

# ChatGPT와 KoBERT을 활용한 지리 정보 처리: 개체명 인식과 개체명 연결 성능 비교 연구

## "Comparative Study of Geographical Information Processing with ChatGPT and KoBERT: Performance Comparison of Named Entity Recognition and Named Entity Linking"

최대웅 · 유기운

Dae Woong Choi · Kiyun Yu

서울대학교 공과대학 건설환경공학부 석사과정 (E-mail:cdwoong828@snu.ac.kr)

서울대학교 공과대학 건설환경공학부 교수(E-mail:kiyun@snu.ac.kr)

### 요약

최근 자연어 처리 분야에서 딥 러닝 모델은 많은 관심을 받고 있으며, ChatGPT는 특히 자연어 이해와 생성에 뛰어난 성능을 보여주고 있다. 그러나 지리 정보 처리에 딥 러닝 모델을 적용하는 연구는 아직 제한적이다. 이 논문은 공간 질의 처리 분야에서 Named Entity Recognition(NER)과 Named Entity Disambiguation(NED)을 효과적으로 수행하는 딥 러닝 모델을 개발하는 것을 목표로 한다. ChatGPT와 koBERT 모델을 활용하여 개체명 인식 작업을 수행하고, fuzzy string match를 통해 NED를 수행한다. 학습 데이터셋은 실제 공간 정보를 반영하고 다양한 시나리오를 포함하여 모델의 일반화 능력을 향상시킨다. 그 후 koBERT를 활용한 딥 러닝 모델과 chatGPT 모델의 성능은 정확도, 재현율, F1 점수 등을 기준으로 평가되며, 테스트 데이터 세트를 활용하여 실제 공간 질의에 대한 NER, NED의 정확성과 유용성을 평가한다.

### 1. 서론

최근 자연어 처리 분야에서 딥 러닝 모델은 많은 관심을 받고 있으며, ChatGPT는 특히 자연어 이해와 생성에 뛰어난 성능을 보여주고 있다. 그러나 지리 정보 처리에 딥러닝 모델을 적용하는 연구는 많지 않다. 공간 질문 자연어 처리 분야에서 NER, NED와 같은 Entity Linking은 매우 중요하지만 선행 연구는 많이 찾아볼 수 없다. 따라서 이 연구의 목표는 공간 질의에 대한 NER과 NED를 효과적으로 수행하는 모델을 개발하는 것이다. 현재 사회적으로 이슈가 된 ChatGPT의 API와 한국어 말뭉치로 학습된 koBERT를 fine-tuning해 NER과 NED를 수행한다. 이를 통해 공간 연산을 수행하여 정확한 답변을 생성할 수 있는 시스템의 기반이 되는 모델을 개발 하는 것을 목표로

한다.

### 2. 본론

본 연구에서는 다음과 같은 접근 방식을 활용하여 공간 질의 처리를 개선하고자 한다. 본 논문의 파이프라인을 잘 보여주는 그림 1을 보면 ChatGPT API를 이용해 NER을 수행, 개체명을 인식한 후 ChatGPT API를 이용해 인식된 개체명을 사전 구성된 공간 DB와 매칭시켜 NED를 수행한다. 이와 병렬적으로 사전 한국어 말뭉치로 학습된 koBERT 모델을 fine-tuning해 NER

을 수행, 개체명을 인식한다. koBERT는 한국어 자연어 처리를 위한 사전 학습된 BERT 모델로, ChatGPT와 함께 사용하여 개체명 인식기를 구축하는 데 활용된다. 본 연구에서는 koBERT를 공간 질의 처리 작업에 알맞게 조정하여 개체명 인식기를 개발하고자 한다. koBERT를 활용하여 개체명 인식 작업을 수행하기 위해 이 모델을 공간 질의 데이터의 NER에 더 좋은 성능을 가질 수 있도록 fine-tuning 한다. Fine-tuning은 사전 학습된 모델을 특정 작업에 맞게 세부 조정하는 과정이다. 개체명을 인식한 후 fuzzy string match를 통해 인식된 개체와 공간 dB와의 string match를 통해 NED를 수행한다.

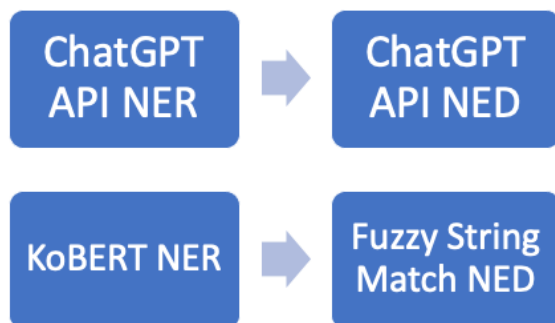


그림 1 학습 절차

학습 데이터셋을 구성하기 위해 공간 관련 질문 데이터를 수집한다. 이 데이터셋은 실제 공간 정보를 반영하고 다양한 시나리오를 포함하여 모델의 일반화 능력을 향상시킨다. 예를 들어, 지도 관련 질문 데이터, 위치 기반 서비스의 질문 데이터, POI 질문 데이터를 활용할 수 있다.

실험을 통해 개발한 koBERT를 활용한 딥 러닝 모델과 ChatGPT의 성능을 비교 분석한다. 성능 평가는 정확도, 재현율, F1 점수 등을 기준으로 수행된다. 또한, 테스트 데이터 세트를 활용하여 실제 공간 질의를 처리하고, 답변의 정확성과 유용성을 평가한다.

### 3. 결론

본 연구는 공간 질의 처리를 위한 NER과 NED

를 효과적으로 수행하는 koBERT 모델을 개발하였고 이를 ChatGPT의 NER 수행 성능과 비교하였다. 이를 통해 공간 연산을 수행하여 정확한 답변을 생성할 수 있는 시스템의 기반이 되고 사용자들에게 정확하고 유용한 공간 정보를 제공하며, 지리 정보와 지식 베이스를 결합한 KBQA 시스템을 발전시킬 수 있는 기반을 마련할 것이다. 끝으로, 이러한 결과는 공간 질의 처리 분야에서 딥 러닝 모델의 유효성과 중요성을 입증하는 역할을 한다. 또한, 공간 정보를 기반으로 한 자연어 처리 작업에서의 NER과 NED의 중요성을 강조하며, 공간 질의에 대한 정확한 답변을 제공하는 시스템의 필요성을 제시한다. 따라서 이 연구는 공간 질의 처리에 딥 러닝 모델을 적용하는 데 있어서 중요한 기여를 하였으며, 향후 더 많은 연구와 응용 분야에서 활용될 수 있을 것이다.

### 감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음 (과제번호 RS-2022-00143336).

### 참고 문헌

- 현지원, 이준일 and 조현권. (2022). KoBERT를 이용한 기업관련 신문기사 감성 분류 연구7). 회계학연구, 47(4), 33-54.
- Kamath, U., Graham, K. L., & Emara, W. (2022). Bidirectional encoder representations from Transformers (Bert). *Transformers for Machine Learning*, 43-70. <https://doi.org/10.1201/9781003170082-3>
- Li, X., Qian, Y., & Dou, L. (2021a). Bert named entity recognition with self-attention mechanism. *2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)*. <https://doi.org/10.1109/tocs53301.2021.9688590>