_y 는 각각 신호 x 와 y 의 분산이며 이는 signal의 대비를 반영한다. σ_{xy} 는 signal x 와 y 사이의 공분산이며 이는 두 signal 간의 상관도를 반영한다. 이러한 휘도, 명암비, 구조도는 독립적이고 각각 1에 가까울수록 원본 이미지와 가깝고, 0에 가까울수록 원본 이미지와 다르다. 이처럼 정지영상에 기반을 둔 SSIM을 활용하여 특허의 평면도나 디자인권의 사시도와 육면도와 같은 이미지의 구조적 유사도를 측정하여 구축된 기술-디자인 트리를 검증할 수 있다. 이때, SSIM 값에 임계값을 주어 일정 수준 이하(δ_2)인 트리에 대해서만 관련 분야의 전문가의 분석을 의뢰하여 전문가의 개입을 최소화한다. 이로써 제안 방법론은 Algorithm 4로 정리된다. Algorithm 4의 라인 3부터 라인 6은 <Figure 3>의 Step 1인 기술 트리 구축과정이고, 라인 8부터 라인 11은 <Figure 3>의 Step 2인 디자인 트리 구축 과정이다. 라인 13와 라인 14는 구축된 트

리의 도면 정보 간 이미지 구조 유사도를 평가하여 임계값(δ_2)을 초과하는 경우 기술-디자인 트리의 대표 키워드 간 코사인 유사도를 반환함으로써 <Figure 3>의 Step 3를 수행한다.

-디자인 트리를 구축함으로써 요구기능과 제품 디자인 니즈에 관한 정보를 구조화하고, 이들 간의 관계를 분석함으로써 기술기회 발굴을 지원하기 위한 하나의 실시 예를 보인다.

Algorithm 4. Proposed Methodology

SSP : Semantic similarity matrix of between patents
 SSD : Semantic similarity matrix of between design rights
 TS : Technical similarity matrix
 PS : Product similarity matrix
 N : The number of documents
 C^{tech} : Technology cluster
 C^{prod} : Product cluster
 PI : Image of utility patent
 DI : Image of design right
 δ_2 : Threshold

1 **Input** : SSP, SSD, TS, PS
2
3 **for** $i = 1$ to N_{SSP} **do**:
4 **for** $j = 1$ to N_{SSP} **do**: # Step 1
5 **return** $C_{ij}^{tech} \leftarrow SSP_{ij} \times TS_{ij}$
6 $keyword_{C^{tech}} \leftarrow KeyBERT(C^{tech})$
7
8 **for** $i = 1$ to N_{SSD} **do**:
9 **for** $j = 1$ to N_{SSD} **do**: # Step 2
10 **return** $C_{ij}^{prod} \leftarrow SSD_{ij} \times PS_{ij}$
11 $keyword_{C^{prod}} \leftarrow KeyBERT(C^{prod})$
12
13 **if** $\frac{1}{l_{PI}} \sum SSIM(PI, PI) > \delta_2$:
14 **if** $\frac{1}{l_{DI}} \sum SSIM(DI, DI) > \delta_2$:
15 **return** $cos(keyword_{C^{tech}}, keyword_{C^{prod}})$ # Step 3
16 **else**:
17 **continue**
18 **else**:
19 **continue**
20 **end**
21
22 **Output**: Similarity between Technology tree & Design tree

4.1 분석사례 선정

본 실험에서는 제안 방법론의 기여점을 부각하기 위하여, 기존 연구(Song *et al.*, 2023)와 동일한 ‘차량 인스트루먼트 패널’을 분석 사례로 선정하였다. 또한 동일한 기술 분야-동일한 디자인 분야 내에서도 서로 상이한 트렌드를 갖는 시사점을 도출하거나 유사한 트렌드를 도출하기 위해 특허와 디자인권 모두 <Table 9>와 같은 동일한 검색식을 사용하였다. KIPRIS(Korea Intellectual Property Rights Information Service)의 IP 검색 서비스를 활용하여 연도 제한을 두지 않고 분석 사례를 수집한 결과 총 1,206건의 특허와 22건의 디자인권이 수집되었다. 특허의 경우 디자인권에 비해 월등히 많은 출원 건수를 보이고 있지만, <Figure 8>과 같이 출원 트렌드는 2017년 이후 급격히 감소하는 추세를 보인다. 반면, 디자인권의 경우 특허에 비해 절대적인 출원 건수는 적지만, <Figure 9>와 같이 2015년 이후 꾸준히 증가하는 추세를 보인다. 이는 기술 개발의 고도화로 기업들의 기술 수준이 평준화되면서 기술 중심 개발보다 먼저 디자인 중심 개발을 선행하는 경향이 반영된 것으로 사료된다.

Table 9. Search Formulae for Patent and Design Right

Database	Search formulae (KIPRIS DB)
Patent	(tractor OR equipment OR car OR automotive OR automobile OR vehicle OR truck OR machine OR "city bus" OR shuttle OR taxis) AND ("instrument panel" OR dashboard OR "instrument cluster" OR "dashboard organizer" OR "dashboard cluster" OR "dashboard assembly" OR "instrument organizer" OR "instrument assembly")
Design right	"dashboard assembly" OR "instrument organizer" OR "instrument assembly")

수집된 특허의 CPC의 사전분석 결과는 <Figure 10>과 같다. 본 실험에서는 기술 트리를 구성하는 기술의 세부 단위까지 고려하기 위해 <Table 6>과 <Table 7>을 통해 살펴본 것과 같이 CPC의 서브그룹 단위에서 분석을 진행하였다. 수집된 특허들은 3,053개의 고유한 기술 분야를 가지는 것으로 확인된다. 또한 각 특허는 평균 6개, 최대 43개의 기술 분야가 융합되어 출원된 최신 기술의 보고(cutting-edge technology)라고 할 수 있다. 이를 통해 특허의 CPC를 기술혁신과 연구개발의 동향에 대한 정량적 대리변수로 삼기에 적합함을 확인한다. 또한 수집된 디자인권의 LOC에 대한 사전분석 결과는 <Table 10>과 같다. 수집된 디자인권들은 대부분 제12류에 분포되어 있다. 이는 ‘차량 인스트루먼트 패널’ 분야에서 대부분의 디자인이 ‘운송 또는 승강 수

4. 제안 방법론의 실시 예

본 장에서는 한국의 특허와 디자인권 데이터를 기반으로 기술

단'을 목적으로 하는 제품인 것을 알 수 있다. 또한 디자인권의 도면은 육면도와 사시도 같은 정형화된 형태로 디자인 컨셉을 담고 있다. 이를 통해 유사한 제품의 디자인 트렌드를 식별하거나, 기술 트렌드와 상이하게 흘러가는 디자인 트렌드를 식별할 수도 있다.

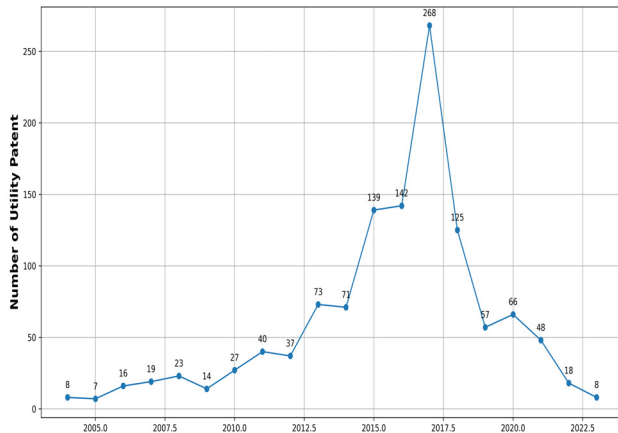


Figure 8. Patent Application Trends

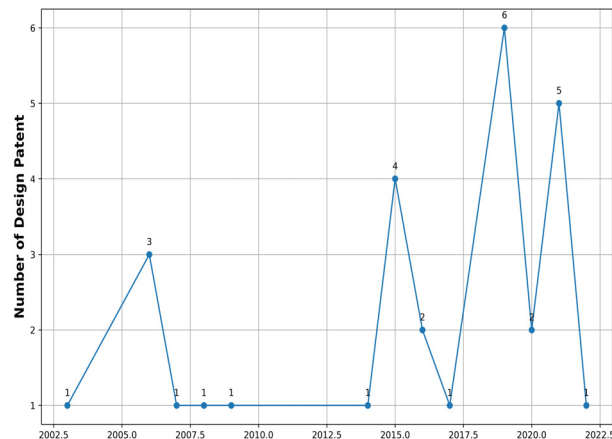


Figure 9. Design Application Trends

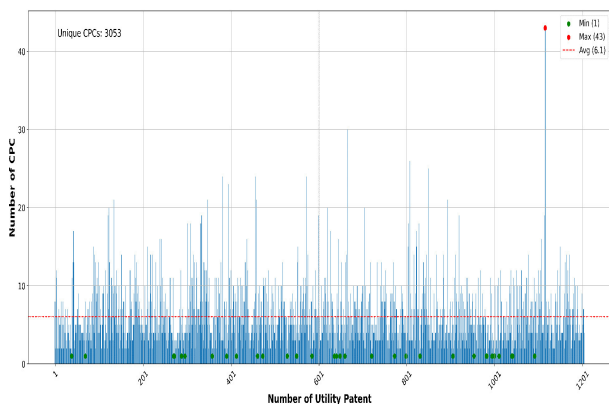


Figure 10. Pre-Analysis Results of CPC

Table 10. Pre-Analysis Results of LOC

Class	Description	Distribution
2	Articles of clothing and haberdashery	2
12	Means of transport or hoisting	13
13	Equipment for production, distribution or transformation of electricity	2
14	Recording, telecommunication or data processing equipment	4
21	Games, toys, tents and sports goods	1

4.2 기술 트리 구축 결과

본 절에서는 특허의 ‘발명의 명칭, 요약, 청구항’ 텍스트 정보로부터 의미적 유사도 분석을 위해 Python의 KeyBERT 라이브러리에서 제공하는 ‘distilbert-base-nli-mean-tokens’ 임베딩 모델을 활용한다. 키워드 추출 과정에서 하이퍼 파라미터인 추출할 상위 키워드 갯수(top_n)는 1로, 추출된 키워드의 N-gram 범위(keyphrase_ngram_range)는 (5,5)로 설정한다. 즉, 각 특허별로 5개의 단어로 구성된 핵심 구문을 1개씩 추출하여 의미적 유사도를 계산한다. 각 특허의 핵심 구문 간 코사인 유사도가 평균 이상인 경우(δ_1)를 시각화 하면 <Figure 11>의 (a)와 같다. 이때, 임계값(δ_1)는 분석하고자 하는 기술 분야나 전문가의 의견에 따라 조정할 수 있다. x 축과 y 축은 각 특허의 인덱스이고 대각선을 제외하고 노란색으로 표시된 부분이 각 특허가 가진 기술적 의미가 유사한 기술군이 된다. 다음으로, CPC를 활용한 기술적 유사도 분석 결과는 <Figure 11>의 (b)와 같다. 각 특허별로 공유한(intersection) 기술 분야가 있는 경우 노란색으로 표시된 것을 확인할 수 있다. 본 실험에서는 기존 방법(Sung *et al.*, 2023)과 동일한 조건으로 CPC 공유 여부(δ_2)를 임계값으로 두고 이진 벡터값(1, 0)을 곱하여 동일한 CPC의 공유가 없다면 유사도 값은 0으로 도출한다. 이때도 임계값은 분석 목적이나 도메인 전문가의 판단에 따라 조정될 수 있다. 마지막으로, <Figure 11>의 (a)와 (b)를 곱하여 도출된 (c)는 기술적 의미와 기술 분야가 모두 유사한 영역이다. 이 과정에서 코사인 유사도로 계산된 의미적 유사도와 자카드 유사도로 계산된 기술적 유사도의 값의 스케일이 다르기 때문에 0에서 1 사이의 값으로 정규화하였다. 이로써 ‘차량 인스트루먼트 패널’ 분야에서 구축된 기술 트리는 <Table 11>과 같이 구축된다. 구축된 기술 트리는 각 그룹을 구성하는 특허의 도면 이미지 구조적 유사도를 평가하여 검증된다. 예를 들어, <Table 11>의 두 번째 그룹의 출원번호 ‘10-2015-0155055’는 ‘운전자 보조 시스템 및 이를 포함하는 차량, 운전자 보조 시스템의 제어방법’에 관한 특허이고 ‘10-2017-0074902’은 ‘차량, 및 그 제어방법’에 관한 특허이다. 이들 특허는 ‘decide whether car offer control’ 키워드로 대표되는 기술적 의미와 ‘B60W40(특정 하위 유닛의 제어와 관련이 없는 도로 차량 주행 제어 시스템에 대한 {비직접 측정 가능} 주행 파라미터의 추정 또는 계산)’

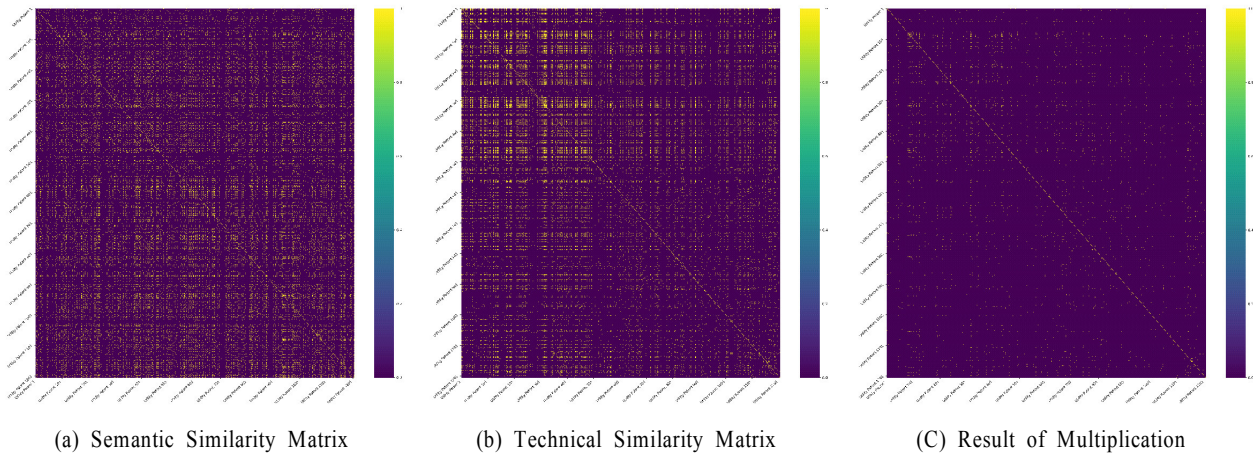


Figure 11. Visualization Results of the Technology Tree Construction Process

Table 11. Results of Technology Tree Construction

Technically Similar Group		SSIM	CPC	Main keyword
10-2013-0167820	10-2013-0106227	0.4897	G06F 3/011 G06F 3/048	express automotive instrument panel display
10-2015-0155055	10-2017-0074902	0.4555	B60W40/076 B60W40/072 B60W40/10 B60W40/06 B60W2556/50 B60W2556/60 B60W2050/0052	decide whether car offer control
⋮		⋮	⋮	⋮
10-2021-7031781	10-2019-7012462	0.9961	B60Q 1/52 B60Q 1/46 B60Q 1/44 B60R 16/023	light work strobing method microcontroller

과 ‘B60W2556(데이터와 관련된 입력 매개변수)’같은 CPC를 공유하는 기술적 유사도를 만족하는 유사한 기술군이 된다.

4.3 디자인 트리 구축 결과

본 절에서는 기술 트리 구축 과정과 동일한 실험 조건에서 디자인 트리를 구축한다. 수집된 디자인권의 ‘물품명칭, 창작

내용, 창작내용’ 텍스트 정보로부터 의미적 유사도 분석을 통해 유사도가 평균 이상인 경우(δ_1)를 시각화 하면 <Figure 12>의 (a)와 같다. 이때, 임계값 δ 는 분석하고자 하는 디자인 분야나 전문가의 의견에 따라 조정할 수 있다. x 축과 y 축은 디자인 권별 인덱스이고 대각선을 제외하고 노란색으로 표시된 부분이 디자인 컨셉이 유사한 디자인군이 된다. 다음으로, LOC를 활용하여 제품적 유사도를 비교한 결과는 <Figure 12>의 (b)와

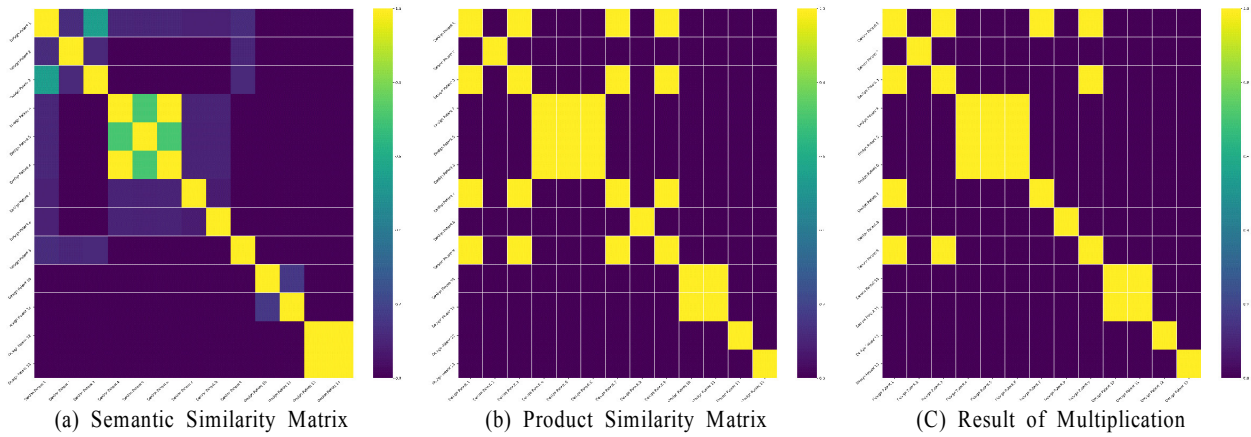




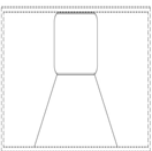
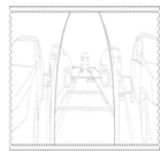
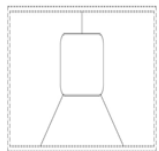




Figure 12. Visualization Results of the Technology Tree Construction Process

Table 12. Results of Design Tree Construction

Product Similar Group				SSIM	LOC	Main keyword
30-2009-0010785 	DM/208393 	30-2022-0000855 	DM/096472 	0.7784	12-16	instrument panel bottom display metal
30-2015-0014387 	30-2015-0016807 	30-2015-0014120 				
DM/205850 	DM/207095 			0.8165	13-03	car dashboard design indicator actuator

같다. 기술 트리 구축 과정과 마찬가지로 LOC 공유 여부(δ_1)를 임계값으로 두고 이진 벡터값(1, 0)을 곱하여 동일한 LOC의 공유가 없다면 유사도 값은 0으로 도출한다. 그러나 의미적 유사도 분석과 마찬가지로 분석 목적이나 도메인 전문가의 판단에 따라 조정될 수 있다. 마지막으로, <Figure 12>의 (a)와 (b)를 곱하여 도출된 (c)는 디자인 컨셉과 디자인 분야가 일치가 모두 유사한 영역이라고 할 수 있다. ‘차량 인스트루먼트 패널’ 분야에서 구축된 디자인 트리는 <Table 12>와 같다. 구축된 디자인 트리는 각 그룹을 구성하는 디자인권의 육면도나 사시도 이미지의 구조적 유사도를 평가하여 검증한다. <Table 12>의 세 번째 그룹의 국제등록번호 ‘DM/205850’은 ‘자동차 대시보드용 액추에이터’에 관한 디자인권이고 ‘DM/207095’은 ‘전기 액추에이터’에 관한 디자인권이다. 이들 디자인권은 ‘car dashboard design indicator actuator’ 키워드로 대표되는 디자인적 컨

셉과 ‘13-03(전력 공급 및 제어기기)’같은 LOC를 공유하는 제품적 유사도를 만족하는 유사한 디자인군이 된다.

4.4 개발 기회 영역 도출 결과

본 절에서는 디자인 트리와 특히 기술 트리를 비교하여 개발 기회 영역을 도출한다. 이를 위해서 디자인 컨셉을 담은 대표 키워드와 기술적 의미를 담은 대표 키워드간 코사인 유사도를 통해 트리 간 유사성을 분석한다. 예를 들어, <Table 13>의 LOC ‘12-16’에 해당하는 디자인군은 ‘운송 또는 승강수단(Class 12)’ 분야에서도 ‘다른 류나 군에 포함되지 않는 운송수단용 부품, 장비 부속품(Subclass 16)’을 목적으로 하는 디자인 영역이다. 해당 디자인 영역의 컨셉은 ‘instrument panel bottom display metal’로 하고 있으므로, CPC ‘B60W50/14(운전자에게 알리기 위한 수단, 운전자

Table 13. Result of Similarity Analysis between Technology-tree and Design-tree

Design-tree		Similarity		Technology-tree	
LOC	Main keyword	Measure	Value	CPC	Main keyword
12-16	instrument panel bottom display metal	High	0.8156	B60K35/00 B60W2050/146 B60W30/14 B60W30/18163 B60W10/20 B60W40/02 B60W40/105 B60Y2400/92 B60R16/02	display visual information execution screen
		Low	0.4623	B60W50/14	video driver working device assistant
14-04	image design car display panel	High	0.7844	B60K35/00	display system claim vibrating plate
		Low	0.4662	C07D209/82 C09K11/06	use chemical dibenzo carbazole synthesis
13-03	car dashboard design indicator actuator	High	0.7903	B60R2011/0057	contain guide body car door
		Low	0.4899	B60W50/14	video driver working device assistant

에게 경고하기 위한 수단 또는 운전자의 개입을 촉구하기 위한 수단)로 분류된 기술군보다는 ‘display visual information execution screen’과 같은 기술적 의미를 담고 있는 ‘B60W2050/146 (디스플레이 수단)’, ‘B60W30/14(속도 제한)’, ‘B60R16/02(차량의 상대적으로 움직이는 부품사이의 것, 예. 핸들과 칼럼간에 있어서의 것)’ 등과 같은 보다 구체적인 기술이나 그러한 신기술의 개발이 요구된다. 또한, CPC ‘B60R2011/0057(자기 수단을 이용하는 것)’에 해당하는 기술군은 ‘전기의 발전, 공급 또는 변전을 위한 장치(Class 13)’ 중에서도 ‘전력 공급 및 제어기기(Subclass 03)’과 관련된 신제품의 개발이 요구되고, 개발 과정에서 ‘car dashboard design indicator actuator’을 컨셉으로 한 디자인 영역의 트렌드를 참고할 수 있을 것이다.

5. 결론 및 향후 연구

5.1 결론

최근 기술 수준의 고도화와 평준화로 기술 중심 개발보다 디자인 중심 개발을 선행하는 경향이 있다. 이에 특허와 디자인 인권의 분석을 통해 경쟁자의 연구개발을 이해하고, 기업의 시장 기회를 파악하는 것이 필요하다. 그러나 특허와 디자인 권을 동시에 분석한 연구가 거의 없고, 기존의 방법으로는 분석 과정에서 전문가 의존도가 높거나 개입이 빈번하여 기술기회발굴을 지원하는 데 한계가 있다. 특히 속지주의를 따르는 산업재산권의 특성상 국가별로 데이터의 구조와 형태가 달라

텍스트 데이터가 복잡한 구조와 형태일 경우 해당 문서를 대표하는 주요 키워드를 추출하는 데 어려움이 있다. 본 연구에서는 텍스트 데이터의 구조와 형태가 복잡한 경우에도 전문가의 개입을 최소화하는 텍스트 분석 기반의 기술-디자인 트리 구축 방법을 제시한다. 이를 위해서 각 특허와 디자인권에서 KeyBERT 모델을 활용하여 문맥을 고려한 주요 키워드를 추출하여 의미론적 유사도를 분석한다. 다음으로, 특허의 CPC와 디자인권의 LOC를 활용하여 기술적·제품적 유사도를 통해 기술-디자인 트리를 구축한다. 마지막으로, 구축된 트리는 특허와 디자인권의 도면 이미지의 구조적 유사도를 평가하여 검증한다. 실험 결과 ‘차량 인스트루먼트 패널’ 분야에서 요구기능과 제품 디자인 니즈를 계층적으로 구조화할 수 있다. 또한 이들 간의 유사성을 분석함으로써 기술 개발에 국한되지 않고, 유사한 제품 디자인의 트렌드 영역도 파악할 수 있다. 이를 통해 학술적 관점에서 제안 방법론의 유용성을 보이고, 실무적 관점에서는 연구개발 주체의 보유역량(기술/제품)에 기반한 신기술 또는 신제품 기획의 발굴을 지원할 수 있음을 시사한다.

5.2 향후 연구

향후 연구에서는 제안 방법론의 한계를 보완하고, 더욱 타당한 분석을 위한 노력이 필요할 것으로 사료된다. 먼저, 제안 방법론에서 활용된 모델의 특성상 분석 분야(domain)가 변하거나 새로운 특허, 디자인권 데이터가 추가될 때 기술 트리와 디자인 트리를 실시간으로 구축하는 데 어려움이 있다. 따라서 추후 연구에서는 AutoML(Automated Machine Learning)을 활용한다면 기술과 디자인 트렌드를 실시간으로 반영한 기술-디자인 트리를 구축할 수 있을 것이다. 또한 기술-디자인 트리 간 유사성을 비교하는 과정에서 요구기능과 니즈를 대표하는 단어들의 가중치가 다른 문제도 있다. 따라서 유사도 계산 과정에서 대표 키워드의 가중치를 고려하여 더욱 정교한 정렬이 요구된다. 나아가 특허나 디자인권 외에도 다양한 IP를 분석에 활용한다면, 더욱 타당하고 유용성 있는 기술기회 발굴 지원이 가능할 것으로 기대된다.

참고문헌

- Ba, Z., Meng, K., Ma, Y., and Xia, Y. (2024), Discovering technological opportunities by identifying dynamic structure-coupling patterns and lead-lag distance between science and technology, *Technological Forecasting and Social Change*, **200**, 123147.
- Chae, S. H. and Gim, J. (2018), A Study of CPC-based Technology Classification Analysis Model of Patents, *The Journal of the Korea Contents Association*, **18**(10), 443-452.
- Chen, A. and Chen, R. (2007), Design patent map: An innovative measure for corporative design strategies, *Engineering Management Journal*, **19**(3), 14-29.
- Chen, R. (2009), Design patent map visualization display, *Expert Systems with Applications*, **36**(10), 12362-12374.
- Choi, D. I. and Lim, C. (2023), Analysis of trends in deep learning and reinforcement learning, *Journal of The Korea Society of Computer and Information*, **28**(10), 55-65.
- Choi, J., Jeong, B., and Yoon, J. (2019), Technology opportunity discovery under the dynamic change of focus technology fields: Application of sequential pattern mining to patent classifications, *Technological Forecasting and Social Change*, **148**, 119737.
- Choi, S., Yoon, J., Kim, K., Lee, J. Y., and Kim, C. H. (2011), SAO network analysis of patents for technology trends identification: A case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells, *Scientometrics*, **88**(3), 863-883.
- Cozzens, S., Gatchair, S., Kang, J., Kim, K. S., Lee, H. J., Ordóñez, G., and Porter, A. (2010), Emerging technologies: Quantitative identification and measurement, *Technology Analysis & Strategic Management*, **22**(3), 361-376.
- Feng, L., Niu, Y., Liu, Z., Wang, J., and Zhang, K. (2019), Discovering technology opportunity by keyword-based patent analysis: A hybrid approach of morphology analysis and USIT, *Sustainability*, **12**(1), 136.
- Hu, J., Li, S., Yao, Y., Yu, L., Yang, G., and Hu, J. (2018), Patent keyword extraction algorithm based on distributed representation for patent classification, *Entropy*, **20**(2), 104.
- Huang, S. H., Ke, H. R., and Yang, W. P. (2008), Structure clustering for chinese patent documents, *Expert Systems with Applications*, **34**(4), 2290-2297.
- Jeong, B., Yoon, J., and Lee, J. M. (2019), Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis, *International Journal of Information Management*, **48**, 280-290.
- Kang, W., Lee, S., and Choi, S. (2023), A matrix factorization-based recommendation approach with SBERT embeddings, *The Journal of Korean Institute of Information Technology*, **21**(11), 203-211.
- Kim, C. H. (2010), Research for reformulating protection scope of design right, *Korean Intellectual Property Office*, 217.
- Kim, C. and Shin, J. (2013), Linking capability-based technology opportunities analysis to R&D planning, *Fall Conference of Korean Institute of Industrial Engineers*, 1249-1258.
- Kim, C., Lee, H., Seol, H., and Lee, C. (2011), Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach, *Expert Systems with Applications*, **38**(10), 12559-12564.
- Kim, S. (2012), Trends for design industry in industrial property statistics: On changes and promotion system of design and trademark registration, *Journal of Integrated Design Research*, **11**(4), 9-22.
- Lee, C., Hong, S., and Kim, J. (2021), Anticipating multi-technology convergence: A machine learning approach using patent information, *Scientometrics*, **126**, 1867-1896.
- Lee, H., Kim, C., Cho, H., and Park, Y. (2009), An ANP-based technology network for identification of core technologies: A case of telecommunication technologies, *Expert Systems with Applications*, **36**(1), 894-908.
- Lee, S., Yoon, B., and Park, Y. (2009), An approach to discovering new technology opportunities: Keyword-based patent map approach, *Technovation*, **29**(6-7), 481-497.
- Li, J., Fan, Q. N., and Zhang, K. (2007), Keyword extraction based on tf/idf for chinese news document, *Wuhan University Journal of Natural Sciences*, **12**(5), 917-921.

- Mihalcea, R. and Tarau, P. (2004, July), TextRank: Bringing order into text. *In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 404-411.
- Park, J., Sim, W., Lee, S., Ko, B., and Noh, H. (2022), A study on automatic CPC classification based on Korean patent sentence: A deep learning approach using artificial intelligence language model KorPatBERT, *The Journal of Intellectual Property*, **17**(3), 209-256.
- Park, Y. and Yoon, J. (2017), Application technology opportunity discovery from technology portfolios: Use of patent classification and collaborative filtering, *Technological Forecasting and Social Change*, **118**, 170-183.
- Reimers, N. and Gurevych, I. (2019), Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.
- Rodriguez, A., Tosiya, A., Kim, B., Choi, J., Lee, J. M., Coh, B. Y., and Jeong, M. K. (2016), Patent clustering and outlier ranking methodologies for attributed patent citation networks for technology opportunity discovery, *IEEE Transactions on Engineering Management*, **63**(4), 426-437.
- Rose, S. J., Cowley, W. E., Crow, V. L., and Cramer, N. O. (2012), U.S. Patent No. 8,131,735. Washington, DC: U.S. Patent and Trademark Office.
- Seo, W. (2021), A study on technology opportunity discovery by analyzing patent-based technology themes, *The Journal of Intellectual Property*, **16**(2), 205-248.
- Seo, W. (2022), Discovering opportunities for technological convergence between technology topics: Focusing on wearable technology cases, *The Journal of Intellectual Property*, **17**(4), 141-146.
- Song, K., Ahn, J., Park, S., Lee, S., and Kim, J. (2023), Discovering new technology opportunities from convergence analysis of utility patents and design patents, *Journal of the Korean Institute of Industrial Engineers*, **49**(5), 395-405.
- Trappey, C. V., Wu, H. Y., Taghaboni-Dutta, F., and Trappey, A. J. (2011), Using patent data for technology forecasting: China RFID patent analysis, *Advanced Engineering Informatics*, **25**(1), 53-64.
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004), Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing*, **13**(4), 600-612.
- Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C., and Nevill-Manning, C. G. (1999, August). KEA: Practical automatic keyphrase extraction, *In Proceedings of The Fourth ACM Conference on Digital Libraries*, 254-255.
- Yoon, J. and Kim, B. (2024), Text analytics-based business area identification for patent-trademark linkage business intelligence, *Journal of the Korean Institute of Industrial Engineers*, **50**(1), 47-63.
- Yoon, J. and Kim, K. (2012), Detecting signals of new technological opportunities using semantic patent analysis and outlier detection, *Scientometrics*, **90**(2), 445-461.
- Yoon, J., Park, H., Seo, W., Lee, J. M., Coh, B. Y., and Kim, J. (2015), Technology opportunity discovery (TOD) from existing technologies and products: A function-based TOD framework, *Technological Forecasting and Social Change*, **100**, 153-167.
- Yoon, J., Seo, W., Ko, B., and Lee, J. (2013), Technological capability-based technology opportunity discovery, *Fall Conference of Korean Institute of Industrial Engineers*, 1171-1179.
- Yoon, W. and Kim, J. (2023), Changes in automobile interior design based on design patent classification, *The Journal of Industrial Design Studies*, **17**(1), 63-70.
- Yu, E. J., Kim, J. C., Lee, C. Y., and Kim, N. G. (2012), Using ontologies for semantic text mining, *The Journal of Information Systems*, **21**(3), 137-161.
- Zhang, C. (2008), Automatic keyword extraction from documents using conditional random fields, *Journal of Computational Information Systems*, **4**(3), 1169-1180.

저자소개

윤주호: 명지전문대학 산업시스템경영공학과에서 학사학위를 취득하고, 한양대학교 ERICA 산업경영공학과 석박사통합과정에 재학 중이다. 주요 관심분야는 통계 학습, 기계 학습, 데이터마이닝, 기술경영이다.

권세훈: 한양대학교 산업경영공학과에서 학사학위를 취득하고 한양대학교 ERICA 산업경영공학과 석사과정에 재학 중이다. 주요 관심분야는 통계 학습, 신뢰성 공학이다.

김병훈: 렛저스 대학교에서 산업시스템공학 박사학위를 취득하고, 현재 한양대학교 ERICA 산업경영공학과에 재직 중이다. 주요 관심분야는 통계 데이터 마이닝 방법론 개발, 반도체 제조 공정용 데이터 마이닝 모델 개발, 그래프 마이닝이다.