

배관계장도 검토를 위한 OntoGPT 기반 OntoCAPE 동적 확장에 관한 연구

OntoGPT-based Dynamic OntoCAPE Extension System for P&ID Review

○ 엄 미 영* 윤 희 준** 한 주 호** 이 강***
Uhm Miyoung Youn, Heejun Han, Juho Lee, Ghang

Abstract

Large multimodal models (LMMs) for reviewing piping and instrumentation diagrams (P&IDs) require domain knowledge, but the Ontology for Computer Aided Process Engineering (OntoCAPE) has remained static since 2009, limiting its applicability to automatically review safety issues. To address this gap, this paper extends OntoCAPE by integrating the Ontology-guided Generative Pre-trained Transformer (OntoGPT) framework with rule-based parsing. The method embeds Korean chemical plant safety regulations, covering 86 sentences and yielding 46 class mappings with 44 Resource Description Framework (RDF) triples. Validation using the Semantic Web Rule Language (SWRL) confirms logical consistency and enables automatic detection of safety violations for compliance automation.

키워드 : OntoGPT, OntoCAPE, 배관계장도, 온톨로지, 안전 규정

Keywords : OntoGPT, OntoCAPE, P&ID, Ontology, Safety Regulations

1. 서론

1.1 연구의 배경 및 목적

배관계장도 (P&ID, Piping and Instrumentation Diagram)는 장비 배치, 배관 연결, 계측기기 위치 등 필수적인 정보를 담은 핵심 설계 문서이며, P&ID 검토는 플랜트의 안전성과 운전 신뢰성 확보를 위한 핵심 업무에 해당한다. 그러나, 현재 P&ID 검토 작업은 단순한 도면 해석을 넘어 복잡한 안전 규정과 설계 기준에 대한 깊은 이해를 요구하기에 여전히 전문가의 경험과 지식을 바탕으로 한 종합적 판단이 필요하다. 이러한 전문가 의존적 검토 방식은 높은 신뢰성을 보장하지만, 상당한 시간과 인력이 소요될 뿐 아니라, 검토자 간 일관성 부족과 오류 누락 가능성의 문제가 발생할 수 있다.

거대멀티모달모델(LMMs, Large Multimodal Models)은 객체 인식이나 패턴 매칭을 넘어 복잡한 도면 구조를 이해하고 다양한 형태의 정보를 종합적으로 처리할 수 있는 능력으로 인해 P&ID 자동 검토 도구로 가능성이 논의되고

있다. 그러나, 일반 텍스트와 이미지 데이터로 학습된 LMM을 화학공정과 같은 고도의 전문 분야에 적용하기 위해서는 안전 규정, 설계 기준, 산업 표준 등의 도메인 특화 지식 부족의 문제가 남아있다. 따라서, LMM을 화학공정 분야에서 실현하기 위해서는 온톨로지와 같이 체계적으로 구조화된 도메인 지식의 통합이 필수적이다.

그러나 온톨로지의 개발과 수정 작업은 엄격한 구조 정의와 도메인 전문성이 요구될 뿐 아니라, 지속적으로 변화하는 산업 표준과 규정 반영을 위한 지속적인 업데이트를 위해서는 자동화 기술 도입이 시급하다. OntoGPT는 LMM 에이전트를 활용해 텍스트로부터 지식을 자동으로 추출하고, 이를 기존 온톨로지 구조에 맞게 확장할 수 있는 동적 확장 능력을 제공한다는 점에서 이러한 자동화 요구에 대한 유력한 해결책으로 주목받고 있다 (Palagin, O. et al. 2023). 다만, OntoGPT는 현재까지 생명과학 및 의료분야에 집중되어 있으며, 화학 공정과 P&ID 검토와 같은 산업 현장에서의 응용은 거의 연구된 바 없어 OntoGPT 기반 온톨로지 확장에 대한 산업 도메인 적용 지식이 부족한 실정이다.

이에 본 연구는 OntoGPT를 활용하여 화학공정 분야의 대표적 온톨로지인 OntoCAPE(Ontology for the domain of Computer-Aided Process Engineering)를 P&ID 검토에 특화된 지식 구조로 확장하는 체계적 방법론을 개발하고, 이를 통해 P&ID 자동 검토 시스템을 구축하여 그 실용적 가능성을 검증하고자 한다.

* 연세대학교 건축공학과 박사후연구원, 공학박사

** 연세대학교 건축공학과 연구원

*** 연세대학교 건축공학과 교수, 공학박사

(Corresponding author : Department of Architecture and Architectural Engineering, Yonsei University, glee@yonsei.ac.kr)

이 연구는 2024년도 국토교통부의 재원으로 국토교통과학기술진흥원의 지원을 받아 수행한 연구임 (RS-2024-00407028)

2. 기존 연구 고찰

2009년 제안된 OntoCAPE는 화학공학 분야의 지식 표현 및 교환을 위해 개발된 온톨로지로, 공정 설계 및 시뮬레이션 분야의 표준 참조 모델로 기능해왔다. OntoCAPE는 62개의 개별 OWL(Web Ontology Language) 모듈을 통해 플랜트의 구성요소를 위계적으로 정의하며, HAZOP(Hazard and Operability) 자동화와 같은 공정 모델링 분야에서 온톨로지 기반 지식 표현의 유효성을 입증하였다 (Jan, M., et al. 2007). 그러나 OntoCAPE는 공정의 설계 및 시뮬레이션 단계를 위해 개발되었기에, P&ID 검토와 같은 공정 운영 및 안전 관리에 필수적인 지식 체계는 포함되어 있지 않다. 또한, 2009년 이후 갱신이 안 되어, 변화하는 산업 표준 및 안전 요구사항을 적시에 반영하지 못하고 있다.

이러한 온톨로지의 동적 확장 연구는 다양한 접근법을 통해 시도되었다. 초기에는 사전에 정의된 어휘·구문 패턴을 활용하는 규칙 기반 방법론이 사용되었으며, 이후 대규모 코퍼스를 활용하는 통계적 기법과 개체-관계 추출을 자동화하는 기계학습 기반 방법론으로 발전하였다 (Asim, M.N. et al. 2018). 최근에는 온톨로지를 지식 그래프로 해석하고, 임베딩 기법을 통해 잠재적 관계를 추론하여 온톨로지를 확장하는 접근법이 시도되고 있다 (Chen, J. et al. 2025). 그러나, 이러한 기존 방법론들은 명시적인 도메인 지식에 대한 높은 의존성, 확장된 지식의 낮은 해석가능성, 급변하는 산업 환경에 대한 유연성 부족이라는 본질적 한계를 가진다.

이러한 기존 방법론들의 한계를 해결하기 위해 등장한 OntoGPT는 대형 언어모델과 SPIRES (Structured Prompt Interrogation and Recursive Extraction of Semantics) 프레임워크를 결합한 새로운 접근법을 제시한다. 이 시스템은 도메인별 추출 템플릿과 few-shot learning을 활용하여 소수의 예시만으로도 정확한 지식 추출을 수행할 수 있으며, ontology grounding 메커니즘을 통해 추출된 지식을 기존 온톨로지 스키마와 자동으로 정렬한다 (Palagin, O. et al. 2023). 더욱이 OWL reasoner와의 통합을 통해 확장 과정에서 발생할 수 있는 논리적 불일치를 사전에 탐지하고 해결할 수 있어 기존 방법론들이 가진 일관성 문제를 효과적으로 극복한다. 실제로 OntoGPT의 우수성은 생명과학 및 의료분야에서 구체적인 성과를 보여, Gene Ontology(GO) 확장에서는 precision 0.89, recall 0.84의 성능을 달성하였으며, UMLS (Unified Medical Language System) 기반 임상 온톨로지 매핑에서는 기존 규칙 기반 시스템 대비 27% 향상된 정확도를 보였다 (Caufield, J. et al. 2024). 이러한 성과는 멀티모달 처리 능력을 통해 텍스트뿐만 아니라 의료 이미지나 구조화된 데이터로부터도 일관된 품질의 지식 추출이 가능하기 때문이다.

하지만 화학공학 도메인으로의 적용에는 생명과학 분야와는 본질적으로 다른 복잡성이 존재한다. 우선 화학공정의 안전 규정은 법령적 특성상 복잡한 조건부 논리와 다층적 예외 조항을 포함하고 있어 단순한 패턴 추출로는

완전한 의미 파악이 어렵다. 또한, KOSHA (Korea Occupational Safety and Health Agency), API (American Petroleum Institute), ASTM (American Society for Testing and Materials) 등 표준 기관별로 상이한 표기법과 용어 체계가 혼재하여 일관된 지식 추출에 장애가 된다. 더 중요한 것은 화학공학 도메인 특유의 기술적 요구사항들이다. 압력, 온도, 거리 등의 수치적 제약 조건은 정확한 추출과 단위 변환이 필수적이며, 설계 중심의 기존 온톨로지를 운영 및 검토 중심으로 확장할 때는 근본적인 개념적 불일치와 스키마 진화 문제가 불가피하게 발생한다. 이러한 다차원적 도전 과제들로 인해 OntoGPT의 화학공학 분야 적용은 기존 성공 사례의 단순한 확장이 아닌, 도메인 특화적 새로운 방법론의 개발이 필요하다.

3. OntoGPT 기반 OntoCAPE 동적 확장

3.1 제안 방법

OntoCAPE를 OntoGPT에 기반한 동적 검토 온톨로지로 확장하기 위해 데이터 수집 및 지식 추출, 온톨로지 매핑, 통합 검증의 3단계 방법론을 설계하였다. 1단계에서는 국내외 안전 기준 문서로부터 P&ID 검토 관련 문장을 수집하고 전문가 검토를 통해 핵심 안전 규정을 선별하여 5개 범주로 분류된 텍스트 코퍼스를 구축한다. 2단계에서는 OntoGPT SPIRES 방법론을 활용하여 구축된 안전 기준을 기존 OntoCAPE 구조와 자동 매핑하고 새로운 검토 관련 개념을 추출한다. 3단계에서는 OntoGPT 내장 검증 기능과 SWRL (Semantic Web Rule Language) 기반 논리적 검증을 통해 확장된 온톨로지의 일관성과 완전성을 보장할 수 있도록 설계하였다. 이러한 3단계 방법론은 범용 LLM이 화학공정과 같은 고도의 전문성을 요구하는 도메인 규정도 정확히 해석하여 온톨로지로 구조화할 수 있음을 입증하였으며, 특히 규정 개정이나 신규 안전 기준 도입 시 전문가의 수동 개입 없이 온톨로지를 자동으로 확장·업데이트함으로써 지식베이스의 최신성을 유지하고 유지보수 비용을 절감한다는 점에서 실무적 가치를 갖는다.

3.2 데이터 수집 및 지식 추출

P&ID 자동 검토 시스템 구축을 위한 안전 기준 지식베이스 확보를 위해 KOSHA 기준 4종 (D-5 화학공정의 시스템 디자인 크리테리아, D-29 공정배관계장도 작성, D-37 화학설비 등의 공정설계, D-52 배관계통의 공정설계 기술 지침)과 국제 기준 3종 (API 5L, API-610, ASME B16.34)을 선정하였다. 각 문서를 대상으로 P&ID 검토와 연관성을 가지는 안전 규정, 설계 기준, 검사 요구사항이 포함된 문장을 수동 추출하여 총 258개의 후보 문장을 확보하였다. 화학공학 분야 경력 10년 이상 전문가 3명을 통해 수집된 258개를 검토하여 3명 중 2명 이상이 적용 가능하다고 평가한 문장만을 선별한 결과 중복 20개 문장을 제외한 최종 86개의 문장을 확정하였다.

확정된 86개 문장은 표 1과 같이 P&ID 검토 프로세스의 단계별 특성에 따라 AR(Annotation Review, 주석 및

국제 표준), BR(Basic Review, 기본 정보), CR(Conditional Review, 조건부 설계 규칙), DR(Design Review, 상세 설계 원칙), ER(Engineering Review, 엔지니어링 계산)의 5개 범주로 분류하였다.

표1. 단계별 P&ID 검토 기준 문장 및 예시

구분	빈도	예시
AR	24 (28.6%)	도면에는 공정 또는 지역 등을 포함하여 고유의 도면번호를 부여하여야 한다 (D-29)
BR	29 (34.5%)	배관은 호칭지름 20mm(3/4 inch) 이상을 사용한다 (D-5)
CR	10 (11.9%)	운전온도가 215℃를 넘을 경우 전기 트레이싱은 사용하지 않는다 (D-29)
DR	13 (15.1%)	벤트는 타워(Tower)나 수직 용기의 꼭대기에 설치하여야 한다 (D-5)
ER	10 (11.6%)	설계온도는 최고운전온도에 30℃를 더한 값으로 한다 (D-5)

분류된 86개 문장을 OntoGPT SPIRES 방법론에 최적화된 형태로 변환하기 위한 전처리 작업을 수행하였다. Python 환경에서 pandas 라이브러리를 활용하여 데이터 정제를 진행하고, NLTK 라이브러리를 통해 특수문자 제거, 문장 토큰화, 불용어 제거 등의 자연어 전처리를 순차적으로 적용하였다. 최종적으로 아래와 같이 각 범주별 안전 규정과 함께 출처, 적용 조건, 중요도 등의 메타데이터를 포함하는 JSON 형식의 구조화된 텍스트 코퍼스를 완성하여 온톨로지 매핑 단계의 입력 데이터로 활용하였다.

```

"data": [{
  "id": "AR_2034",
  "category": "AR",
  "category_description": "Annotation Review",
  "original_korean": "도면에는 공정 또는 지역 등을 포함하여 고유의 도면번호를 부여하여야 한다",
  "translated_english": "The drawing shall be given a unique drawing number, including a process or region.",
  "cleaned_text": "The drawing shall be given a unique drawing number, including a process or region.",
  "processed_text": "drawing shall given unique drawing number including process region",
  "tokens": ["the", "drawing", "shall", "be", "given", "a", "unique", "drawing", "number", ",", "including", "a", "process", "or", "region", "."],
  "metadata": {"source": "D-29-5.2-(2)"}
}]

```

3.3 온톨로지 매핑

기존 OntoCAPE 온톨로지에 안전 검토 기능을 통합하기 위해 OntoGPT SPIRES 방법론을 기반으로 3단계에 걸쳐 동적 검토 온톨로지를 구현하였다. 첫 번째 단계에서는 OntoCAPE의 기존 아키텍처 분석을 통해 확장가능한 영역을 식별하고 이에 적합한 SPIRES 스키마를 정의하였다. 그림 2와 같이 OntoCAPE는 ProcessEquipment를 중심으로 hasProperty, hasSubsystem, isConnectedTo 등의 관계를 정의하였다. 예를 들어, 기존 Diameter 클래스 하위에는

Outside Diameter, InsideDiameter, HoleDiameter의 3가지 직경 유형이 정의되고 있어, 86개 안전 기준 중 “배관은 호칭지름 20mm 이상을 사용한다”와 같은 규정을 처리할 때 호칭지름(Nominal Diameter)을 기존 InsideDiameter 클래스에 매핑하여 최소값 제약조건을 추가하는 방식으로 매핑 스키마를 설계하였다 (그림 1).

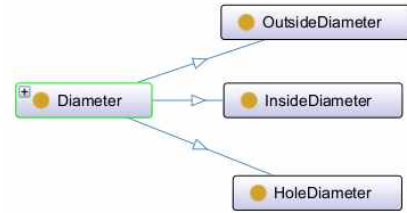


그림1. OntoCAPE 내 파이프 규격 정보

두 번째 단계에서는 정의된 스키마를 기반으로 OntoGPT SPIRES를 활용하여 지식을 추출하였다. 86개 안전 검토 기준 문장 중 AR 범주의 문서화 및 표준 참조 규정, BR의 기존 요구 사항, DR의 설계 규정은 OntoGPT가 직접 처리 가능하였으며, ER 범주의 공학 계산 규정을 포함해 총 79개 문장 (91.9%)이 성공적으로 추출되었다.

반면, 처리에 어려움을 겪은 7개 문장은 주로 CR 범주의 복합 조건문으로, “운전온도가 215℃를 넘을 경우 전기 트레이싱은 사용하지 않는다”와 같이 조건과 결과가 복합적으로 결합된 구조로 인해 OntoGPT의 단순 주어-동사-목적어 트리플 구조로 변환이 어려웠다. 이러한 문장들은 조건과 결과를 분리하는 규칙 기반 파싱을 통해 (ElectricalTracing, hasTemperatureCondition, Above215C), (ElectricalTracing, hasRestriction, Prohibited) 형태로 변환함으로 문제를 해결하였다.

세 번째 단계는 추출된 지식과 기존 OntoCAPE 관계 구조와의 통합으로, 그림 2와 같이 새로운 안전 관련 엔티티들은 기존 OntoCAPE 클래스들과 hasProperty, hasSubsystem과 같은 표준 관계를 통해 연결되었다.

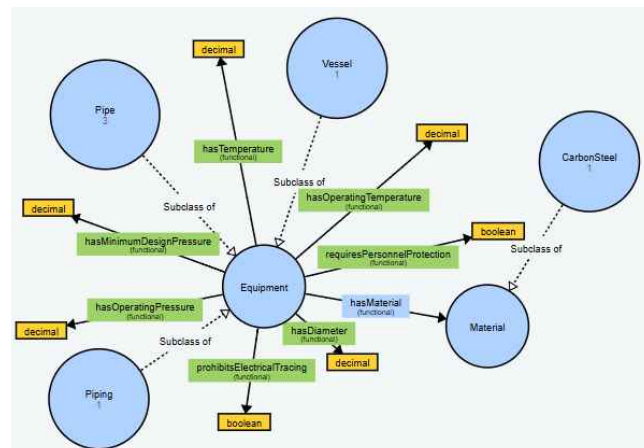


그림 2. 확장된 OntoCAPE 온톨로지 구조

추출된 지식은 기존 OntoCAPE를 안전 제약이 포함된 속성으로 확장하였다. 표 2의 내용과 같이 hasDiameter 속성은 20mm 이상 제약을 포함한 hasMinimumDiameter로, hasDesignTemperature는 '운전온도 + 30℃' 안전 마진을

반영한 hasSafetyDesign- Temperature로 확장하였다. 이를 통해 총 46개의 OntoCAPE 클래스와 44개의 RDF (Resource Description Framework) 트리플이 추가되어, 일반적인 공정 설계 온톨로지에서 안전 준수 검토가 가능한 확장된 온톨로지로 발전하였다.

표 2. 단계별 P&ID 검토 기준 문장 및 예시

	기존 OntoCAPE	확장된 OntoCAPE
직경	hasDiameter	hasMinimumDiameter
두께	hasThickness	hasThickness + 안전 마진
온도	hasTemperature	hasTemperatureSafetyLimit
설계온도	hasDesignTemperature	hasSafetyDesignTemperature (+30°C)

3.3 온톨로지 통합 및 검증

확장된 OntoCAPE 온톨로지는 그림 3과 같이 Protégé의 SWRL 규칙을 적용하여 안전 제약 조건의 논리적 검증을 수행하였다.

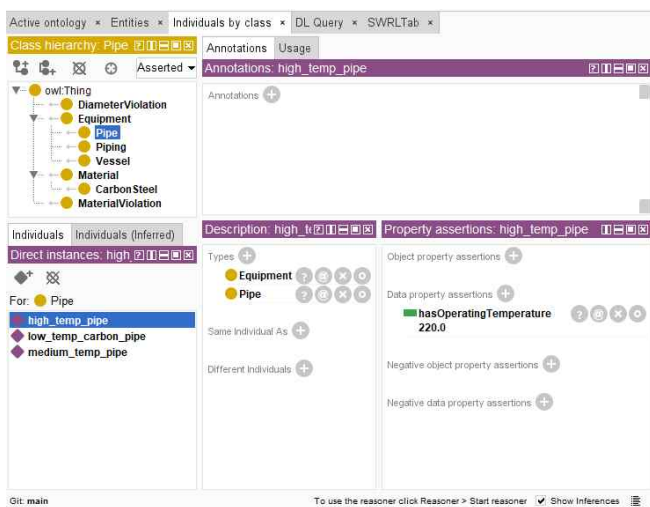


그림 3. Protégé에서의 SWRL 규칙 기반 온톨로지 검증 결과

예를 들어, Pipe, Vessel, SafetyValve 등의 안전 장비 클래스들과 hasOperatingTemperature, hasDiameter, hasMaterial 등의 속성들을 기존 온톨로지 구조와 일관성 있게 통합하였는데, 이러한 구조적 일관성 검증을 위해 다음과 같은 SWRL 규칙들을 정의하여 적용하였다:

```

온도 안전 제약:      Pipe(?p)      ^
hasOperatingTemperature(?p,      ?temp)      ^
swrlb:greaterThan(?temp,      215)      →
prohibitsElectricalTracing(?p, true)

```

```

최소 직경 검증: Piping(?p) ^ hasDiameter(?p,
?dia) ^ swrlb:lessThan(?dia, 20) →
DiameterViolation(?p)

```

그림 3은 SWRL Rules 탭에서 검토 결과로, Protégé 내 환경에서 정상적으로 실행되었으며, HermiT 추론기를 통한 추론 결과, 온도, 압력, 직경 등의 안전 제약 조건들이 논리적 모순 없이 온톨로지에 통합되었음을 확인하였다.

4. 결론

본 연구는 OntoGPT를 활용하여 화학공정 분야의 대표적 온톨로지인 OntoCAPE를 P&ID 안전 검토에 특화된 지식 구조로 확장하는 체계적 방법론을 제시하였다. 86개 한국 안전 규정 문장을 분석하여 46개 클래스 매핑과 44개 RDF 트리플을 생성하였으며, OntoGPT의 자연어 처리 능력과 규칙 기반 파싱을 결합하여 복잡한 조건문과 메타데이터를 포함한 모든 규정을 성공적으로 온톨로지화하였다. 이 연구의 핵심 기여는 LLM이 도메인 전문가의 개입 없이도 복잡한 공학 규정을 이해하고 기존 온톨로지 구조에 통합할 수 있음을 실증한 것이다. 특히 조건부 규정을 조건-결과 구조로 분리하여 처리함으로 LLM의 한계를 보완하면서도 온톨로지의 논리적 표현력을 유지하는 실용적 해결책을 제공하였다. 이러한 LLM 기반 온톨로지 확장 방법은 P&ID 설계 검토 시 안전 규정 위반을 자동으로 탐지할 수 있게 함으로써, 설계 단계에서부터 안전성을 확보하는 사전 예방적 접근을 가능하게 한다. 나아가 지속적으로 개정되는 규정을 온톨로지에 신속하게 반영할 수 있어 지식베이스의 최신성을 유지할 수 있으며, 규정 준수가 요구되는 다른 엔지니어링 분야에도 적용 가능한 범용적 프레임워크를 제공한다는 점에서 의의가 있다.

참고문헌

- Palagin, O., Kaverinskiy, V., Litvin, A., and Malakhov, K. (2023). OntoChatGPT information system: ontology-driven structured prompts for ChatGPT meta-learning, arXiv preprint arXiv:2307.05082. DOI: <https://doi.org/10.47839/ijc.22.2.3086>.
- Jan, M., Yang, A., and Marquardt, W. (2007). OntoCAPE —a large-scale ontology for chemical process engineering, Engineering Applications of Artificial Intelligence, 20(2) 147-161. DOI: <https://doi.org/10.1016/j.engappai.2006.06.010>
- Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, W., & Abbasi, H. M. (2018). A survey of ontology learning techniques and applications, Database, 2018: bay101. DOI: <https://doi.org/10.1093/database/bay101>.
- Chen, J., Mashkova, O., Zhapa-Camacho, F., Hoehndorf, R., He, Y., & Horrocks, I. (2025). Ontology embedding: a survey of methods, applications and resources, IEEE Transactions on Knowledge and Data Engineering 37(7), 4193-4212. DOI: 10.1109/TKDE.2025.3559023.
- Caufield, J. H., Hegde, H., Emonet, V., Harris, N. L., Joachimiak, M. P., Matentzoglou, N., Kim, H., Sierra Moxon, S., Reese, J., Haendel, M., Robinson, P., MungallMungall, C. J. (2024). Structured prompt interrogation and recursive extraction of semantics (SPIRES): A method for populating knowledge bases using zero-shot learning, Bioinformatics, 40(3), btac104. DOI: <https://doi.org/10.1093/bioinformatics/btac104>.