Journal of Next-generation Convergence Information Services Technology Vol.13, No.2, April (2024), pp.277-285
ISSN 2384-101X(p) 2672-1163(e) *http://dx.doi.org/10.29056/jncist.2024.04.12*

BERT를 활용한 기술 특허 분류 성능 평가

Classification of Technology Patents Using BERT

이우식[1]

Woosik Lee[1]

요 약

최근 RPA과 인공지능을 결합한 인텔리전트 오토메이션의 도래로 재무, 보험, 인적자원 관리, 회계, 제조, 공급망 관리, IT 관리, 고객 서비스 등 다양한 분야에 광범위한 영향을 미치고 있음에도 불구하고, 기술 특허 분석에 딥러닝 기반의 자연어 처리를 적용한 IPA 연구는 충분히 발전하지 못한 상황이다. 본 연구는 이산화탄소 포집·활용에 대한 특허 데이터, 자연어 전처리 기법 그리고 BERT와 BERT 파생 모형 기반의 기술 특허 분류 시스템을 설계하고, 정확도, 카파 상관계수 그리고 F1-점수를 비교·분석하였다. 주요 결과를 요약·정리하면 다음과 같다. 첫째, 다섯 가지 CCU 기술 분류에서 BERT 모형이 ELECTRA 모형보다 더 좋은 성능을 보였으며, 이는 BERT모형의 MLM 방식이 전체 문맥 정보를 더 효과적으로 이해할 수 있음을 시사한다. 둘째, 특허 요약 분류에서는 제1 청구항 분류보다 더 높은 성능을 나타냈는데, 이는 언어 모형들이 다양한 텍스트 유형으로 학습되고, 일반적인 언어 사용과 문맥을 기반으로 학습되기 때문에, 전체 특허 내용을 요약하는 특허 요약을 더 효과적으로 처리할 수 있다고 판단된다. 본 연구는 BERT와 BERT 파생 모형을 기술 특허 분류에 적용한 의미 있는 IPA 연구로 비즈니스 전략 수립과 기술경쟁력 강화에 중대한 영향을 미칠 수 있는 가능성을 제시한다.

핵심어 : 비즈니스 애널리틱스, 지능형 프로세스 자동화, 자연어 처리, 언어모델, 비즈니스 의사 결정

Abstract

Despite the widespread impact on various fields such as finance, insurance, human resource management, accounting, manufacturing, supply chain management, IT management, and customer service due to the advent of Intelligent Process Automation(IPA) combining RPA and artificial intelligence, research on IPA applying deep learning-based natural language processing to technology patent analysis has not sufficiently developed. This study designs a technology patent classification system based on patent data on carbon dioxide capture and utilization, natural language preprocessing techniques, and BERT and BERT-derived models, and compares and analyzes accuracy, kappa coefficient, and F1-score. The main results are summarized as follows: First, among the five CCU technology classifications, the BERT model showed better performance than the ELECTRA model, suggesting that the BERT model's MLM method can understand the overall context information more effectively. Second, in patent summary classification, it showed higher performance than the first claim classification, which is judged because the language models are trained with various types of texts and based on general language use and context, thus more effectively processing patent summaries that summarize the entire patent content. This study presents the potential to significantly impact business strategy development and enhance technological competitiveness as

1 College of Business Administration, Gyeongsang National University, Jinju, Korea [Professor]
e-mail: woosiklee@gnu.ac.kr

Received(February 19, 2024), Review Result(1st: March 8, 2024), Accepted(April 5, 2024), Published(April 30, 2024)

© 2024 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/4.0/.

meaningful IPA research applying BERT and BERT-derived models to technology patent classification.

Keyword : Business Analytics, Intelligent Process Automation, Natural Language Processing, Language Model, Business Decision-Making

1. 서론

빅데이터, 인공지능, 사물인터넷, 가상현실 등 4차 산업혁명 기술이 기업 경영에 널리 이용되고 있다. 하지만 최근 기업이 디지털 변환을 추진하고 경쟁력을 강화하는 중요한 수단으로 로봇 프로세스 자동화(Robotic Process Automation, RPA) 도입이 중요한 수단이 되고 있다. RPA는 재무, 보험, 인적자원 관리, 회계, 제조, 공급망 관리, IT 관리, 고객 서비스 등의 다양한 분야에서 고정된 규칙에 기반한 반복적이고 정형화된 작업을 자동화하는 기술이다. 예를 들어, IT 관리의 경우, 정기적인 시스템 유지보수, 보고서 생성, 데이터 백업 및 복구 작업 등 IT 관련 작업을 자동화하여 IT 부서의 효율성을 높일 수 있고, 재무 및 회계 분야에서는 송장 처리, 지급 처리, 재무 보고, 예산 편성과 같은 작업에 RPA를 적용하여 처리 시간을 단축하고 정확도를 향상시킬 수 있다. 세계적인 회계 컨설팅 기업인 프라이스워터하우스쿠퍼스(PwC)는 RPA를 도입하여 700만 시간의 수작업을 자동화하고, 이를 통해 컨설턴트의 역량을 향상시켰다 [1].

인공지능은 데이터 분석, 의사 결정 지원, 예측 모델링 등 복잡한 문제 해결에 초점을 맞춰 경영 관리와 의사 결정 과정의 복잡성을 줄이고 전략적 인사이트를 제공하는 데 중점을 두고 개발되었다. 이를 통해 기업은 시장 변화를 미리 예측하고, 리스크를 관리하며, 경쟁력을 강화할 수 있다. 반면, RPA는 사용자의 일상적인 반복 업무를 자동화함으로써 시간을 절약하고 오류를 줄이는 데 초점을 맞춰, 직원들이 더 창의적이고 전략적인 작업에 집중할 수 있는 환경을 조성하는 데 기여하도록 설계되었다. 이는 인간의 노동을 디지털 노동으로 전환함을 의미하며, RPA가 저부가가치의 반복적인 업무를 처리함으로써 유휴 인력이 고부가가치 업무 및 창의적인 비즈니스 가치 발굴에 집중할 수 있도록 하는 노동 유연성 측면에서 긍정적인 효과를 제공한다 [2].

최근 RPA과 인공지능을 결합한 인텔리전트 오토메이션(Intelligent Process Automation, IPA) 이 단순 반복 작업을 넘어서 복잡한 의사 결정과 예측, 자연어 처리 등을 포함한 기업 업무 효율성 향상의 핵심기술로 빠르게 부상하고 있으며, 기업은 주 52시간제 근무 방식 확산의 대안으로 IPA 도입을 검토하고 이를 통해 업무 시간 단축과 비용 절감을 꾀하고 있다. 또한, 이는 직원들의 일과 생활의 균형을 개선하고, 기업의 경쟁력을 강화하는 데에도 기여할 수 있다. 예를 들어, 기계학습과 인공지능 기반의 자연어 처리 기능을 통합한 IPA는 고객 질문에 자동으로 응답하고, 고객 서비스 프로세스를 자동화할 수 있다. 이를 통해 고객 만족도를 높이는 동시에 인적 자원의 부담을 줄일 수 있다. 더불어 기계학습과 인공지능을 통한 수요 예측 모델과 RPA의 결합으로, 제조업체는

재고 관리, 주문 처리, 공급망 관리 등을 자동화하고 최적화할 수 있다. 이는 비용 절감과 운영 효율성 증대로 이어질 수 있다. 이렇듯 RPA/IPA는 직원의 업무 생산성과 편의성을 향상시키며, 현장 업무에서의 활용 사례가 점차 증가함에 따라 그 활용이 가속화되고 있다. 생산성 향상과 새로운 가치 창출이 기업의 생존력과 경쟁력 확보에 점점 더 중요해지는 상황에서, RPA/IPA의 성공적 안착과 지속적 발전을 위한 연구가 필요하다. RPA/IPA 도입에 대한 공감대는 형성되었으나, 대부분의 연구는 도입 사례에 집중되어 있고, 아직 RPA/IPA에 대한 연구가 충분히 이루어지지 않고 있다.

과학기술과 관련, 특허는 핵심기술을 담고 있는 기술문서로서 기술경쟁력 유지와 미래 신성장 동력 구축에 있어 중요한 역할을 한다. 그러나 매년, 선진 5개 특허청을 중심으로 선진 특허 분류(Cooperative Patent Classification, CPC)를 위해 많은 예산과 인력이 투입되고 있다 [3]. 최근에는 인공지능을 활용한 자동화된 분류 모델의 연구 및 개발이 추진되고 있으나, 이러한 기술은 아직 실무 분야에서는 제한적으로 활용되고 있다. 특히 특허의 자동 분류는 국가의 기술 경쟁력 확보 차원에서도 매우 도전적이며 중요한 과제로 여겨지고 있다 [3]. 이에 따라, 이에 따라, 특허 RPA/IPA 솔루션의 양적·질적 성장을 촉진하기 위한 측면에서 축적된 특허 데이터를 기반으로 한 분석이 필요한 시점이다.

최근 국내에서 수행된 특허 RPA/IPA 관련 연구는 장지모 외 2인의 연구 [4] 와 박진우 외 4인의 연구 [3] 등이 있으나, 특허 분야에 대한 RPA 연구는 아직 충분하지 않은 상태다. 장지모 외 2인의 연구 [4]는 특허 분야에 최적화된 사전학습 언어모델인 ‘KorPatELECTRA’를 개발하고, NER(Named Entity Recognition), MRC(Machine Reading Comprehension) 그리고 문서 분류 등을 통해 특히 특허 분야에 대해 성능이 우수함을 보여주었다. 박진우 외 4인의 연구 [3]는 KorPatBERT와 CPC 코드별 데이터를 활용하여 CPC 서브클래스 분류를 가능케 하는 시스템을 개발하였다. 그러나 선행 연구들은 주로 한글 특허 언어모형 개발과 CPC 코드별 데이터에 국한되어 있다. 이에 본 연구는 BERT 모형과 BERT 파생 모형 그리고 ‘제1 청구항’과 ‘요약’ 데이터를 기반으로 한 기술 특허 분류를 통해 선행연구와의 차별성을 강조하고자 한다.

본 논문은 다음과 같이 4개 장(章)으로 구성되어 있다. 제1장 서론에서는 이 연구를 수행하게 된 배경과 필요성에 대해 설명한다. 제2장에서는 본 연구의 핵심인 딥러닝 기반의 자연어 처리에 대한 이론적 근거를 검토하고 요약하여 제시한다. 제3장에서는 실증 연구를 위한 방법론을 정립하고, 실제 분석을 통해 얻은 결과를 밝힌다. 마지막으로, 제4장에서는 연구 결과의 의의와 시사점을 도출함으로써 연구의 결론을 마무리한다.

2. 이론적 배경

2.1 딥러닝 기반의 자연어 처리

본 연구에서는 텍스트 마이닝을 활용한 문서 분류 방법을 살펴본다. 텍스트 마이닝은 자연어로 이루어진 데이터에서 숨겨진 구조나 연관성을 도출하여 유용한 지식을 발굴하는 방법이며, 이 과정에서 자료의 핵심 구성요소를 도출하고 벡터 형태로 나타낸다. 문서 요약 방식은 주로 추출요약(Extraction Summary)과 생성요약(Abstraction Summary)으로 구분된다. 추출요약은 원본 문장들의 중요성을 기준으로 내용을 선택하는 반면, 생성요약은 문서의 의미를 파악하고 새로운 요약문을 작성한다. 추출요약은 구문적 특징을 추출하는 데의 어려움과 새로운 문서에 적용할 때의 제약이 있으며, 생성요약은 시퀀스-투-시퀀스(Sequence-to-Sequence) 모형을 활용하지만 장거리 의존성 문제에 직면한다. 이러한 문제를 해결하기 위해 어텐션 메커니즘(Attention Mechanism)을 적용한 트랜스포머(Transformer) 모형이 사용되고 있으며, BERT(Bidirectional Encoder Representations from Transformers)는 이러한 트랜스포머의 인코더(Encoder)를 사용하여 양방향 문맥정보를 효과적으로 이해하는 데 강점을 가진다. 본 연구에서는 BERT 모형을 사용하였다.

2.1.1 BERT 모형

BERT 모형은 2018년 구글에 의해 소개된 언어 구조로, 사전 학습 단계에서 MLM(Masked Language Model)과 이어지는 문장 추론(Next Sentence Prediction, NSP)을 통해 문맥 속 뜻을 이해한다. MLM은 문장의 일부 어휘를 '가림' 처리한 다음, 이 가려진 어휘를 양쪽 방향에서 학습하여 추론하는 방법이며, NSP는 두 문장이 실질적으로 이어지는지를 학습한다. BERT의 주요 구동 원리는 Dot-Product Attention이고, 그 연산 과정은 식(1)과 동일하다 [5].

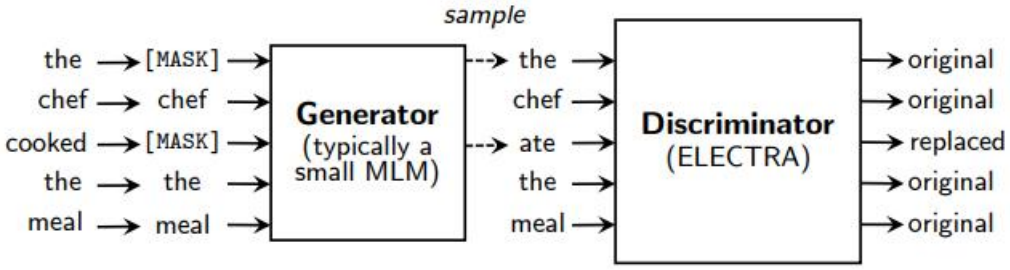
$$\text{Scaled Dot - Product Attention}(Q,\ K,\ V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

Dot-Product Attention은 입력 행렬 X와 가중치 행렬인 W_i^Q, W_i^K, W_i^V을 각각 곱하여, 단어 벡터를 행으로 하는 Q, K, V 행렬을 생성한다 [5]. Q와 K 사이에 관련성을 파악하기 위해 내적을 수행한 후, 모형의 안정성을 높이기 위해 $\sqrt{d_k}$(k는 키벡터의 차원 수) 로 나누고, 소프트맥스(SoftMax)함수를 적용하여 각 단어의 가중치(확률)를 계산한다. 특정 Q와 K가 문맥상에서 중요한 연관성을 지낼 경우, 내적 값을 증가시켜 해당 연관성에 더 많은 비중 V(Attention Score)를 두도록 학습된다. 가중치 행렬 W_i^Q, W_i^K, W_i^V는 학습 과정을 통해 최적화되어, 정확한 Q, K, V 행렬

을 생성한다 [5]. 그 다음 단계의 파인튜닝에서는 이미 의미론적이며 구문론적 연계성이 습득된 BERT를 활용하여, 목표에 부합하도록 BERT 모형을 조정한다 [5]. 이 방법은 대규모 말뭉치로 비지도학습을 통해 사전 학습된 언어 표현을 담은 모형을 사용하기 때문에, 제한된 자원과 소규모 데이터로도 다양한 도메인과 작업에 적용할 수 있으며, 학습 시간을 단축시킬 수 있는 장점이 있다 [5]. 그러나 BERT 모형은 입력 시퀀스에서 약 15%의 토큰을 무작위로 마스킹하고, 이를 복원하는 '마스크 언어 모델링(Masked Language Modeling, MLM)' 작업을 통해 학습한다. 이 학습 과정에서 전체 토큰 중 마스킹된 15%에 대해서만 손실이 계산되는데 이 방식은 비효율적일 수 있다. 또한, 학습 시에 사용되었던 [MASK] 토큰은 실제 상황의 데이터를 처리 추론 과정에서는 나타나지 않는데, 이는 모델의 한계점으로 작용할 수 있다 [6].

2.1.2 ELECTRA 모형

BERT 모형의 MLM 작업의 한계를 극복하기 위해, [그림 1]에서 보듯이 사전 학습 작업인 RTD(Replaced Token Detection)를 활용하는 ELECTRA 모형이 제안되었다 [7]. RTD는 생성기(Generator)를 사용하여 입력 시퀀스의 일부 토큰을 대체 가능한 토큰으로 교체한 후, 분류기(Discriminator)가 해당 토큰이 진짜인지 또는 생성기에 의해 생성된 가짜 토큰인지를 이진 분류하는 과정을 통해 학습을 진행한다. 이러한 방식은 BERT이 비해 더 효율적이고 효과적이라 할 수 있다. [6].



[그림 1] RTD의 구조

[Fig. 1] A Structure of RTD[7]

2.2 평가지표

본 연구에서는 분류 모형의 성능 척도로 정확도(Accuracy), 카파 계수(Kappa Coefficient), F1-값(Score)를 고려하였다 [8]. 정확도는 모든 측정값이 모형의 추정치와 얼마나 동일한지를 나타내는 비율이며, 카파 계수는 실측값과 예상치의 일치 수준을 평가하는 데 활용된다. 카파 상관계수의 등급은 [표 1]에서 보듯이 Landis와 Koch가 제시한 해석을 따른다 [9]. F1-값은 재현율과 정밀도의 조

화 평균을 사용하여 분류 모형의 성능을 측정하는 척도이다.

[표 1] 코헨의 카파 통계량을 통한 해석
[Table 1] Interpretation of Cohen's Kappa Statistic

코헨의 카파 통계량	해석
< 0	Poor
0.0 ~ 0.2	Slight
0.2 ~ 0.4	Fair
0.4 ~ 0.6	Moderate
0.6 ~ 0.8	Substantial
0.8 ~ 1.0	Almost Perfect

3. 실증분석

3.1 자료의 구성

실증 분석에 활용된 자료는 이산화탄소 포집·활용(Carbon Capture Utilization, CCU) 기술을 화학적, 전기화학적, 광화학/광전기 화학적, 생물학적 그리고 광물탄산화 방식으로 다섯 가지 주요 범주로 분류하였다 [8]. 그리고 특허 검색 프로그램인 윕스 온(WIPS ON)을 활용하여 CCU 기술과 관련된 선진 특허 분류(Cooperative Patent Classification, CPC)와 '발명의 명칭'에 '이산화탄소', 'Carbon Dioxide' 등 이산화탄소의 동의어 및 유의어를 포함하는 특허로 범위를 한정하여 총 1,190건의 특허를 수집하였다 [8]. 이후 노이즈 데이터를 제거한 뒤, 최종적으로 대한민국에서 등록된 특허 329건의 '제1 청구항'과 '요약'을 수집하였다 [8].

[표 2] 기술 분류 요약 예시
[Table 2] Example of Technical Classification Summary

기술 분류	요약
화학적 전환	본 발명은 탄소 지지체 및 상기 탄소 지지체에 담지된 나노금속 촉매를 포함하는 CO2의 메탄화 반응용 촉매로서, 상기 탄소 ...
전기화학적 전환	본 발명은 전기화학적 이산화탄소 환원 장치용 유동 플레이트에 관한 것이다. 상기 유동 플레이트는 일정 형상의 몸체부 ...
광화학/광전기 화학적 전환	【요약】 【과제】 유기 용매 중 및 수계 용매 중에서 낮은 바이어스에 있어서 전기화학적 또는 광촉매적인 이산화탄소 환원...
생물학적 전환	본 발명은 이산화탄소 포집 및 전환용 생물반응기에 관한 것으로, 보다 상세하게는 미생물을 이용한 생물학적인 발효공정과...
광물탄산화 전환	감마-이칼슘실리케이트 조성물을 포함하는 이산화탄소...

3.2 모형의 추정 및 분석

분석을 위해 BERT와 ELECTRA 모형에 각각 전체 데이터의 20%를 검증 세트로 할당하고, 레이블 분포를 균일하게 유지했다. 훈련 세트와 검증 세트의 텍스트를 토큰화하는 과정에서 한글 BERT 모형에는 'kcbert-base'와 한글 ELECTRA 모형에는 'koelectra-base-discriminator'를 사용했다. 이때 모든 텍스트 문장이 토큰의 최대 개수인 256에 도달하도록 패딩 처리를 진행했다. 모형의 미세조정(Fine-Tuning)시, 설정한 주요 하이퍼 파라미터값의 범위는 다음과 같다. 훈련 반복 횟수(Epoch) 는 최대 5회까지로 하고, 검증 데이터셋에 가장 높은 성능을 보인 시점의 모델을 평가 데이터셋에 적용하였다. 배치 사이즈(Batch Size)는 3으로 하였고 최적화 함수는 AdamW을 사용하였다. 학습률(Learning rate)은 1e-5, 엡실론(Epsilon)을 1e-8로 설정하고, 선형 스케줄러를 통해 점진적으로 조절하였다.

마지막으로 [표 3]에서는 BERT와 ELECTRA 모델의 성능을 정확도, 카파 상관계수, F1-점수를 기준으로 비교할 수 있다. 특허 요약과 제1 청구항 부문에서, BERT 모델이 더 높은 정확도를 보여주었다. 이 결과는 BERT의 MLM 방식이 전체 문맥의 깊은 이해를 통해 우수한 학습 성능을 제공함을 확인할 수 있었다. 또한, 특허 요약 분류에서 제1 청구항 분류보다 더 높은 성능을 보인 것은, 언어 모델들이 다양한 종류의 텍스트로 학습되며, 일반적인 언어 사용과 문맥을 기반으로 하는 학습이 이루어지기 때문에, 전체 특허 내용을 요약하는 특허 요약을 더 효과적으로 처리할 수 있다고 판단된다.

[표 3] 분류 모형의 성능 평가
[Table 3] Performance of Classification Models

요약	Accuracy	Kappa	F1-Score
BERT	0.8181	0.7624	0.8027
KoELECTRA	0.5974	0.4572	0.5140

제1 청구항	Accuracy	Kappa	F1-Score
BERT	0.7532	0.6786	0.7465
KoELECTRA	0.5844	0.4401	0.5123

4. 결론

최근 RPA가 재무, 보험, 인적자원 관리, 회계, 제조, 공급망 관리, IT 관리, 고객 서비스 등의 다양한 분야에서 큰 영향을 미치고 있음에도 불구하고, 인공지능을 활용한 IPA연구는 아직 충분히 발전하지 못한 상황이다. 본 연구는 CCU(탄소 포집 및 활용) 기술 특허 데이터, 한국어 문장 처리

를 위한 한글 사전학습 언어 모형, 그리고 BERT 모형 기반의 문서 분류 시스템을 설계하고, 이를 정확도, 카파 상관계수, F1-점수로 비교·분석하였다. 주요 결과를 요약하면 다음과 같다: 첫째, 다섯 가지 CCU 기술 분류에서 BERT 모형이 ELECTRA 모형보다 더 좋은 성능을 보였으며, 이는 BERT 모형의 MLM 방식이 전체 문맥 정보를 더 효과적으로 이해할 수 있음을 시사한다. 둘째, 특허 요약 분류에서는 제1 청구항 분류보다 더 높은 성능을 나타냈는데, 이는 언어 모형들이 다양한 텍스트 유형으로 학습되고, 일반적인 언어 사용과 문맥을 기반으로 학습되기 때문에, 전체 특허 내용을 요약하는 특허 요약을 더 효과적으로 처리할 수 있다고 판단된다.

본 연구는 BERT와 BERT 파생 모형과 '제1 청구항'과 '요약' 데이터를 기반으로 기술 특허 분류에 적용한 의미 있는 연구이다. 비즈니스 측면에서 볼 때, 특허 분류는 경쟁 분석 강화, 시장 기회 발견, 기술 동향 이해, 리스크 관리, 기술 협력 및 라이선싱 기회 찾기, 그리고 R&D 투자의 효율성 증대 등 다양한 방면에서 기업에 도움이 될 수 있다. 따라서, IPA의 발전은 이러한 분류 과정을 자동화하고 최적화하는 데 크게 기여할 수 있으며, 이는 비즈니스 전략 수립과 경쟁력 강화에 중대한 영향을 미칠 수 있는 가능성을 제시한다. 그러나 본 연구가 중요한 결과와 의미를 제시하고 있음에도 불구하고, 여전히 앞으로의 연구에서 다루어야 할 보완점이 몇 가지 존재한다. 본 연구에서 활용된 특허 데이터는 요약과 제1 청구항으로 구성되어 있으나, 특허 분석의 질적 향상을 위해서는 선행기술, 도면, 그리고 국제 특허 분류(International Patent Classification)를 포함하는 다양한 데이터를 기반으로 한 분석과 예측이 필요하다. 또한, 한글 토큰화 및 한글 사전 학습 모형의 토큰화가 문장 임베딩에 미치는 영향에 대한 연구 분석이 필요하다.

References

[1] J. K. Lee, D. H. Min, J. K. Yoo, "Perception Gaps between Managers and Workers in SMEs about the Effect of RPA(Robotic Process Automation) : A Co-orientation Analysis", The e-Business Studies, vol. 23, no. 5, October 2022, pp. 85-103, doi: 10.20462/tebs.2022.10.23.5.85.

[2] T. Kim, J. Lee. "A Study on the Impact and Response of RPA Adoption to the Customs Broker Industry in the Future", The Journal of Korea Research Society for Customs, vol. 21, no. 3, March 2020, pp. 43-66.

[3] J. Park, W. Sim, S. Lee, B. Ko, H. Noh, "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, vol. 17, no. 3, September 2022, pp. 209-256, doi: 10.34122/jip.2022.17.3.209.

[4] J. Jang, J. Min, H. Noh, "KorPatELECTRA : A Pre-trained Language Model for Korean Patent Literature to improve performance in the field of natural language processing(Korean Patent ELECTRA)", Journal of the Korea Society of Computer and Information, vol. 27, no. 2, February 2022, pp. 15-23, doi: 10.9708/jksci.2022.27.02.015.

[5] J. Shin, J. Shin, J. E. Jo, Y. Yoon, J. Jung, "Comparison of BERT-based Model Performance in CBCA

Criteria Classification", Journal of KIISE, vol. 49, no. 9, September 2022, pp. 727-734, doi: 10.5626/JOK.2022.49.9.727.

[6] J. Oh, H. Oh, "Detects depression-related emotions in user input sentences", Journal of the Korea Institute of Information and Communication Engineering, vol. 26, no. 12, December 2022, pp. 1759-1768.

[7] K. Clark, M Luong, Q. V. Le, C. D. Manning, "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generatorsm", The International Conference on Learning Representations, April 26-30, 2020, pp. 1-18, doi: 10.48550/arXiv.2003.10555.

[8] W. Lee, Y. J. Lee, "Classification of Technology Patents Using Natural Language Processing and Machine Learning Models", Journal of Next-generation Convergence Information Services Technology, vol. 13, no. 1, February 2024, pp. 93-102, doi: 10.29056/jncist.2024.02.09.