

식품 개체명 인식을 위한 트랜스포머 기반 모델과 대규모 언어 모델의 앙상블 기법

임소희*, 김건우**

Ensemble Approach of Transformer-based Models and Large Language Models for Food Named Entity Recognition

So-Hee Lim*, Gun-Woo Kim**

본 논문은 2022년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 (NRF-2021R1G1A1006381) 및 2025년도 산업통상자원부 및 한국산업기술기회평가원(KEIT)의 연구비 지원(RS-2025-02633048)에 의한 연구결과임

요약

식품 도메인에서의 개체명 인식(NER)은 정확한 정보 추출을 위해 필수적이며, 희소 개체명의 낮은 출현 빈도는 학습에 어려움을 준다. 본 연구는 트랜스포머 기반 모델(BERT, ELECTRA)과 대규모 언어 모델(LLM, GPT-3.5-turbo, LLaMA 2 7B)의 성능을 비교하고 최적 조합을 탐색하였다. ELECTRA+CRF+bi-LSTM은 희소 개체명에서, LLaMA 2 7B는 문맥 보완 측면에서 강점을 보였다. ELECTRA+CRF는 희소 개체명에서 F1-score 0.89를 달성했으며, LLaMA 2 7B와의 가중치 결합(Weighted Voting)으로 0.91까지 향상되었다. 일반 개체명에서는 0.93의 F1-score를 기록했다. 실험 결과, 트랜스포머 모델과 LLM의 조합이 식품 도메인 개체명 인식 성능 향상에 효과적임을 입증하였다.

Abstract

Named Entity Recognition (NER) in the food domain is essential for accurate information extraction, but the low frequency of sparse entities makes training challenging. This study compares the performance of transformer-based models (BERT, ELECTRA) and large language models (LLMs, GPT-3.5-turbo, LLaMA 2 7B) to find the optimal combination. ELECTRA combined with CRF and bi-LSTM showed strong performance for sparse entities, while LLaMA 2 7B was effective in providing contextual support. ELECTRA+CRF achieved an F1-score of 0.89 for sparse entities, which improved to 0.91 when combined with LLaMA 2 7B using weighted voting. For general entities, the F1-score reached 0.93. The results demonstrate that combining transformer-based models with LLMs effectively enhances NER performance in the food domain.

Keywords

named entity recognition, NER, transformer, large language model, LLM, ensemble

* 경상국립대학교 컴퓨터공학과 학사과정
- ORCID: <https://orcid.org/0009-0003-1095-8351>
** 경상국립대학교 컴퓨터공학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-5643-4797>

• Received: May 14, 2025, Revised: Jun. 27, 2025, Accepted: Jun. 30, 2025
• Corresponding Author: Gun-Woo Kim
Dept. of Computer Science and Engineering, College of IT Engineering,
Gyeongsang National University, Jinju, Korea
Tel.: +82-55-772-3323, Email: gunwoo.kim@gnu.ac.kr

1. 서 론

현대 사회에서 건강에 대한 관심이 점차 높아지면서, 식품은 인간의 건강과 밀접하게 연결된 중요한 요인으로 주목받고 있다[1]. 식품은 단순한 에너지 공급원을 넘어 생존의 물질적 기초이자 건강 유지에 필수적인 역할을 하며, 개인의 유전자, 생활 방식, 질병 상태 등 다양한 요인과 상호작용하여 같은 식품이라도 사람마다 상반된 영향을 미칠 수 있다[2][3]. 예를 들어, 특정 식품은 약물 복용이나 특정 질병과의 상호작용으로 인해 부작용을 유발하거나, 개인의 면역 체계에 영향을 미쳐 알레르기나 약물 효과 변화와 같은 예기치 못한 결과를 초래할 수 있다[1].

따라서 식품이 개인 건강에 미치는 영향을 정확히 파악하기 위해서는 식품에 대한 정량적이고 구조적인 정보를 포괄적으로 분석할 수 있는 기술이 필요하다. 이를 위해 최근에는 레시피, 건강 기록, 식품 라벨 등 비정형 텍스트 데이터를 정제하여, 식품명 외에도 단위(Unit), 양(Quantity), 부위(Part)와 같은 세부 속성 정보를 함께 추출하려는 시도가 이어지고 있다. 이러한 정보는 조리 과정의 정량화, 영양소 계산, 개인 맞춤형 식단 설계 등 다양한 실용적 응용에 핵심적인 역할을 할 수 있다.

D. Choi et al.[4]은 레시피 문맥에서 재료의 양과 단위를 예측하기 위한 KitchenScale 모델을 제안하여, 측정 유형 분류 및 단위 추정, 양 회귀 등의 세부 태스크를 수행하고 조리법의 정량화에 기여하였다. S. Batra et al.[5]은 RecipeDB를 구축하여 재료명, 단위, 양, 상태 등 식재료의 다양한 속성을 체계적으로 구조화함으로써 정밀한 정보 추출을 위한 기반을 마련하였다. B. Popovski[6]는 FoodIE라는 규칙 기반 개체명 인식 시스템을 통해 식품명뿐만 아니라 부위와 같은 속성 정보를 함께 추출하였고, F. Mohbat and M. J. Zaki[7]는 KERL 시스템을 통해 식품 지식 그래프와 대규모 언어 모델을 결합함으로써 부위별 미세 영양 정보를 활용한 개인화된 식단 추천을 시도하였다. 이처럼 단위, 양, 부위 등의 정량적 속성에 대한 정교한 추출은 식품 데이터 분석의 정밀도와 활용 가능성을 높이는 데 핵심적인

요소가 된다.

이러한 작업을 가능하게 하는 핵심 기술은 개체명 인식(NER, Named Entity Recognition)이며, 이는 텍스트 내에서 특정 개념이나 실체를 인식하고 분류하는 자연어 처리(NLP, Natural Language Processing) 기술로, 정보 추출(IE, Information Extraction)의 핵심 구성 요소로 활용된다[8][9]. 최근에는 트랜스포머(Transformer) 구조를 기반으로 하는 BERT[9], ELECTRA 등의 사전 학습 모델들이 다양한 도메인에서 NER 성능을 크게 향상시켰으며, GPT, LLaMA와 같은 대규모 언어 모델(LLM, Large Language Model)은 도메인 특화 학습 없이도 문맥 기반 인식에 뛰어난 성능을 보이고 있다[10].

그러나 식품 도메인에서는 일반적인 개체명 외에도 단위, 부위와 같은 특수 개체명이 데이터 내에 희소하게 등장하는 경우가 많아, 기존 NER 모델들이 해당 라벨에 대해 낮은 인식 성능을 보이는 문제가 발생한다. 또한 식품 텍스트는 그 자체로 도메인 특유의 어휘와 맥락 구조를 가지므로, 일반적인 NER 태스크와는 차별화된 접근이 요구된다.

본 연구에서는 이러한 문제 해결을 위해 식품 도메인에 특화된 레이블 정보를 포함하는 식품 데이터셋인 TASTEset[11]을 활용하였다. 해당 데이터셋은 식품, 맛, 색, 양 등 다양한 속성 태그를 포함하고 있으며, 본 연구에서는 이 중 단위(Unit), 양(Quantity), 부위(Part)와 같은 정량적·구조적 정보를 중심으로 필수 레이블을 선별하였다. 이후 이를 BIO(Begin, Inside, Outside) 태깅 방식으로 재정의하고, 보다 정밀한 식품 개체명 인식 태스크로 재구성하였다.

그러나 식품 도메인에서는 일반적인 개체명 외에도 단위, 부위와 같은 특수 개체명이 데이터 내에 희소하게 등장하는 경우가 많아, 기존 NER 모델들이 해당 라벨에 대해 낮은 인식 성능을 보이는 문제가 발생한다. 또한 식품 텍스트는 그 자체로 도메인 특유의 어휘와 맥락 구조를 가지므로, 일반적인 NER 태스크와는 차별화된 접근이 요구된다.

이에 본 연구는 다음 세 가지 주여 기여를 중심으로 식품 도메인 특화 NER 모델을 설계하였다. 본 연구의 주요 기여는 다음과 같다.

첫째, 식품 도메인에서 출현 빈도가 낮은 단위, 부위 등의 희소 개체명에 대해 데이터 재구성과 모델 구조 개선을 통해 인식 성능을 향상시키는 전략을 제안하였다. 둘째, 트랜스포머 기반 모델(BERT, ELECTRA)의 도메인 적합성과 GPT 및 LLaMA 2 7B와 같은 대규모 언어 모델(LLM)의 문맥 일반화 능력을 결합하여, 상호 보완적인 앙상블 모델을 구현하였다. 셋째, 식품 도메인에 특화된 TASTeSet을 기반으로 단위(Unit), 양(Quantity), 부위(Part) 중심의 BIO 태깅 체계를 재정의하고, 이를 활용한 정밀 개체명 인식 태스크를 구성하였다.

이를 통해, 본 연구는 기존의 단순한 식품명 추출을 넘어서 구조화된 식품 정보를 정밀하게 인식할 수 있는 기반을 마련하였으며, 조리법 정량화, 질환 맞춤형 식단 설계, 식품 안전 모니터링, 식품-질환 연계 분석 등 다양한 실제 응용 분야로의 확장 가능성을 제시한다.

II. 배경 지식

2.1 개체명 인식(NER)

개체명 인식(NER)은 자연어 처리(NLP)에서 핵심적인 과제로, 텍스트에서 특정 개체를 식별하고 분류하는 작업이다. NER은 정보 추출, 질의 응답, 기계 번역 등 다양한 NLP 응용 분야에서 중요한 역할을 하며[7][11], 문장에서 인물, 장소, 조직, 도메인 특화 개체(예: 식품명, 단위 등)를 정확히 인식하는 데 사용된다. 전통적으로 NER은 규칙 기반 접근 방식과 통계적 기법을 활용했으나, 최근에는 딥러닝 모델을 적용함으로써 성능이 크게 향상되었다[12]. 특히, 트랜스포머 기반 모델이 등장하면서 문맥 정보를 효과적으로 반영할 수 있게 되었고, 대규모 언어 모델(LLM)과의 결합을 통해 더욱 정교한 개체 인식이 가능해졌다[13][14].

2.2 BIO 태깅

BIO 태깅은 NER에서 널리 사용되는 엔티티 경계 표시 방법으로, 각 토큰을 B(Begin, 개체 시작),

I(Inside, 개체 내부), O(Outside, 개체가 아닌 단어)로 태깅하는 방식이다[13][15]. 이 방법은 개체의 시작과 끝을 명확히 구분할 수 있도록 하여 다중 토큰으로 구성된 개체(예: "올리브 오일")도 효과적으로 처리할 수 있게 한다[14]. BIO 태깅은 모델 학습 과정에서 여러 가지 이점을 제공한다. 첫째, 개체 경계를 명확하게 정의하여 모델이 엔티티 구조를 더욱 정밀하게 학습할 수 있도록 돕는다. 둘째, 연속된 개체들을 효과적으로 구별할 수 있어, 문장에서 서로 다른 개체가 혼동되는 문제를 줄인다. 셋째, BIO 태그를 추가적인 입력 특성으로 활용하면 모델의 표현력을 향상시킬 수 있으며, 실제로 BIO 태그 임베딩을 적용하면 모델 성능이 개선된다는 연구 결과가 보고되었다[12].

2.3 BERT

BERT(Bidirectional Encoder Representations from Transformers)는 2018년 Google AI 연구팀이 발표한 트랜스포머 기반 언어 모델로, 양방향 문맥을 고려하여 단어의 의미를 더욱 정교하게 이해할 수 있도록 설계되었다[7][16]. BERT는 대규모 코퍼스를 사용하여 사전 학습되며, 마스킹된 단어 예측(Masked language model)과 다음 문장 예측(Next sentence prediction) 태스크를 통해 문맥을 깊이 있게 학습한다[12][16].

NER 태스크에서 BERT의 성능을 향상시키기 위해 다양한 기법이 제안되었다. 단어 임베딩에 품사 태그 임베딩, 문자 수준 임베딩 등을 결합하여 입력 표현을 풍부하게 만들 수 있으며, 특정 도메인의 텍스트로 추가적인 사전 학습을 수행하면 도메인 특화 지식을 획득할 수 있다. 또한, CRF(Conditional Random Fields) 층을 결합하면 출력 레이블 간의 의존성을 모델링할 수 있으며, bi-LSTM을 추가하면 장거리 문맥 정보를 효과적으로 반영할 수 있다[12].

2.4 ELECTRA

ELECTRA(Efficiently Learning an Encoder that Classifies Token Replacements Accurately)는 BERT의

성능을 유지하면서도 훈련 효율성을 크게 개선한 모델이다. BERT가 마스킹된 토큰을 예측하는 방식과 달리, ELECTRA는 생성자(Generator)와 판별자(Discriminator) 네트워크를 활용하여 대체된 토큰을 식별하는 방식으로 학습한다[17].

이러한 학습 방식의 차이로 인해 ELECTRA는 모든 토큰에 대해 학습 신호를 제공하므로, BERT보다 학습 효율성이 높고 빠르게 최적의 성능에 도달할 수 있다. 결과적으로 NER 태스크에서도 ELECTRA는 상대적으로 적은 계산 자원으로도 BERT와 유사하거나 더 나은 성능을 보인다[17].

2.5 GPT

GPT(Generative Pre-trained Transformer) 시리즈는 OpenAI에서 개발한 대규모 언어 모델로, 텍스트 생성 능력이 뛰어나다. GPT-3.5-turbo는 GPT-3의 개선된 버전으로, 다양한 NLP 작업에서 높은 성능을 보인다[18]. 특히, NER 작업에서는 fine-tuning 없이도 few-shot 학습을 통해 우수한 성능을 보인다. 예를 들어, 몇 개의 예제 문장을 제공하면 모델이 해당 패턴을 학습하여 새로운 입력에서도 적절한 개체를 예측할 수 있다[19]. 이러한 특성을 활용하면 식품 도메인에서도 제한된 학습 데이터로 효과적인 개체명 인식이 가능하다[19].

2.6 LLaMA

LLaMA는 Meta AI에서 개발한 오픈소스 대규모 언어 모델로, 상대적으로 적은 파라미터로도 높은 성능을 달성할 수 있도록 최적화되었다. 특히, 토큰 혼합(Token mixing) 기법을 사용하여 더 긴 시퀀스를 효과적으로 처리할 수 있으며, 이는 개체의 의미가 주변 문맥에 의존적인 NER 태스크에서 유용하다[20][21]. 또한, LLaMA는 다국어 데이터를 활용하여 학습되었으며, 다양한 도메인에 적용할 수 있다. 이러한 특성으로 인해 NER 작업에서도 fine-tuning을 수행하면 높은 정확도를 달성할 수 있으며, 특히 저자원 언어나 특수 도메인에서도 좋은 성능을 보인다[20].

2.7 앙상블 기법과 가중치 투표

앙상블 기법은 여러 모델의 예측을 결합하여 성능을 향상시키는 방법이다[22]. 그중 가중치 투표(Weighted voting)는 각 모델의 예측에 가중치를 부여하여 최종 결정을 내리는 방식으로, 개별 모델보다 높은 성능을 기대할 수 있다[22].

가중치 부여 방식은 다양한 전략을 활용할 수 있다. 예를 들어, 특정 개체(예: '단위', '양')에 대한 인식 성능이 뛰어난 모델에 더 높은 가중치를 부여하는 방식으로 조정할 수 있으며, 학습 과정에서 각 모델의 예측 신뢰도를 동적으로 조정하는 방법도 적용할 수 있다[13].

NER 작업에서는 트랜스포머 모델과 대규모 언어 모델(LLM)의 장점을 결합함으로써 성능을 더욱 향상시킬 수 있다[14]. BERT와 같은 트랜스포머 모델은 양방향 문맥을 고려하여 개체를 정밀하게 인식하며, 특정 도메인에 대한 fine-tuning이 용이하다[9]. 반면, GPT와 같은 LLM은 방대한 지식을 바탕으로 높은 일반화 성능을 보이며, 적은 학습 데이터로도 효과적인 예측이 가능하다[19].

III. 관련 연구

3.1 BERT 기반 식품 개체명 인식 연구

최근 BERT 기반 모델을 활용한 식품 개체명 인식(Food NER, Food Named Entity Recognition)이 주목받고 있다. R. Stojanov et al.[23]의 연구에서는 사전 훈련된 BERT 모델을 세 가지 서로 다른 식품 개체명 데이터셋에서 미세 조정(Fine-tuning)하여 FoodNER 모델을 개발하였다. 본 연구는 식품 정보를 추출하는 여러 가지 태스크(식품 vs. 비식품, 세부 식품 그룹 분류 등)에 대한 성능을 평가하였으며, 가장 높은 경우 94.31%의 F1-score를 기록하며 기존 식품 정보 추출 방법 대비 우수한 성능을 보였다. 또한, FoodNER 모델은 FoodOn, Hansard, SNOMED-CT 등 여러 식품 시맨틱 태그를 적용하여 개체명을 분류할 수 있도록 확장되었다.

3.2 대규모 언어 모델을 활용한 식품 개체명 인식

최근 ChatGPT와 같은 대규모 언어 모델(LLM)이 다양한 자연어 처리 태스크에서 활용되고 있으며, 식품 개체명 인식에도 적용되고 있다. M. Ogrinc et al.[24]의 연구에서는 ChatGPT-3.5 및 ChatGPT-4 모델을 활용하여 식품 개체명 인식과 개체명 연결(NEL, Named Entity Linking) 작업을 수행하였다. 본 연구에서는 ChatGPT 모델이 기존의 학습 데이터 없이도(ZSL, Zero-Shot Learning) 상당한 성능을 발휘할 수 있음을 입증하였으나, 개체명 연결 성능은 상대적으로 낮았다. 특히, ChatGPT의 성능은 식품 개체명을 추출하는 작업에서는 비교적 신뢰할 수 있는 결과를 제공했지만, 특정 개체를 식별하고 지식베이스에 연결하는 작업에서는 정확도가 낮아지는 한계를 보였다.

3.3 ChatGPT와 앙상블 학습을 통한 개체명 인식 성능 향상

T. C. Chen et al.[25]의 연구에서는 ChatGPT를 앙상블 학습에서 조정자(arbitrator)로 활용하여 불연속 개체명 인식(DNER, Discontinuous Named Entity Recognition)의 성능을 개선하는 방법을 제안하였다. 본 연구에서는 기존 개체명 인식 모델 다섯 가지를 조합하고, ChatGPT를 활용하여 최종 예측을 조정하는 방식으로 F1-score를 기존 모델 대비 1.13%~0.67%까지 향상시켰다. 특히, ChatGPT가 다양한 모델의 출력을 조합하여 보다 정밀한 개체명 인식을 수행할 수 있도록 하는 프롬프트 엔지니어링 기법이 연구되었다. 이러한 연구는 ChatGPT와 기존 NER 모델을 결합하여 성능을 향상시킬 수 있는 가능성을 제시하며, 식품 개체명 인식에서도 적용 가능성이 있다.

IV. 식품 도메인 개체명 인식 모델 및 시스템

기존의 개체명 인식(NER) 모델들은 일반적인 개체명 인식을 수행하는 데 초점을 맞추고 있으나, 식

품 도메인에서의 개체명 인식은 특정 도메인의 전문적인 개체(예: 식품명, 단위, 양 등)를 보다 정밀하게 인식해야 한다. 본 연구에서는 식품 도메인 특화 개체명을 효과적으로 인식하기 위해 트랜스포머 기반 모델과 대규모 언어 모델(LLM)을 활용하여 최적의 개체명 인식 시스템을 개발하였다.

본 연구에서는 BERT 및 ELECTRA 모델을 활용하여 개체명 인식을 수행하였으며, 성능 향상을 위해 CRF(Conditional Random Fields)와 bi-LSTM을 결합하는 실험을 진행하였다. 또한, GPT-3.5-turbo와 LLaMA 2 7B를 활용하여 대규모 언어 모델을 기반으로 한 개체명 인식 실험을 수행하고, 최종적으로 ELECTRA+CRF 모델과 LLaMA 2 7B 모델을 결합하여 최적의 개체명 인식 시스템을 구축하였다.

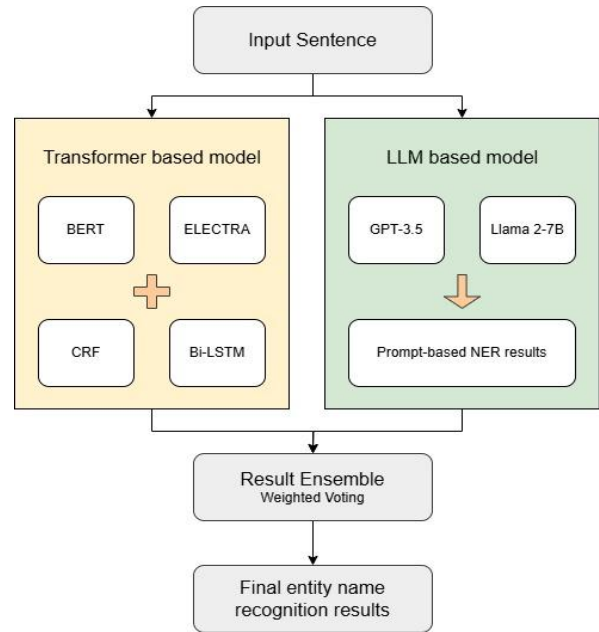


그림 1. 시스템 아키텍처
Fig. 1. System architecture

V. 실험

5.1 데이터 전처리

본 연구에서는 개체명 인식 모델의 성능을 평가하기 위해 기존 TASTESet에서 필수적인 태그(식품 Food, 단위 Unit, 양 Quantity, 부위 Part)만 추출한 데이터셋을 활용하였다.

이 데이터셋에는 BIO 태깅(Begin-Inside-Outside) 방식이 적용되었으며, 학습용과 검증용 데이터로 분할하여 실험을 진행하였다. 데이터셋을 분할할 때, 일부 레이블(B-PART, I-UNIT, I-PART)은 전체 데이터에서 출현 빈도가 낮아 희소 라벨(Sparse label)로 구분하였다. 이러한 희소 라벨이 포함된 문장과 일반 문장 두 가지 8:2 비율로 분리하여 학습이나 검증 데이터 중 한 곳에 희소 라벨이 집중되거나 부족한 현상을 막기 위해 학습 데이터에 균형을 맞추었다. 또한 희소 라벨을 포함한 문장들이 전체 학습 과정에서 충분히 반영될 수 있도록, k-fold(k=3) 교차 검증을 적용하여 모델이 다양한 분할에 대해 학습하고 평가되도록 하였다. BIO 태깅이 적용된 데이터셋의 분포는 표 1과 그림 1에 제시하였다.

표 1. 데이터셋 분포
Table 1. Dataset distribution

Tag(Label)	Counts
B-FOOD	3,245
B-QUANTITY	3,108
O(others)	2,945
B-UNIT	2,625
I-FOOD	1,851
I-QUANTITY	336
B-PART	48
I-UNIT	29
I-PART	7

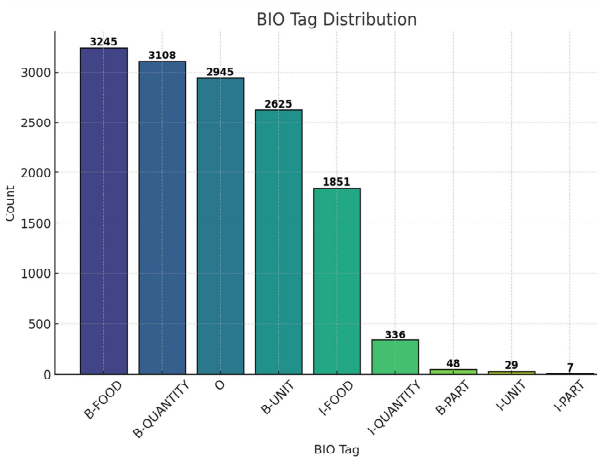


그림 2. 데이터셋 분포
Fig. 2. Dataset distribution

희소 라벨은 전체 데이터셋에서 차지하는 비율이 1% 미만으로 매우 낮아, 모델이 해당 개체명을 효과적으로 학습하지 못하거나 데이터 분할 과정에서 학습 데이터에서 제외될 위험이 있었다. 이를 방지하기 위해, 희소 라벨이 포함된 문장과 일반 문장을 분리한 후, 데이터를 분할하였다.

이러한 전처리 과정을 통해, 모델이 일반적인 개체명뿐만 아니라 희소 개체명도 효과적으로 학습할 수 있도록 데이터 균형을 조정하였다. 데이터셋의 총 문장 수는 2,995개이며, 이 중 희소 라벨을 포함한 문장은 73개, 일반 라벨만 포함된 문장은 2,922개이다.

5.2 트랜스포머 기반 개체명 인식 모델 학습

본 연구에서는 트랜스포머 기반 개체명 인식 모델로 BERT 및 ELECTRA 모델을 활용하여 개체명 인식을 수행하였다. 두 모델은 자연어 처리(NLP)에서 널리 사용되며, 문맥을 이해하고 개체명을 학습하는 데 강점을 가진다. 특히, ELECTRA는 BERT보다 학습 효율성이 높은 것으로 알려져 있으며, 적은 데이터로도 높은 성능을 달성할 수 있다.

BERT 및 ELECTRA 모델의 성능을 평가하기 위해 Precision(정밀도), Recall(재현율), F1-score를 측정하였으며, 개체명 태그별 성능은 표 2와 표 3에 제시되었다.

표 2. BERT 모델 성능 비교
Table 2. BERT model performance comparison

Label	Precision	Recall	F1-score
B-FOOD	0.84	0.83	0.83
B-PART	1	0.5	0.67
B-QUANTITY	0.87	0.9	0.88
B-UNIT	0.88	0.88	0.88
I-FOOD	0.79	0.76	0.78
I-PART	0	0	0
I-QUANTITY	0.82	0.85	0.84
I-UNIT	1	0.33	0.5
O	0.82	0.86	0.84
macro avg	0.8	0.66	0.71
weighted avg	0.83	0.84	0.83

표 3. ELECTRA 모델 성능 비교

Table 3. ELECTRA model performance comparison

Label	Precision	Recall	F1-score
B-FOOD	0.87	0.85	0.85
B-PART	0.7	0.8	0.74
B-QUANTITY	0.9	0.91	0.9
B-UNIT	0.91	0.91	0.91
I-FOOD	0.83	0.86	0.84
I-PART	0	0	0
I-QUANTITY	0.88	0.85	0.86
I-UNIT	0.33	0.33	0.33
O	0.82	0.86	0.83
macro avg	0.69	0.7	0.7
weighted avg	0.86	0.75	0.87

BERT와 ELECTRA 모두 일반 라벨에 대해서는 전반적으로 좋은 성능을 보였으나, 희소 라벨에 대해서는 다소 낮은 성능인 것을 확인할 수 있었다. 특히, 개수가 적은 I-PART에 대해서는 전혀 예측되지 않았다. 전체 점수를 통해서는 희소 라벨을 잘 구분하는지 성능을 파악하기 어려워, 별도로 희소 라벨과 일반 라벨을 구분하여 macro avg를 도출하였다. 표 4는 BERT와 ELECTRA의 희소 라벨과 일반 라벨, 전체 라벨에 대해 평균을 구한 성능을 나타내고 있다.

표 4. BERT와 ELECTRA의 희소 라벨 및 일반 라벨 macro avg 성능 비교

Table 4. Macro average performance comparison of sparse and general Labels in BERT and ELECTRA

Model	Label	Precision	Recall	F1-score
BERT	Sparse	0.34	0.37	0.35
	General	0.86	0.87	0.87
	All	0.8	0.66	0.71
ELECTRA	Sparse	0.52	0.58	0.55
	General	0.86	0.87	0.87
	All	0.83	0.84	0.83

전체적인 모델의 성능을 더욱 향상시키기 위해서는 희소 라벨의 판별하는 성능을 높이는 것이 중요하다. BIO 태깅을 적용하였기에 B 태그 다음에 I 태그가 오는 규칙이 있다. 이를 탐지할 수 있는 CRF (Conditional Random Fields)와 bi-LSTM (Bidirectional Long Short-Term Memory)을 추가하는 실험을 수행

하였다. CRF는 개체명 간의 의존성을 학습하여 모델이 일관된 예측을 할 수 있도록 보완한다. 특히 희소 라벨이 포함된 문장에서 역손된 개체명 태깅 오류를 줄이는 데 기여한다. bi-LSTM은 문맥 정보를 장거리로 반영할 수 있어, 다중 토큰으로 구성된 개체명을 인식할 때 유리하다. 이전 문맥과 이후 문맥을 모두 고려하여 문장 내 태깅 오류를 최소화할 수 있다. 이를 반영하여 BERT와 ELECTRA에 CRF와 bi-LSTM을 추가하여 학습을 진행하였다. 표 5는 CRF와 bi-LSTM을 추가한 후, BERT와 ELECTRA의 희소 라벨 및 일반 라벨, 전체 라벨에 대한 성능 지표표를 나타낸다.

표 5의 결과를 통해, CRF를 추가한 ELECTRA 모델이 희소 라벨과 일반 라벨 모두에서 가장 높은 성능을 기록했음을 확인할 수 있다. 이러한 결과를 바탕으로 CRF 및 bi-LSTM을 추가한 후의 성능 변화에 대해 다음과 같이 분석할 수 있다.

표 5. CRF와 bi-LSTM을 추가한 BERT와 ELECTRA의 희소 라벨 및 일반 라벨 macro avg 성능 비교

Table 5. Macro Average Performance Comparison of Sparse and General Labels in BERT and ELECTRA with CRF and bi-LSTM

Model	Label	Precision	Recall	F1-score
BERT +CRF	Sparse	0.82	0.56	0.51
	General	0.92	0.90	0.90
	All	0.87	0.73	0.70
BERT +bi-LSTM	Sparse	0.31	0.56	0.38
	General	0.93	0.93	0.93
	All	0.62	0.74	0.65
ELECTRA +CRF	Sparse	0.93	0.86	0.89
	General	0.92	0.93	0.92
	All	0.92	0.89	0.90
ELECTRA +bi-LSTM	Sparse	0.83	0.57	0.64
	General	0.91	0.93	0.91
	All	0.87	0.75	0.78

BERT와 ELECTRA 모두 CRF를 추가한 경우 희소 라벨 판별 성능이 향상됨을 확인할 수 있다. ELECTRA+CRF 모델은 희소 라벨에서 F1-score 0.89을 기록하며 다른 모든 모델보다 높은 성능을 보였다. BERT+CRF 모델은 BERT만을 사용하였을 때보다 성능이 향상하였지만 희소 라벨에서 F1-score

0.51을 기록하여 ELECTRA+CRF에 비해 상대적으로 낮은 성능을 보였다. 이는 ELECTRA 모델이 효과적으로 패턴을 학습하며, CRF가 개체명 간의 문맥적 의존성을 활용하여 희소 라벨의 예측 성능을 더욱 향상시켰음을 의미한다. bi-LSTM 추가 시 일반 라벨의 성능은 향상되었지만, 희소 라벨에서는 큰 영향을 미치지 못했다. BERT+bi-LSTM 모델은 일반 라벨에서 F1-score 0.93으로 가장 높은 성능을 보였으나, 희소 라벨에서는 F1-score 0.38로 낮은 성능을 기록하였다. ELECTRA+bi-LSTM 모델도 희소 F1-score가 0.64로 비교적 낮은 성능을 기록하였다. 이는 bi-LSTM이 문맥 정보를 장거리로 반영하는데 강점을 가지지만, 데이터가 부족한 희소 개체명 인식에는 큰 영향을 미치지 못함을 시사한다.

본 연구에서는 희소 라벨을 효과적으로 판별하기 위해 CRF와 bi-LSTM을 추가한 실험을 수행하였으며, ELECTRA+CRF 모델이 가장 우수한 성능을 기록함을 확인하였다. 특히 CRF를 추가한 경우 희소 라벨의 성능이 크게 향상되었으며, bi-LSTM은 일반 개체명 인식에서는 유리하지만 희소 개체명 판별에서는 큰 효과를 보이지 않았다. 이를 바탕으로 본 연구에서는 ELECTRA+CRF 모델을 대규모 언어 모델과 결합하여 성능을 더욱 향상시키는 실험을 진행하였다.

5.3 대규모 언어 모델 기반 개체명 인식 모델 학습

본 연구에서는 트랜스포머 기반 개체명 인식 모델(BERT, ELECTRA)과 함께, 대규모 언어 모델(LLM)의 개체명 인식 성능을 분석하기 위해 GPT-3.5-turbo(OpenAI)와 LLaMA 2 7B(Meta AI)를 활용하여 실험을 수행하였다.

기존의 트랜스포머 기반 사전 학습(pre-trained)된 모델을 fine-tuning하여 개체명을 학습하는 방식으로 개체명 인식을 수행한다. 대규모 언어 모델 또한 사전 학습된 모델을 활용하되, 본 연구의 데이터셋을 이용하여 직접 fine-tuning을 수행함으로써 개체명 인식 성능을 평가하였다.

5.3.1 ChatGPT 학습

표 6은 GPT-3.5-turbo(ChatGPT)의 희소 라벨 및 일반 라벨의 macro avg 성능을 나타낸다.

특히, 희소 라벨의 경우 GPT-3.5-turbo의 F1-score는 0.27에 불과하여, 일반 개체명 인식에 비해 성능이 낮음을 확인할 수 있다.

5.3.2 LLaMA 학습

표 6의 결과를 보면, LLaMA 2 7B가 GPT-3.5-turbo보다 희소 라벨과 일반 라벨에서 전반적으로 더 높은 성능을 기록한 것을 확인할 수 있다. 특히, 희소 라벨의 경우 GPT-3.5-turbo의 F1-score는 0.27에 불과한 반면, LLaMA 2 7B는 0.42를 기록하며 상대적으로 높은 성능을 보였다.

희소 라벨에서의 Recall은 GPT-3.5-turbo(0.24)에 비해 LLaMA 2 7B(0.42)가 더 높은 값을 기록하며, 이는 희소 개체명을 상대적으로 더 잘 검출함을 의미한다. 일반 라벨에서는 두 모델의 F1-score는 큰 차이를 보이지 않는다. 일반적인 개체명(B-FOOD, B-UNIT, B-QUANTITY 등)의 경우, 두 모델 모두 충분한 학습이 이루어진 것을 알 수 있다. 하지만 LLaMA 2 7B의 Precision(0.78)이 GPT-3.5-turbo(0.67)보다 높아, 보다 정확한 개체명 판별이 가능했음을 확인할 수 있다.

표 6. GPT-3.5-turbo와 LLaMA 2 7B의 희소 라벨 및 일반 라벨의 macro avg 성능 비교

Table 6. Macro average performance comparison of sparse and general labels in GPT-3.5-turbo and LLaMA 2 7B

Model	Label	Precision	Recall	F1-score
GPT-3.5-turbo	Sparse	0.31	0.24	0.27
	General	0.67	0.72	0.69
	All	0.49	0.48	0.48
LLaMA 2 7B	Sparse	0.75	0.42	0.42
	General	0.78	0.75	0.76
	All	0.76	0.58	0.59

본 연구에서 수행한 트랜스포머 기반 개체명 인식 모델(BERT, ELECTRA)과의 성능을 비교한 결과, ELECTRA+CRF 모델이 희소 라벨에서 가장 높은 성능을 기록(F1-score 0.89)하였으며, LLaMA 2

7B(0.42)와 GPT-3.5-turbo(0.27)는 상대적으로 낮은 성능을 보였다. 일반 라벨에서는 ELECTRA+CRF 모델이 0.92의 F1-score를 기록하며 가장 높은 성능을 보였으며, LLaMA 2 7B(0.76)와 GPT-3.5-turbo(0.69)도 비교적 높은 성능을 기록하였다.

이러한 결과는 대규모 언어 모델이 일반적인 개체명을 인식하는 데 효과적일 수 있으나, 희소 개체명에 대한 학습 능력은 상대적으로 떨어지는 한계를 보였음을 의미한다. 특히, 희소 개체명의 경우 트랜스포머 기반 모델이 더 높은 성능을 기록하였으며, 이는 대규모 언어 모델이 적은 데이터로 학습된 개체명을 인식하는 데 어려움을 겪을 수 있음을 시사한다. 따라서, 트랜스포머 기반 모델의 정밀한 희소 개체명 인식 성능과 대규모 언어 모델의 넓은 개체명 학습 능력을 결합하면, 각각의 강점을 보완하여 보다 향상된 개체명 인식 성능을 기대할 수 있다.

5.4 최종 앙상블 모델 구축 및 성능 분석

5.4.1 앙상블 가중치 설정

최종 개체명 예측 값은 ELECTRA+CRF 모델과 LLaMA 2 7B 모델의 예측 확률 값을 기반으로 가중치를 적용하여 계산된다. 이를 수식으로 표현하면 다음과 같다.

$$P_{final} = w_1 \cdot P_{ELECTRA+CRF} + w_2 \cdot P_{Llama2} \quad (1)$$

여기서 P_{final} 은 최종 개체명 예측 확률을 의미하며, $P_{ELECTRA+CRF}$ 와 P_{Llama2} 는 각각 ELECTRA+CRF 모델과 LLaMA 2 7B 모델이 예측한 확률 값을 나타낸다. 또한, w_1 및 w_2 는 각각 ELECTRA+CRF 모델과 LLaMA 2 7B 모델에 부여된 가중치이며, 두 가중치의 합은 다음 조건을 만족한다.

$$w_1 + w_2 = 1 \quad (2)$$

즉, 각 모델의 예측 확률을 서로 다른 비율로 반영하여 최종 예측 값을 결정하는 방식을 사용하였다.

이를 통해 ELECTRA+CRF 모델의 강점을 유지하면서도, LLaMA 2 7B가 문맥적 보완 역할을 수행할 수 있도록 가중치를 조정하였다.

본 연구에서는 ELECTRA+CRF 모델과 LLaMA 2 7B 모델을 결합하여 최종 앙상블 모델을 구축하였다. 기존 실험 결과에서 ELECTRA+CRF 모델은 희소 개체명(B-PART, I-UNIT 등)에 대해 높은 성능을 보였으며, 특히 희소 라벨의 F1-score가 0.89로 가장 우수하였다. 일반 개체명 인식에서도 ELECTRA+CRF 모델이 LLaMA 2 7B 모델보다 높은 성능을 기록하였으나, 대규모 언어 모델은 트랜스포머 기반 모델이 놓칠 수 있는 문맥적 정보를 학습하는 데 장점이 있었다. 특히, 데이터에서 등장하지 않았던 개체명을 예측하는 능력(Out-of-Vocabulary, OOV 처리)이 트랜스포머 모델보다 더 강할 가능성이 있다. 따라서, 가중치를 LLaMA 2 7B에 과도하게 부여하지 않고 ELECTRA+CRF 중심으로 결합하되 일부 문맥적 보완 역할을 수행하도록 조정하였다.

희소 개체명의 경우 ELECTRA+CRF 모델이 압도적으로 성능이 우수하므로, ELECTRA+CRF에 더 높은 가중치를 부여하였다. 일반 개체명의 경우에도 ELECTRA+CRF의 성능이 더 좋았으나, LLaMA 2 7B의 문맥적 보완 가능성을 고려하여 적절히 낮은 가중치로 반영하였다. 최종적으로 적용된 가중치는 다음 표 7과 같다.

표 7. 최종 앙상블 모델의 희소 라벨 및 일반 라벨 가중치 설정

Table 7. Weight allocation for sparse and general labels in the final ensemble model

Label	w_1 (ELECTRA+CRF)	w_2 (LLaMA 2 7B)
Sparse	0.85	0.15
General	0.75	0.25

5.4.2 가중치 적용 후 전체적인 모델 성능

이러한 가중치 결합 방식을 적용한 최종 앙상블 모델의 성능을 분석하기 위해, ELECTRA+CRF 및 LLaMA 2 7B 단일 모델과 비교하였다. 표 8은 가중치 결합을 적용한 앙상블 모델의 희소 라벨 및 일반 라벨에 대한 성능을 나타낸다.

표 8. 최종 앙상블 모델의 희소 라벨 및 일반 라벨의 macro avg 성능 비교

Table 8. Macro average performance comparison of sparse and general labels in the final ensemble model

Model	Label	Precision	Recall	F1-score
ELECTRA+CRF	Sparse	0.93	0.86	0.89
	General	0.92	0.93	0.92
	All	0.92	0.89	0.90
LLaMA 2 7B	Sparse	0.75	0.42	0.42
	General	0.78	0.75	0.76
	All	0.76	0.58	0.59
Ensemble	Sparse	0.94	0.88	0.91
	General	0.93	0.94	0.93
	All	0.92	0.91	0.92

표 8의 결과를 보면, 가중치 결합을 적용한 앙상블 모델이 희소 라벨과 일반 라벨 모두에서 높은 성능을 기록하였음을 확인할 수 있다. 희소 라벨의 F1-score는 0.89에서 0.91로, 전체 성능은 0.90에서 0.92로 향상되었으며, 일반 라벨에서도 소폭 개선된 수치를 보였다. 이는 ELECTRA+CRF 단독 모델과 비교했을 때 수치상으로는 근소한 차이지만, 희소 라벨과 같이 출현 빈도가 낮고 학습이 어려운 개체에 대해 문맥 보완 효과가 작용했을 가능성을 시사한다.

특히, LLaMA 2 7B는 단독 성능은 낮았지만, ELECTRA+CRF가 간과할 수 있는 문맥의 뉘앙스를 반영하는 데 기여함으로써, 기존 모델의 강점을 유지하면서도 일부 예외적 상황에서의 오류를 보완할 수 있는 구조적 이점을 제공하였다. 실제로 일반 개체명에서는 ELECTRA+CRF가 이미 높은 성능을 보였기 때문에, LLaMA 2 7B의 가중치를 낮게 부여하더라도 일정 수준의 보완 효과가 있었던 것으로 보인다.

이러한 관점에서, LLM의 추가는 전체 성능을 비약적으로 끌어올리는 목적보다는, 기존 모델이 놓치기 쉬운 희소 개체나 문맥적으로 모호한 사례에서의 보완적 역할에 의의가 있으며, 이는 정확성이 중요한 도메인 특화 NER 태스크에서 유용하게 활용될 수 있다.

VI. 결론 및 향후 과제

본 연구에서는 식품 도메인에서의 개체명 인식 성능을 개선하기 위해, 트랜스포머 기반 모델과 대

규모 언어 모델을 결합하는 방법을 탐색하였다.

이를 위해, 식품 개체명 데이터를 BIO 태깅 방식으로 정리하고, 희소 개체명(B-PART, I-UNIT 등)을 구분하여 학습하는 방식을 적용하였다.

실험 결과, ELECTRA+CRF 모델이 희소 개체명(B-PART, I-UNIT 등)에서 가장 높은 성능을 기록하였으며, LLaMA 2 7B 모델은 일반 개체명(B-FOOD, B-QUANTITY 등)에서 보완적인 역할을 수행할 수 있음을 확인하였다. 이를 바탕으로, ELECTRA+CRF 모델을 중심으로 하되, LLaMA 2 7B 모델의 문맥적 보완 효과를 반영하는 가중치 결합 방식(Weighted voting)을 적용하여 최종 앙상블 모델을 구축하였다.

최종 실험 결과, 희소 개체명의 F1-score가 0.89(ELECTRA+CRF)에서 0.91(앙상블 모델)로 향상되었으며, 일반 개체명의 F1-score도 0.92에서 0.93으로 증가하여 전반적인 성능 개선을 달성하였다. 이를 통해, 식품 도메인 개체명 인식에서 트랜스포머 기반 모델과 대규모 언어 모델의 결합이 효과적임을 실험적으로 검증하였으며, 희소 개체명 판별 성능 향상을 위한 앙상블 기법의 가능성을 확인하였다.

본 연구에서는 ELECTRA + CRF 모델과 LLaMA 2 7B 모델을 결합하여 개체명 인식 성능을 향상시키는 방법을 탐색하였으며, 실험 결과 해당 방법이 효과적임을 확인하였다. 그러나, 개체명 인식 성능을 더욱 개선하기 위해 몇 가지 연구 방향을 고려할 필요가 있다.

우선, 가중치 조정 방식의 최적화가 필요하다. 본 연구에서는 ELECTRA+CRF와 LLaMA 2 7B의 가중치를 고정된 값(희소 개체명: 0.85, 0.15 / 일반 개체명: 0.75, 0.25)으로 설정하였다. 향후 연구에서는 각 개체명 태그별 성능을 동적으로 조정하는 방식(Adaptive weighting)을 적용하여, 가중치를 최적화하는 방법을 탐색할 필요가 있다. 또한, 다른 대규모 언어 모델과의 비교를 검토할 필요가 있다. 본 연구에서는 GPT-3.5-turbo 및 LLaMA 2 7B를 실험 대상으로 선정하였으나, GPT-4, LLaMA 2 13B, PaLM 등 다양한 대규모 언어 모델과의 비교 실험을 수행할 경우, 보다 효과적인 모델 조합을 탐색할 수 있을 것이다. 개체명 인식 모델의 확장 가능성을 고려하는 것도 중요한 연구 방향이다.

본 연구에서 제안한 방법은 식품 개체명 인식에 초점을 맞추었으나, 의료, 화학, 농업 등 특정 도메인의 개체명 인식 문제에도 적용 가능성을 확장할 수 있다. 향후 연구에서는 도메인별 데이터셋을 활용하여 본 연구에서 제안한 모델의 확장성을 검토할 필요가 있다.

이러한 연구 방향을 통해, 보다 정밀한 개체명 인식이 가능해질 것이며, 식품 도메인을 넘어 다양한 응용 분야에도 활용할 수 있을 것으로 기대된다. 특히, 본 연구 결과는 실제 산업 현장에서의 활용 가능성도 지닌다. 예를 들어, 식품 이력 추적 시스템에서 원재료명 및 유통정보 자동 추출, 영양 성분 관리 시스템에서 성분명 및 단위 인식, 나아가 개인 맞춤형 식단 분석에서 질환/약물과 상호작용이 있는 식품 성분의 판별 등에 적용할 수 있다. 이러한 응용 가능성은 트랜스포머 기반 NER 모델과 대규모 언어 모델의 결합이 가져올 수 있는 실제적 효과를 뒷받침하며, 향후 연구에서는 도메인별 시나리오 기반 성능 평가 및 확장성 검토가 요구된다.

References

- [1] R. Stojanov, G. Popovski, G. Cenikj, B. Koroušič Seljak, and T. Eftimov, "A Fine-Tuned Bidirectional Encoder Representations From Transformers Model for Food Named-Entity Recognition: Algorithm Development and Validation", *J. Med. Internet Res.*, Vol. 23, No. 8, pp. e28229, Aug. 2021. <https://doi.org/10.2196/28229>.
- [2] Q. Zhang, M. Li, W. Dong, M. Zuo, S. Wei, S. Song, and D. Ai, "An Entity Relationship Extraction Model Based on BERT-BLSTM-CRF for Food Safety Domain", *Comput. Intell. Neurosci.*, Vol. 2022, pp. 7773259, Apr. 2022. <https://doi.org/10.1155/2022/7773259>.
- [3] J. M. Ordovas, L. R. Ferguson, E. S. Tai, and J. C. Mathers, "Personalised nutrition and health", *BMJ*, Vol. 361, pp. k2173, Jun. 2018. <https://doi.org/10.1136/bmj.k2173>.
- [4] D. Choi, M. Gim, S. Badreddine, H. Kim, D. Park, and J. Kang, "KitchenScale: Learning to predict ingredient quantities from recipe contexts", *Expert Systems with Applications*, Vol. 224, pp. 120041, Aug. 2023. <https://doi.org/10.1016/j.eswa.2023.120041>.
- [5] D. Batra, N. Diwan, U. Upadhyay, J. S. Kalra, T. Sharma, A. K. Sharma, D. Khanna, J. S. Marwah, S. Kalathil, N. Singh, R. Tuwani, and G. Bagler, "RecipeDB: a resource for exploring recipes", *Database*, Vol. 2020, pp. baaa077, Nov. 2020. <https://doi.org/10.1093/database/baaa077>.
- [6] G. Popovski, T. Eftimov, D. Mladenčić, and B. Koroušič Seljak, "Ontology alignment using Named-Entity Recognition methods in the domain of food", *Proc. SiKDD 2020 — Slovenian KDD Conference on Data Mining and Data Warehouses*, Ljubljana, Slovenia, pp. 1-4, Oct. 2020.
- [7] F. Mohbat and M. J. Zaki, "KERL: Knowledge-Enhanced Personalized Recipe Recommendation using Large Language Models", *arXiv preprint arXiv:2505.14629*, May 2025. <https://doi.org/10.48550/arXiv.2505.14629>.
- [8] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification", *Linguisticae Investigationes*, Vol. 30, No. 1, pp. 3-26, Jan. 2007. <https://doi.org/10.1075/li.30.1.03nad>.
- [9] X. Zhou, X. Zhang, and X. Hu, "MaxMatcher: Biological Concept Extraction Using Approximate Dictionary Lookup", in *Proc. PRICAI 2006: Trends in Artificial Intelligence (Lecture Notes in Computer Science*, Vol. 4099), pp. 1145-1149, Aug. 2006. https://doi.org/10.1007/978-3-540-36668-3_150.
- [10] Restack, "Fine-tuning BERT for Named Entity Recognition", <https://www.restack.io/p/fine-tuning-answer-bert-named-entity-recognition-cat-ai>. [accessed: Feb. 03, 2025]
- [11] A. Wróblewska, et al., "TASTESet--Recipe Dataset and Food Entities Recognition Benchmark", *arXiv preprint, arXiv:2204.07775*, Apr. 2022. <https://doi.org/10.48550/arXiv.2204.07775>.

- [12] I. Keraghel, S. Morbieu, and M. Nadif, "Recent Advances in Named Entity Recognition: A Comprehensive Survey and Comparative Study", arXiv, Vol. 2401.10825, Dec. 2024. <https://doi.org/10.48550/arXiv.2401.10825>.
- [13] K. Pakhale, "Comprehensive Overview of Named Entity Recognition: Models, Domain-Specific Applications and Challenges", arXiv, Vol. 2309.14084, Sep. 2023. <https://doi.org/10.48550/arXiv.2309.14084>.
- [14] W. Ye, B. Li, R. Xie, Z. Sheng, L. Chen, and S. Zhang, "Exploiting Entity BIO Tag Embeddings and Multi-task Learning for Relation Extraction with Imbalanced Data", arXiv, Vol. 1906.08931, Jun. 2019. <https://doi.org/10.48550/arXiv.1906.08931>.
- [15] L. A. Ramshaw and M. P. Marcus, "Text chunking using transformation-based learning", Natural Language Processing Using Very Large Corpora, Vol. 11, pp. 157-176, 1999. https://doi.org/10.1007/978-94-017-2390-9_10.
- [16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding", Proc. NAACL-HLT, Minneapolis, MN, USA, pp. 4171-4186, Jun. 2019. <https://doi.org/10.18653/v1/N19-1423>.
- [17] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, "ELECTRA: Pre-training text encoders as discriminators rather than generators", Proc. ICLR, Addis Ababa, Ethiopia, Mar. 2020. <https://doi.org/10.48550/arXiv.2003.10555>.
- [18] T. B. Brown, et al., "Language models are few-shot learners", Proc. NeurIPS, online, pp. 1877-1901, May 2020. <https://doi.org/10.48550/arXiv.2005.14165>.
- [19] W. S. Rustad, G. F. Shidik, E. Noersasongko, Purwanto, Muljono, and D. R. I. M. Setiadi, "Systematic Literature Review on Named Entity Recognition: Approach, Method, and Application", Stat. Optim. Inf. Comput., Vol. 12, No. 4, pp. 907-942, Feb. 2024. <https://doi.org/10.19139/soic-2310-5070-1631>.
- [20] H. Touvron, et al., "LLaMA: Open and efficient foundation language models", arXiv preprint arXiv:2302.13971, Feb. 2023. <https://doi.org/10.48550/arXiv.2302.13971>.
- [21] Z. Liu, F. Jiang, Y. Hu, C. Shi, and P. Fung, "NER-BERT: A Pre-trained Model for Low-Resource Entity Tagging", arXiv, Vol. 2112.00405, Dec. 2021. <https://doi.org/10.48550/arXiv.2112.00405>.
- [22] T. G. Dietterich, "Ensemble methods in machine learning", Proc. Int. Workshop Multiple Classifier Syst., Cagliari, Italy, Vol. 1857, pp. 1-15, Jan. 2000. https://doi.org/10.1007/3-540-45014-9_1.
- [23] R. Stojanov, G. Popovski, G. Cenikj, B. K. Seljak, and T. Eftimov, "A Fine-Tuned Bidirectional Encoder Representations From Transformers Model for Food Named-Entity Recognition: Algorithm Development and Validation", Journal of Medical Internet Research, Vol. 23, No. 8, pp. e28229, Aug. 2021. <https://doi.org/10.2196/28229>.
- [24] M. Ogrinc, B. K. Seljak, and T. Eftimov, "Zero-shot evaluation of ChatGPT for food named-entity recognition and linking", Frontiers in Nutrition, Vol. 11, Aug. 2024. <https://doi.org/10.3389/fnut.2024.1429259>.
- [25] T.-C. Chen and W.-Y. Lin, "On Fusing ChatGPT and Ensemble Learning in Discontinuous Named Entity Recognition in Health Corpora", Preprint, National University of Kaohsiung, Taiwan, Dec. 2024. <https://doi.org/10.48550/arXiv.2412.16976>.

저자소개

임 소 희 (So-Hee Lim)



2025년 2월 : 경상국립대학교
컴퓨터공학과(공학사)
2025년 3월 ~ 현재 :
경상국립대학교 컴퓨터공학과
석사과정
관심분야 : 인공지능, 데이터 분석,
자연어 처리

김 건 우 (Gun-Woo Kim)



2006년 12월 : 호주뉴캐슬대학교
컴퓨터공학과(공학사)
2007년 9월 : 호주뉴캐슬대학교
정보공학과(공학석사)
2017년 8월 : 한양대학교
컴퓨터공학과(공학박사)
2021년 9월 ~ 현재 :
경상국립대학교 컴퓨터공학과 부교수
관심분야 : 인공지능, 시멘틱 헬스케어, 데이터마이닝