

특허 NLP 분석 기반 AI 컨택센터 IP-R&D 방안

김정희*, 김영민**

AI Contact Center IP-R&D Scheme based on Patent NLP Analysis

JungHeui Kim*, Young-Min Kim**

요약

4차 산업혁명 시대에 인공지능, 빅데이터 분석기술과 같은 정보통신기술(ICT)의 현명한 사용은 생산성을 향상하고 혁신을 추진하는데 핵심적인 요소가 되어가고 있다. 본 연구에서는 기술의 발전과 변화의 트렌드를 파악하고 R&D 개발 전략을 수립하는데 활용되는 특허 정보를 인공지능 기술을 이용해 분석한다. 자연어처리(NLP, Natural Language Processing) 기술을 이용한 특허분석을 통해 핵심 기술 키워드를 도출하고, 관련된 핵심특허를 효과적으로 식별하는 방안을 제시한다. 전통적인 콜센터에서 발전하여 AI 응용 기술이 적용된 AI 컨택센터에 관련된 특허문서를 수집하고 NLP 분석기를 이용해 핵심 키워드를 추출한 후, 추출된 핵심기술 키워드로 특허문서를 재검색하는 전략을 활용하여, IP-R&D를 위한 핵심 특허를 선별하였다. 이 방법론은 상세 기술 특허 탐색시 어휘 검색의 오탐율을 감소시키는 디지털 R&D 프로세스를 제안한다.

Abstract

In the era of the 4th industrial revolution, smart use of information and communication technologies(ICT), such as artificial intelligence and big data analysis, is becoming a key factor in improving productivity and performance and driving innovation. In this study, we use artificial intelligence technology to analyze patent information used to identify trends in technological development and change and establish R&D development strategies. We derive core technology keywords through patent analysis using Natural Language Processing(NLP) technology, and suggest ways to effectively identify related core patents. We collected patent documents related to AI contact centers that have evolved from traditional call centers and applied AI application technology, extracted key keywords using NLP analyzer, and then utilized a strategy of re-searching patent documents with the extracted core technology keywords to select patents for IP-R&D. This methodology proposes a digital R&D process that reduces the false positive rate of lexical search when searching patents for detailed technology.

Keywords

patent analysis, natural language processing, IP-R&D, contact center

* 한양대학교 기술경영대학원 박사수료
- ORCID: <https://orcid.org/0000-0001-8328-8342>
** 한양대학교 기술경영대학원 부교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-6914-901X>

• Received: Nov. 14, 2023, Revised: Dec. 20, 2023, Accepted: Dec. 23, 2023
• Corresponding Author: Young-Min Kim
Graduate School of Technology & Innovation Management,
Hanyang University, Korea
Tel.: +82-2-2220-2537, Email: yngmnkim@hanyang.ac.kr

I. 서 론

4차 산업혁명 시대에 인공지능의 활용은 생산 프로세스를 최적화하고 기계 고장을 예측하며 디지털 기반의 새로운 스마트 서비스를 제공하며 중요성이 더욱 커지고 있다[1]. 이러한 인공지능, 빅데이터 분석기술과 같은 정보통신기술(ICT)의 현명한 사용은 생산성을 향상하고 혁신을 추진하는데 핵심적인 요소가 되어가고 있다[2].

본 연구에서는 기업의 R&D 개발계획 수립에 필요한 핵심특허 분석을 위해 인공지능을 활용한 기술 분석 방법론을 제시하고, R&D 개발 프로세스의 소요시간을 단축시킬 수 있는 방안을 제시하고자 한다. 연구 사례로는 기업에서 고객 상담 서비스를 제공하는 전통적인 콜센터(Call center) 서비스에 AI 기술을 적용한 형태인 AI 컨택센터(AI contact center)를 대상으로 한다. AI 컨택센터에 적용된 핵심기술을 분석하기 위해 인공지능 자연어처리(NLP, Natural Language Processing) 기술을 이용하여 관련된 유효 특허를 식별하고, 이 유효 특허에서 핵심 키워드를 추출하여, 추출된 키워드로 기술특허를 재분석하여 가치가 있는 핵심특허를 발굴한다. 본 분석 방법론은 기업이 기술 개발에 앞서 시장 및 경쟁사의 기술 동향을 파악하고 기술 개발시의 문제점과 해결을 위한 핵심 기술을 효율적으로 파악할 수 있게 한다. 본 연구를 통해 사전 정보 분석 단계를 거쳐서 기술 개발 전략을 수립한 후 연구개발을 진행하는 디지털 R&D 프로세스의 방법론을 제안하고자 한다.

II. 이론적 배경 및 연구의 필요성

R&D 프로세스와 활동은 신규 상품이나 신규서비스를 시장에 소개하고 기존 제품이나 서비스를 혁신하기 위해 수행하는 활동들로 제품과 서비스 개발을 위한 기획 식별, 계획, 설계, 개발 및 시장 출시에 이르는 단계별 프로세스로 구성된다[3][4]. 다양한 R&D 개발 프로세스가 도입되어 산업에 활용되고 있다. Cooper 와 Edgett는 신제품 개발의 핵심 성공 요소들을 정의하고, 아이디어에서부터 제품 출시까지 신제품 개발 프로젝트 관리 모델로 Stage-Gate 프로세스를 제시했다. Cooper는 새로운

제품과 프로세스 개발 프로젝트에 적합한 Stage-Gate 프로세스를 독자적으로 제안하기도 하였는데, 이는 프로젝트 범위 및 계획 정의 단계인 project scoping, 아이디어의 기술적 타당성을 시연하는 technical assessment, 실험 계획의 구축 단계인 detailed investigation의 3개 stage로 구성이 된다[5]. 첫 번째 단계인 stage 1의 활동에는 기술문헌 조사, 특허와 IP 조사, 대안 평가, 예비 기술평가 등이 포함되는데, 이 때 기술문헌들은 최신 및 경쟁 기술을 탐색하는데 있어 중요한 빅데이터 소스가 된다[6].

지적재산(IP, Intellectual Property)은 인간 지성의 무형적 창조물을 포함하는 재산 범주로, 특허, 저작권, 상표 및 영업비밀 등의 유형이 있다[7]. 이 중 특허는 발명의 공개에 대한 대가로 정해진 기간 동안 다른 사람이 발명을 제작, 사용, 판매하는 것을 배제하도록 법적 권리를 소유권자에게 부여하는 지적재산권의 일종이다[8]. 특허분석은 기술정보를 파악하는데 활용될 뿐만 아니라 기술 경쟁력과 기술 개발 방향성을 연구하는데도 유용하여, 기업 및 국가의 R&D 방향 설계, 투자 사업전략 등 혁신전략을 수립하는데 객관적인 지표로 활용되고 있다 [9]-[12].

특허 데이터를 사용할 때 주요 문제 중 하나는 특정 도메인에서 특허를 검색하는 방법이다. 국제 특허 분류(IPC) 또는 미국 특허 분류(UPC)를 사용하거나 키워드 검색, 직접 특허 문서를 읽는 방법들이 사용될 수 있다. 이때 많은 양의 특허를 식별하기란 어렵고[13][14], 클래스 분류가 존재하지 않는 경우, 키워드 검색 방법이 일반적으로 선택 되어왔다[15][16]. 그러나 기존의 키워드 검색 방식은 문서 내에 해당 키워드가 명시적으로 포함되지 않고, 유사어나 동의어를 사용하는 경우 문서가 검색 결과에 포함되지 않는 오탐지(False negative) 문제를 내재하고 있다. 최근 인공지능 기술의 발전으로 NLP 기법을 활용하면 문맥의 의미를 파악하여 유사어나 동의어 검색이 가능하여 검색에 오탐율을 감소시킬 수 있다. 본 연구에서는 인공지능 기술이 적용된 키워드 추출기를 사용하여, 동일한 단어로 매칭되지 않는 유사어와 동의어를 포함해서 주요 키워드를 추출하여, IP 기반 R&D에 필요한 핵심특허를 효과적으로 식별해내는 방법론을 제시한다.

III. 분석방법론

3.1 분석대상 기술

실증분석에서는 AI 컨택센터 기술을 분석대상 기술로 정의하고 관련된 특허문서들 중 핵심특허를 빠르게 선별하는 방안을 제시한다. 인공지능을 활용한 혁신에 대한 니즈, COVID-19로 인한 비대면 업무의 강화, 고객 경험에 대한 가치 인식 등으로 기존의 콜센터는 AI 기술이 적용된 AI 컨택센터로 전환 및 확장이 빠르게 이루어지고 있다[17]. 표 1은 전통적인 콜센터 기술 목록을 나타낸다. 기반 기술에는 Private Branch Exchange 비즈니스 전화 시스템, Interactive Voice Response 기술, Calling Line Identity, Computer Telephony Integration 등이 포함된다[18]-[20]. AI 컨택센터는 고객 상호 작용을 관리하는 조직으로, 전화로만 고객 요청을 받는 전통적인 콜센터와 달리 전화, 문자, 웹, 채팅, 이메일, 메시징 앱, 소셜 미디어 등 여러 채널을 통해 인바운드 및 아웃바운드 고객 커뮤니케이션을 처리한다. 전통적인 콜센터 인프라를 기반으로 한 가상 상담

원(Virtual agent), 상담 지원 서비스(Virtual assistant), 상담 분석(AI analytics), 챗봇(Chatbot) 등을 활용하는 형태이다[21]-[23].

3.2 데이터 수집 및 처리

분석 대상 데이터는 AI 컨택센터 구현에 필요한 AI기술 특허문서와 특허에 참조된 문헌을 대상으로 한다. 데이터는 특허 데이터베이스인 PATSTATS와 특허 데이터베이스 검색을 제공하는 공개된 인터넷 사이트인 Google Patents를 통하여 2001년~2021년의 특허 정보를 수집하였다.

데이터 수집 단계는 Fig. 1과 같이 총 세 단계로 구성이 된다. 1단계는 AI 컨택센터와 관련된 AI 기술 특허를 검색하기 위해 PATSTAT의 특허 title 정보 테이블과 abstract 테이블을 ‘+artificial +intelligence +((+contact +center) (+call +center))’ 로 키워드 검색을 수행한 후 두 테이블 검색 결과에서 중복을 제거하며 64건의 특허 출원 정보를 수집하였다(그림 2). 이 중 키워드 검색에서 발생할 수 있는 False Positive (예를 들어, ‘call’과 ‘center’가 복합어가 아닌 개별 단어로 검색된 특허가 포함됨)를

표 1. 전통적인 콜센터 기술 목록
Table 1. Technology lists of traditional call center

Category	Technology	Description
Infrastructure technology	Private Branch Exchange(PBX)	A business phone system that provides more features and tools than a typical phone system.
	Interactive Voice Response (IVR)	A technology that asks you to press a button on your phone's keypad to select the desired service and directs the call to the most appropriate agent
	Calling Line Identity(CLI)	Technology that uses CTI software to match a customer's phone number with past call records
	Computer Telephony Integration (CTI)	Technology that automatically combines the customer's voice and detailed data and displays them on the agent's screen when a call is connected
	Call routing	A process designed to ensure that each call is assigned to an agent with appropriate skills and prior knowledge related to the customer's issue.
	Automatic Call Distributor(ACD)	A special phone system that handles inbound calls, recognizes and answers calls, and assigns calls to the most appropriate attendant.
	Call recording	Technology to record and capture phone calls between all customers and agents
	Voice Response Unit(VRU)	Interactive technology that allows humans to communicate with computers through voice or duotone multifrequency
	Document Management System (DMS)	A system that opens and scans emails for electronic document distribution
Customer services	Customer relationship management	An automated system that helps identify needs, interact, and improve sales to provide optimized services for each customer type
	Customer experience management	Procedures for tracking consultations between customers and agents
	Web self-service	Technologies that enable customers and agents to access information and perform tasks over the Internet

제거하기 위해 특허 title 정보 테이블과 abstract 테이블을 더 엄격한 검색 조건으로 키워드 검색을 수행하여 기본 데이터로 사용할 40건의 특허 출원정보를 추출하였다. 이 40건의 출원정보는 다음 단계 분석을 위한 기본 데이터로 사용하므로 출원번호, 패밀리정보 등의 특허출원에 관련된 추가적인 정보를 수집하였다. 1단계 데이터 수집에 사용된 SQL 쿼리문은 표 2 와 같다.

2단계는 PATSTAT에서 수집된 특허 출원정보 및 등록정보 등을 이용하여 Google Patents에서 특허문서 40건을 수집한다(그림 3).

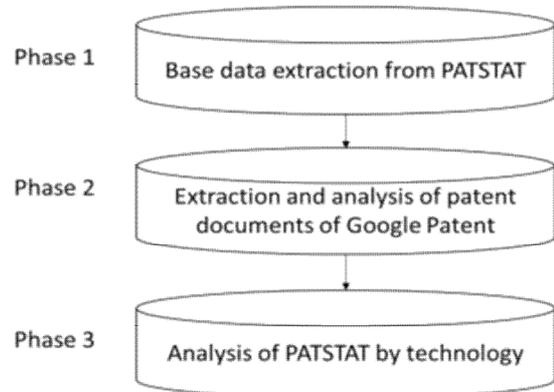


그림 1. 데이터 수집 절차
Fig. 1. Data collection procedure

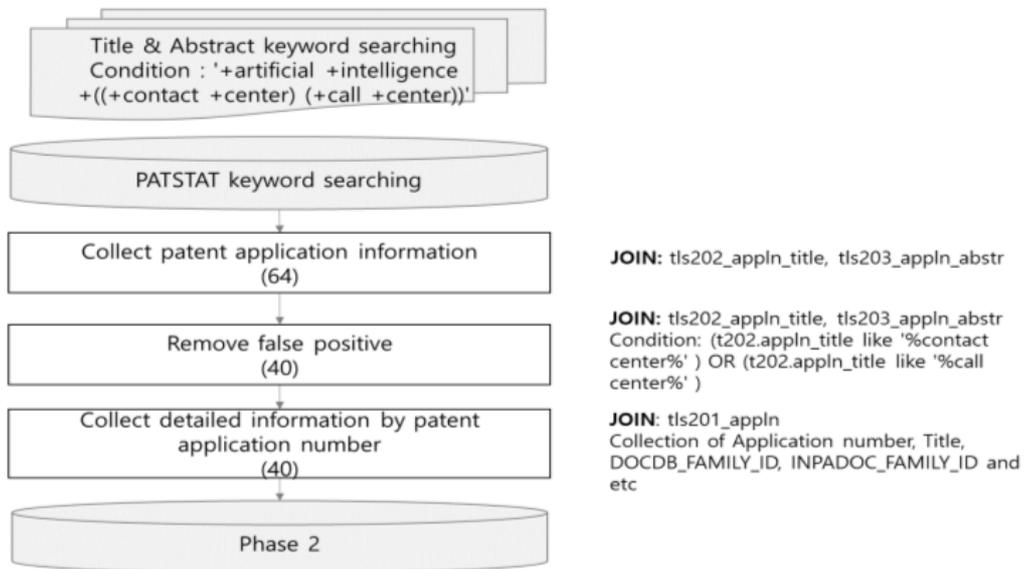


그림 2. 데이터 수집 절차 1단계
Fig. 2. Phase 1 of data collection process

표 2. AI 컨택센터 특허수집 SQL쿼리문
Table 2. SQL queries to collect patents for AI contact center

Matched keyword searching with BOOLEAN MODE	Removed false positive using WHERE predicate
(Extract patent appln_id related to AI contact center through keyword search)	(Search by matching 'call center' and 'contact center' with one word)
<pre> CREATE TABLE students.tbc1 select appln_id from t1s202_appln_title where match (appln_title against ('+artificial +intelligence +((+contact +center) (+call +center))' IN BOOLEAN MODE) group by appln_id UNION select appln_id from t1s203_appln_abstr where match (appln_abstract) against ('+artificial +intelligence +((+contact +center) (+call +center))' IN BOOLEAN MODE) group by appln_id ; </pre>	<pre> CREATE TABLE students.tbc4 select appln_id from students.tbc1 left outer join patstat2021s.t1s202_appln_title t202 using(appln_id) where (t202.appln_title like '%contact center%') OR (t202.appln_title like '%call center%') UNION select appln_id from students.tbc1 left outer join patstat2021s.t1s203_appln_abstr t203 using(appln_id) where (t203.appln_abstract like '%contact center%') OR (t203.appln_abstract like '%call center%') ; </pre>

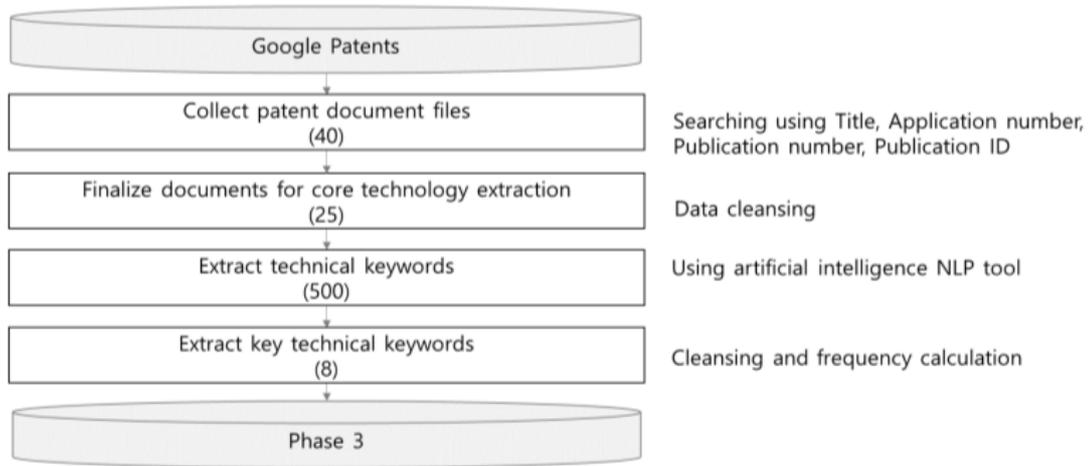


그림 3. 데이터 수집 절차 2단계
Fig. 3. Phase 2 of data collection process

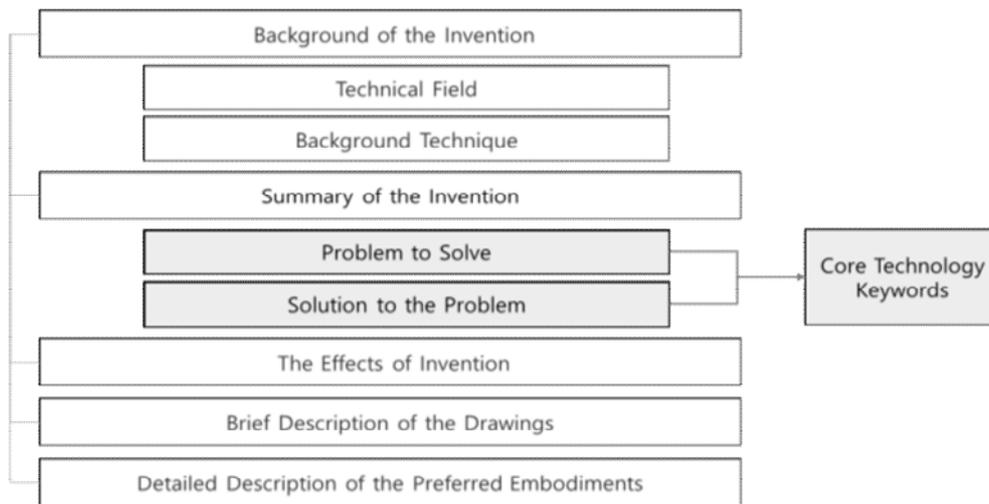


그림 4. 특허 등록 양식의 목차 중 분석에 활용된 항목
Fig. 4. Used lists in contents of the patent registration form

2단계에서는 핵심 기술 키워드를 추출하기 위해 특허 등록 양식의 콘텐츠 중 발명의 내용에 기술된 텍스트 문장을 입력으로 사용하는데, PATSTAT의 데이터베이스에는 발명의 내용을 포함한 특허 별 상세정보가 제공되지 않기 때문에 완전한 특허문서 데이터를 수집하기 위해 Google Patents에서 원본 특허문서를 수집하였다. 이 특허문서들은 기술 키워드 추출을 위해 사용되기 때문에 동일한 특허문서가 이중으로 투입되어 특정 키워드에 가중치가 적용되는 것을 방지할 필요가 있다. 이를 위해 DOCDB Family ID가 동일한 특허 패밀리는 1개의 특허문서만 키워드 추출 대상으로 포함하였다. 40개의 특허문서에서 application title을 수작업으로 점검

하여 False Positive인 Doc#22와 특허 포기된 Doc#38는 기술 키워드 추출대상에서 제외하여 총 25건의 핵심 기술 추출대상 특허문서를 확정하였다(표 3).

수집된 특허문서들은 표준특허 등록 양식의 목차를 따르며 작성되어 있는데(그림 4), 25건의 특허문서에 대해 표준특허 등록 양식의 목차 중 ‘Summary of the Invention’ 하위수준의 목차인 [Problem to Solve] 및 [Solution to the Problem]에 해당하는 텍스트를 인공지능 NLP 분석기에 입력에 각 문서별로 20개씩 총 500개의 주요 키워드를 추출하였다. 이때 동일한 언어로 키워드 추출을 위해 Google Patent에서 제공하는 영문 텍스트 또는 영문으로 번역된 특허문서 텍스트를 일괄적으로 수집하였다.

표 3. 기술 추출 대상 특허문서 목록

Table 3. List of patent documents subject to technology extraction

Doc#	Eligibility/ reason for exclusion	Application title	Application ID
1	Y	Double-channel session information association system and method	539122481
2	Y	METHOD AND SYSTEM FOR SOFT SKILLS-BASED CALL ROUTING IN CONTACT CENTERS	531647825
3	#2 family	METHOD AND SYSTEM FOR SOFT SKILLS-BASED CALL ROUTING IN CONTACT CENTERS	542264399
4	Y	Recording processing method based on convolutional neural network and connectivity time sequence classification	532246226
5	Y	Call center scheduling method based on customer emotion analysis	530089699
6	Y	ARTIFICIAL INTELLIGENCE DRIVEN SENTIMENT ANALYSIS DURING ON-HOLD CALL STATE IN CONTACT CENTER	544115949
7	Y	High-efficiency telephone calling system	519311495
8	#2 family	METHOD AND SYSTEM FOR SOFT SKILLS-BASED CALL ROUTING IN CONTACT CENTERS	540807028
9	Y	Telephone network-based interaction system combining artificial intelligence and manual customer service	518520750
10	Y	ARTIFICIAL INTELLIGENCE DRIVEN CALL ROUTING SYSTEM	538593069
11	Y	intelligent outbound system and method	515024308
12	Y	Artificial intelligence calling system	512236768
13	Y	Call center resource management method and apparatus, electronic device, and storage medium	511502936
14	Y	Artificial intelligence call center system	504115643
15	Y	MONITORING AGENT OVERSIGHT OF ARTIFICIAL INTELLIGENCE CONTENT IN A CONTACT CENTER	498575150
16	#15 family	MONITORING AGENT OVERSIGHT OF ARTIFICIAL INTELLIGENCE CONTENT IN CONTACT CENTER	509212662
17	#15 family	Monitoring agent oversight of artificial intelligence content in contact center	509280993
18	Y	Method and System For Providing Call Center Sharing Service Based On Big Data and Artificial Intelligence	513914430
19	Y	ARTIFICIAL INTELLIGENCE SELF-LEARNING TRAINING SYSTEM TO AUTONOMOUSLY APPLY AND EVALUATE AGENT TRAINING IN A CONTACT CENTER	525267266
20	Y	SYSTEM AND METHOD FOR PROVIDING REVERSE SCRIPTING SERVICE BETWEEN SPEAKING AND TEXT FOR AI DEEP LEARNING	503586403
21	Y	Person-artificial intelligence collaboration Call Center System and service method	518955321
22	False Positive	DOCTOR MAT	518426944
23	Y	Intelligent automated agent for a contact center	486017805
24	#15 family	Monitoring agent oversight of artificial intelligence content in a contact center	493740691
25	#21 title, claims are the same	Person-artificial intelligence collaboration Call Center System and service method	507998882
26	Y	Connectivity Integration Management Method and Connected Car thereof	507589481
27	Y	SERVER FOR CALL CENTER USING ARTIFICIAL INTELLIGENCE	490453624
28	Y	METHOD FOR PROVIDING AUTOMATIC TEXT MESSAGE RESPONSE SERVICE	448601313
29	#23 family	INTELLIGENT AUTOMATED AGENT FOR A CONTACT CENTER	421987631
30	#23 family	INTELLIGENT AUTOMATED AGENT FOR A CONTACT CENTER	422001169
31	#23 family	Intelligent automated agent for a contact center	446673759
32	#23 family	Intelligent automated agent for a contact center	448126098
33	#23 family	INTELLIGENT AUTOMATED AGENT FOR A CONTACT CENTER	448704794
34	Y	Intelligent automated agent and interactive voice response for a contact center	421956761
35	#23 family	INTELLIGENT AUTOMATED AGENT FOR A CONTACT CENTER	422944875
36	Y	Intelligent voice script	417884907
37	Y	Intelligent voice	417884909
38	abandonment	Intelligent interactive voice response unit	47144291
39	Y	SERVICE REQUEST PROCESSING PERFORMED BY ARTIFICIAL INTELLIGENCE SYSTEMS IN CONJUNCTION WITH HUMAN INTERVENTION	45642915
40	#39 family	Service request processing performed by artificial intelligence systems in conjunction with human intervention	53141346

3.2.1 NLP 분석기 이용한 핵심 키워드 추출

자연어 처리 기반의 인공지능 분석 툴은 연구자가 문서를 직접 읽고 키워드를 추출하는 것 대비 사람의 주관적인 판단의 개입을 최소화하기 때문에 분석에 많이 활용되고 있다[24]. 최근에는 자연어 처리 기술을 이용한 주제, 키워드, 감성 추출 기능의 정확도가 크게 향상되어 교육[25], 의료[26] 등의 많은 분야에서 활용되고 있다. 자연어 처리를 통한 키워드 분석은 언어의 형태학적 분석을 기반으로 하며, 전체 콘텐츠의 컨텍스트를 분석하고 가장 관련성이 높은 키워드를 도출하기 때문에, 누락되거나 간과되는 키워드를 최소화하여 저자의 주장을 보다 정확하게 분석할 수 있다[27]. 실제로 자연어 처리 및 패턴 매칭 기술을 기존의 콘텐츠 분석 기술과 비교한 연구에서는 자연어 처리 기술을 사용할 때 정확도가 2배 높은 것으로 나타났다[28].

핵심 키워드 추출을 위해 본 연구에서는 인공지능 NLP 분석 툴인 IBM Watson NLU(Natural Language Understanding)을 사용하였다. 이 분석 툴은 대량의 비정형 텍스트 데이터로부터 딥러닝 알고리즘을 이용해 의미를 추출하는 Text Analytics 기능이 포함되어 있다. 대량의 문서에서 키워드를 추출하기 위해 기존의 NLP 분석에서 사용되었던 기법은 word-frequency 와 TF-IDF와 같은 빈도수 기반 방식이 있다. 이 기법은 간단하고 효과적인 분석 방법이지만 동의어, 유사어, 그리고 다른 단어와의 연관성을 인식하지 못하는 문제점이 있고, 불용어 처리 등의 데이터 전처리에 수작업의 노력과 시간을 필요로 하는 등 어려움이 있었다. Watson NLU 서비스는 입력받은 텍스트를 분석하여 핵심 키워드 20개를 추출하고 관련성(Relevance) 점수를 제시해주는 방식으로 복잡한 프로그래밍 없이 사용할 수 있다. 또한 컨텍스트 내의 키워드 간 의미 관계를 인식하여 동의어나 유의어를 반영해 키워드를 추출하기 때문에 빈도수 기반 접근방식보다 핵심 기술의 빠른 분석과 유효한 데이터 분석이 가능하다. 500개의 기술 키워드는 클렌징 작업을 거쳐 각 키워드별 빈도수를 산출하여 빈도수 3 이상인 18개의 기술 키워드를 정리하였고, 이중 특허 검색조건으로

사용된 ‘artificial intelligence’, ‘contact center’ 등의 상위레벨 키워드를 제외하고, ‘voice recognition’, ‘speech recognition’ 같이 의미가 유사한 키워드를 수작업을 통해 하나의 키워드 그룹으로 구성하여 핵심기술 키워드 그룹 8개를 얻었다(표 4). 본 8개의 키워드 그룹은 AI Contact Center 기술의 핵심기술 키워드로, 3.3에서 3단계 분석 시 검색 키워드로 다시 사용하여 기술 별 특허문서를 수집하고, 상세 분석을 수행하였다.

표 4. 핵심기술 키워드 추출 목록

Table 4. Core technology keywords extracted by NLP

Seq	Keywords from summary of the invention	Freq.	Group of core tech.
1	voice recognition	12	KEY1
2	artificial intelligence	8	-
3	speech recognition	8	KEY1
4	contact center	7	-
5	intelligent sound script	7	KEY2
6	automated agent	6	KEY3
7	present invention	6	-
8	call center	5	-
9	emotion analysis	5	KEY4
10	real time	5	-
11	conversation process design	4	KEY5
12	following steps	3	-
13	information retrieval	3	KEY6
14	Interactive Voice Response	3	KEY7
15	invention	3	-
16	quality inspection	3	KEY8
17	sentiment analysis	3	KEY4
18	Technical problem	3	-

3.3 분석 프로세스 및 결과

3단계는 2단계에서 도출된 핵심 기술 키워드 그룹을 활용하여 기술별 분석을 수행하였다(그림 5). PATSTAT 에 등록된 특허문서의 abstract 또는 title 에 2단계에서 도출된 핵심 기술 키워드가 포함되는 경우 핵심특허 후보로 추출하고, 이 특허문서들을 분석하였다.

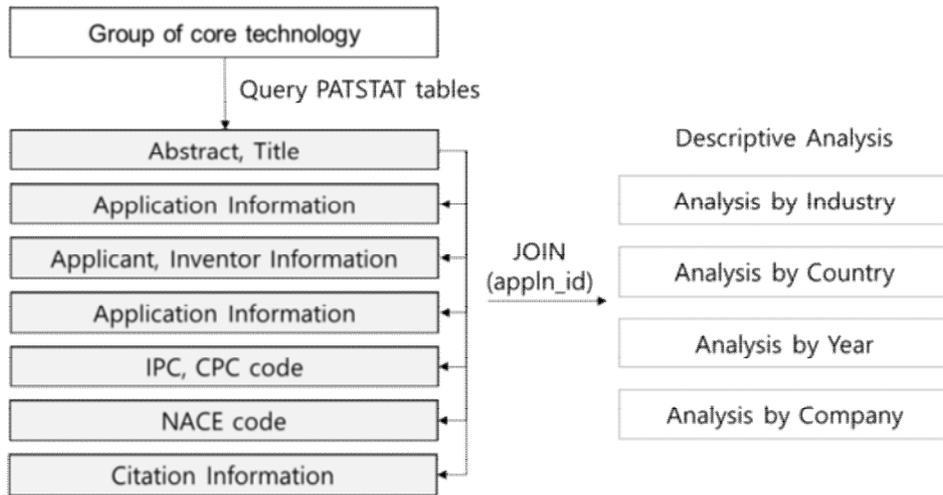


그림 5. 데이터 수집 절차 3단계
Fig. 5. Phase 3 of data collection process

표 5. 핵심기술 키워드 KEY1 관련 특허수집 SQL쿼리문

Table 5. SQL queries to collect patents related to core technology group KEY1

Matched keywords searching with BOOLEAN MODE	Removed the false positives using WHERE predicate
<pre>CREATE TABLE students.tbk1_1 select appln_id from t1s202_appln_title where match (APPLN_TITLE) against ('+artificial +intelligence +((+voice +recognition) (+speech +recognition))' IN BOOLEAN MODE) group by appln_id UNION select appln_id from t1s203_appln_abstr where match (APPLN_ABSTRACT) against ('+artificial +intelligence +((+voice +recognition) (+speech +recognition))' IN BOOLEAN MODE) group by appln_id ;</pre>	<pre>CREATE TABLE students.tbk1_4 select appln_id from students.tbk1_4 left outer join patstat2021s.t1s202_appln_title t202 using(appln_id) where ((t202.appln_title like '%voice recognition%') OR (t202.appln_title like '%speech recognition%')) UNION select appln_id from students.tbk1_4 left outer join patstat2021s.t1s203_appln_abstr t203 using(appln_id) where ((t203.appln_abstract like '%voice recognition%') OR (t203.appln_abstract like '%speech recognition%')) ;</pre>

3.3.1 핵심기술 키워드 그룹 KEY1 분석

핵심기술 키워드 그룹 KEY1 으로 선별된 ‘voice recognition’, ‘speech recognition’ 키워드가 포함된 특허를 대상으로 분석을 수행하였다. KEY1 관련 특허는 표 5와 같이 PATSTATS DB를 SQL 쿼리의 boolean mode 로 매칭 키워드 검색하여 857 건의 특허 application ID 를 수집할 수 있었고, 이 중 False Positive를 제거하기 위해 SQL 쿼리의 WHERE

절을 이용해 더 엄격한 룰로 매칭 키워드 검색을 수행하여 626건의 특허 application ID를 수집하였다. 626 건의 KEY1 특허의 family size는 0~82까지 분포를 나타내며, 17.9%는 family size가 0이고, 60.4%는 1을 갖는다. 특허문서 ‘KNOWLEDGE SYSTEM METHOD AND APPARATUS’는 family size가 82로 그룹 중에 가장 크게 나타나 핵심특허 후보로 판단할 수 있다(표 6).

표 6. 핵심기술 키워드 KEY1 특허 family size
Table 6. Family size of core technology keyword KEY1

Family size	Patents #	%
0	112	17.9
1	378	60.4
2	92	14.7
3~10	38	6.1
>10	6	0.9

KEY1 특허에 대해 연도별 특허 출원 관청을 보면, 1999년에 캐나다에서 1건의 첫 특허가 출원된 후 2017년에 들어서서 출원 건수가 증가하기 시작하고, 전체 626건의 특허 중 전체의 57%인 356건이 중국에서 출원된 특허로 중국이 KEY1에 핵심기술을 보유한 나라임을 확인할 수 있다(그림 6).

Number of patent applications by country, (%)

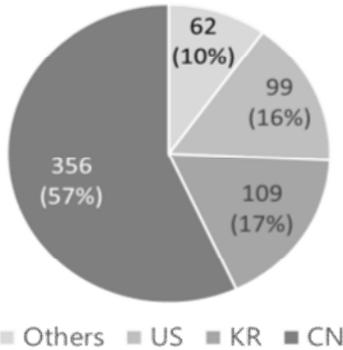


그림 6. 핵심기술 키워드 KEY1 특허의 관청 별 출원수
Fig. 6. Patents by filing office of group KEY1

중국은 전체 특허 출원 비중은 가장 높으나 출원인으로 Top5를 분석하면 가장 많은 특허는 LG ELECTRONICS(143건)로 나타나고, PING AN TECHNOLOGY COMPANY(62건), SAUMSUNG ELECTRONICS COMPANY(38건) 순으로 나타난다(표 7).

표 7. 핵심기술 키워드 KEY1 출원인
Table 7. Core technology group KEY1 applicant

Applicant	Application #
LG ELECTRONICS	143
PING AN TECHNOLOGY COMPANY	62
SAMSUNG ELECTRONICS COMPANY	38
TENCENT TECHNOLOGY (SHENZHEN) COMPANY	35
BAIDU ON-LINE NETWORK TECHNOLOGY (BEIJING) COMPANY	24

IPC class의 4자리 digit으로 분석해 보면(그림 7) G10L(음성분석 또는 합성)이 444건 (38%)로 가장 높은 분포를 나타낸다. IPC class별 설명은 표 8과 같다.

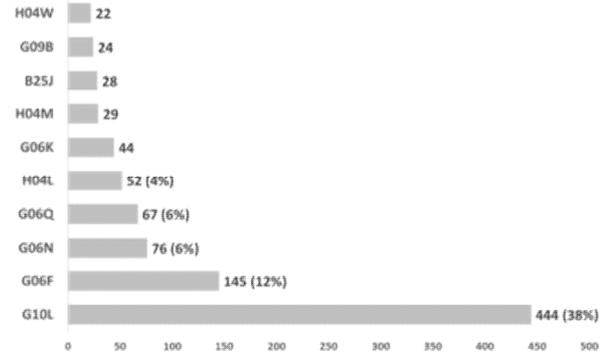


그림 7. IPC 4 digits class별 application 수
Fig. 7. Number of applications by IPC 4-digits class

표 8. IPC class 별 설명
Table 8. Description of IPC class

IPC	Description
G10L	speech analysis or synthesis; speech recognition; speech or voice processing; speech or audio coding or recording
G06F	Electric digital data processing (computer systems based on specific computational models G06N)
G06N	Computer systems based on specific computational models
G06Q	Data processing systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for
H04L	Transmission of digital information, e.g. Telegraphic communication

3.3.2 KEY1 그룹의 핵심특허 선별

family size와 citation 정보는 특허의 가치를 나타내는 지표로[33], 이들 지표의 총합이 높은 특허들은 KEY1 기술을 대표하는 핵심특허로 볼 수 있다. Family size 별, forward citation 별, backward citation 별, NPL citation 별 KEY1 (voice/speech recognition) 그룹의 Top 5 핵심특허 후보 목록을 추출하면 각 지표에 따라 차이가 있다.

이를 통합하여 핵심특허를 선별하기 위해서는 각 항목의 스케일의 차이를 반영한 지표가 필요하다. 본 연구에서는 식 (1)과 같이 min-max feature scaling을 통해 family size와 citation 값을 0에서 1사이의 동일한 스케일 값으로 정규화하였다.

Rescaling (min-max normalization)

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

정규화한 normalized family size(N. F. size), normalized backward citation(N. B. citation), normalized forward citation(N. F. citation), normalized NPL citation(N. NPL citation)의 4가지 항목 총합 (Sum N. score)이 높은 순으로 핵심 특허로 구분하였다. 이때 INPADOC_FAMILY_ID가 동일한 특허는 중복을 제거하여 표 9와 같이 non-family Top 10 핵심특허를 추출하였다.

IV. 연구 결과

새로운 기술인 AI 컨택센터에 관련된 IP R&D 계획수립을 위해 상위 개념 키워드인 ‘Artificial

Intelligence’, ‘Contact Center’, ‘Call Center’를 이용하여 단계1에서 유효특허를 추출하고, 이 유효특허에 대해서 인공지능 NLP로 자연어 텍스트 분석을 수행하여 단계2를 통해 특허 내의 기술 키워드를 수집하고 그룹화하여 핵심기술 그룹 8개를 도출했다. 단계3에서는 핵심기술 그룹 KEY1(‘voice recognition’, ‘speech recognition’)을 상세 분석하기 위해 title 과 abstract에서 KEY1 키워드를 포함하는 핵심특허 후보목록을 선별하고 family size와 citation 정보를 이용해 IP-R&D에 유효하게 활용할 수 있는 핵심특허 Top 10을 추출하였다. 이 방법론은 연구 개발 하고자 하는 기술의 대분류와 상세화된 소분류에 따라 필요한 핵심특허를 단계별로 선별할 수 있는 방안을 제시한다.

본 연구에서 사용된 AI 컨택센터 기술의 경우 단계1에서는 Application ID가 ‘421956761’인 ‘Intelligent automated agent and interactive voice response for a contact center’가 유효특허로 도출되었고(표 10), AI 컨택센터의 상세 기술 키워드 ‘voice recognition’과 ‘speech recognition’을 도출한 단계3에서는 Application ID ‘420202354’인 ‘Method and system for feature detection’이 유효한 핵심특허로 도출되었다 (표 11).

표 9. KEY1 특허의 정규화 된 family size와 citation 수에 따른 Top 10
Table 9. Top 10 patents of KEY1 by normalized family size and citations

Appl. ID	Fam. ID	Application title	N. F. size	N. B. citation	N. F. citation	N. NPL citation	Sum N. score
420202354	378083634	Method and system for feature detection	0.23	1.00	1.00	0.30	2.53
52564024	261748	Knowledge system method and apparatus	1.00	0.23	0.62	0.30	2.15
51759677	422893	Method for noise adaptation in automatic speech recognition using transformed matrices	0.14	0.05	0.03	1.00	1.22
502832593	485734668	Speech recognition method and device based on artificial intelligence	0.01	0.20	0.00	0.60	0.81
496315859	479600491	Method and device for extracting speech feature based on artificial intelligence	0.01	0.05	0.00	0.70	0.76
501927967	501927967	Artificial intelligence voice recognition apparatus and voice recognition	0.02	0.55	0.01	0.20	0.79
531712880	531712880	Artificial intelligence communication system and communication method	-	0.10	-	0.60	0.70
363987549	363987549	Automated personal assistance system	0.01	0.31	0.07	0.30	0.70
274268033	274268033	Language arts game	0.01	0.10	0.03	0.50	0.64
16332912	2734759	METHOD FOR TRANSFORMING LANGUAGE INTO A VISUAL FORM	0.07	0.02	-	0.50	0.59

표 10. 단계1의 AI contact center 핵심특허와 출원인
Table 10. Top patent and applicant of AI contact center in phase 1

Top of core patent		Top of applicant
Appl. ID	Patent title	
421956761	Intelligent automated agent and interactive voice response for a contact center	AVAYA

표 11. 단계3의 핵심기술 KEY1 핵심특허와 출원인
Table 11. Top patent and applicant of KEY1 of AI contact center in phase 3

Top of core patent		Top of applicant
Appl. ID	Patent title	
420202354	Method and system for feature detection	LG ELECTRONICS

V. 결 론

본 연구에서는 세 단계로 특허를 분석하여 R&D 프로세스의 시작 단계에서 신기술에 대한 시장 및 기술 리더를 식별하고, NLP 분석을 통해 향상된 특허 키워드 검색으로 핵심기술을 선별하는 방법론을 제시하였다. 이러한 특허 정보 분석은 R&D 계획수립과 기술 정의 단계 또는 아이디어 창출 단계에서 기술개발에 대한 전략을 빠르게 수립하고, IP 기반 R&D 프로세스를 구축하는데 기반이 된다. 이 방법론은 대량의 비정형데이터를 인공지능 기술을 활용해 분석하는 디지털 방법론으로 R&D 프로세스를 단축시키는데도 기여할 수 있다. 기술이 보호되고 있거나 최신의 혁신적인 기술인 경우, 기술 개발에 참조할 수 있는 핵심 IP 정보를 빠르게 식별하는 방안이 더욱 필요하다. 본 연구에서 제시된 방법론은 공식적으로 공개된 기술 빅데이터인 특허정보를 이용해 R&D 전략을 수립하는 방안을 제시한다.

이 연구 방법론은 특허 뿐 아니라 텍스트 형태의 문서에서 정보를 추출하는데도 다양하게 활용할 수 있다. 본 연구에서는 정보 분석의 대상을 특허문서로 한정하여 분석하였으나, 후속 연구에서는 관련된

특허 뿐 아니라 논문 등의 기술 문서들을 동일한 방법론으로 분석하여 보다 심도 있는 기술정보를 확보하고 정보의 다양성도 확충할 필요가 있다. 또한 연구개발 전체 프로세스에서 인공지능 및 빅데이터 분석을 활용하는 디지털 R&D 구축을 위한 추가 연구가 필요하다.

References

- [1] C. K. Park and Y. B. Seo, "A Study on the Technological Priorities of Manufacturing and Service Companies for Response to the 4th Industrial Revolution and Transformation into a Smart Company", *Journal of Convergence for Information Technology*, Vol. 11, No. 4, pp. 88-101, Apr. 2021. <https://doi.org/10.22156/CS4SMB.2021.11.04.083>.
- [2] A. Park, S. B. Lee, and J. Song, "Application of AI based Chatbot Technology in the Industry", *Journal of The Korea Society of Computer and Information*, Vol. 25, No. 7, pp. 17-25, Jul. 2020. <https://doi.org/10.9708/jksci.2020.25.07.017>.
- [3] R. G. Cooper and S. J. Edgett, "Stage-Gate and the Critical Success Factors for New Product Development", *BPTrends*, Vol. 7, pp. 1-6, Jul. 2006.
- [4] J. Fagerberg, "Innovation: a guide to the literature, in J. Fagerberg, et al. (Eds.)", *The Oxford Handbook of Innovation*. Oxford University Press, pp. 1-26, Oct. 2004. <http://dx.doi.org/10.1093/oxfordhb/9780199286805.003.0001>.
- [5] R. G. Cooper, "Managing Technology Development Projects", *IEEE Engineering Management Review*, Vol. 35, No. 1, pp. 67-76, 2007.
- [6] A. Gibbs, "Strategic use of patents in technological innovation", *Patent21*, Vol. 71, pp. 6-11, 2007.
- [7] World Intellectual Property Organization (WIPO), "Understanding Industrial Property", World Intellectual Property Organization, 2016. <https://doi.org/10.34667/tind.36288>.
- [8] B. H. Hall, "Patents, innovation, and development",

- International Review of Applied Economics, Feb. 2022. <https://doi.org/10.1080/02692171.2021.2022295>.
- [9] G. Kim and J. Bae, "A Novel Approach to Forecast Promising Technology through Patent Analysis", *Technological Forecasting & Social Change*, Vol. 117, pp. 228-237, Apr. 2017. <https://doi.org/10.1016/j.techfore.2016.11.023>.
- [10] S. J. Ahn, et al., "Analysis of patent development contents of climate technology reduction R&D projects using topic modeling", *The Journal of Intellectual Property*, Vol. 15, No. 3, pp. 293-332, 2020.
- [11] D. K. Nam, et al., "Technology Trend Analysis in the Automotive Semiconductor Industry using Topic Model and Patent Analysis", *Journal of Korea Technology Innovation Society*, Vol. 21, No. 3, pp. 1155-1178, Sep. 2018.
- [12] H. Chen, G. Zhang, D. Zhu, and J. Lu, "Topic-Based Technological Forecasting Based on Patent Data: A Case Study of Australian Patents from 2000 to 2014", *Technological Forecasting & Social Change*, Vol. 119, pp. 39-52, Jun. 2017. <https://doi.org/10.1016/j.techfore.2017.03.009>.
- [13] A. Layne-Farrar, "Defining software patents: a research field guide", SSRN, Feb. 2006
- [14] B. H. Hall and M. MacGarvie, "The private value of software patents", *Research Policy*, Vol. 39, No. 7, pp. 994-1009, Sep. 2010. <https://doi.org/10.1016/j.respol.2010.04.007>.
- [15] Z. Xie and K. Miyazaki, "Evaluating the effectiveness of keyword search strategy for patent identification", *World Patent Information*, Vol. 35, No. 1, pp. 20-30, Mar. 2013. <https://doi.org/10.1016/j.wpi.2012.10.005>.
- [16] D. H. McQueen and H. Olsson, "Growth of embedded software related patents", *Technovation*, Vol. 23, No. 6, pp. 533-544, Jun. 2003. [https://doi.org/10.1016/S0166-4972\(02\)00011-1](https://doi.org/10.1016/S0166-4972(02)00011-1).
- [17] B. Fernekees, "AI and the Contact Center: What You Need to Know", *CRM Magazine*, Vol. 24, pp. 19, Mar. 2020.
- [18] Z. Askin, M. Armony, and V. Mehrotra, "The Modern Call Center: A Multi-Disciplinary Perspective on Operations Management Research", *Production and Operations Management*, Vol. 16, No. 6, pp. 665-688, Nov. 2007. <https://doi.org/10.1111/j.1937-5956.2007.tb00288.x>.
- [19] K. Kirkpatrick, "AI in contact centers", *Communications of the ACM*, Vol. 60, No. 8, pp. 18-19, 2017. <http://dx.doi.org/10.1145/3105442>.
- [20] N. Gans, G. Koole, and A. Mandelbaum, "Telephone Call Centers: Tutorial, Review, and Research Prospects", *Manufacturing & Service Operations Management*, Vol. 5, No. 2, pp. 79-141, Apr. 2003. <https://doi.org/10.1287/msom.5.2.79.16071>.
- [21] S. G. Ann and H. C. Ahn, "A Study on User Switching Intention from Contact Center-oriented to AI Chatbot-Oriented Customer Services", *Journal of the Korea Society of Digital Industry and Information Management*, Vol. 19, No. 1, pp. 57-76, Mar. 2023.
- [22] J. P. Park, et al., "Analysis of change in productivity of call center due to introduction of AI system", *Poster of the Korean Society for Quality Management Conference*, Vol. 2020, pp. 144, 2020.
- [23] K.-D. Ryu, J.-P. Park, Y. Kim, D.-H. Lee, and W.-J. Kim, "Development of AI-based Real Time Agent Advisor System on Call Center - Focused on N Bank Call Center", *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 20, No. 2, pp. 750-762, Feb. 2019. <https://doi.org/10.5762/KAIS.2019.20.2.750>.
- [24] H. Baek and H. Lee, "Framework of Socio-Technology Analysis and Prescriptions for a Sustainable Society: Focusing on the Mobile Technology Case", *Technology in Society*, Vol. 65, pp. 101523, May 2021. <https://doi.org/10.1016/j.techsoc.2020.101523>.

- [25] J. Pareek and M. Jhaveri, "DLNEx: A Tool to Automatically Extract Desired Learning Nuggets from Various Learning Materials, in: X.-S. Yang, A.K. Nagar, A. Joshi (Eds.)", Smart Trends in Systems, Security and Sustainability, Springer Singapore, pp. 319-330, Dec. 2017. https://doi.org/10.1007/978-981-10-6916-1_30.
- [26] S. Vidya, "Cross Domain Sentiment Classification Using Natural Language Processing", International Journal of Scientific Research in Computer Science, Vol. 3, No. 3, pp. 348-353, Mar. 2018.
- [27] K. Krippendorff, "Content Analysis: an Introduction to its Methodology", Sage, 2004.
- [28] D. Harhoff, F. M. Scherer, and K. Vopel, "Citations, Family Size, Opposition and the Value of Patent Rights", Research Policy, Vol. 32, No. 8, pp. 1343-1363, Sep. 2003. [https://doi.org/10.1016/S0048-7333\(02\)00124-5](https://doi.org/10.1016/S0048-7333(02)00124-5).

관심분야 : Machine Learning, Probabilistic Graphical Models and Unsupervised Learning

저자소개

김 정 희 (JungHeui Kim)



1998년 : 연세대학교 수학과(학사)
2001년 : 연세대학교 수학교육학과
(석사)
2023년 : 한양대학교
기술경영대학원 박사수료
관심분야 : Deep Learning and AI

김 영 민 (Young-Min Kim)



1999년 : 한양대학교 산업공학과
(학사)
2001년 : 한양대학교 산업공학과
(석사)
2006년 : University Paris 6
Computer Science
(공학석사)

2010년 : University Paris 6 Computer Science
(공학박사)

2016년 ~ 현재 : 한양대학교 기술경영대학원,
산업융합학부 부교수