

NRC데이터정보시스템을 위한 온톨로지 개발 방향 연구

정보통신정책연구원

노희용·정용찬·신우철



NRC데이터정보시스템을 위한 온톨로지 개발 방향 연구

정보통신정책연구원
노희용·정용찬·신우철



목 차

KISDI

CONTENTS



• 국문요약	1
• Abstract	2
• 정책제언	4

제1장

연구 개요

제1 절 연구의 배경과 목적	5
1. 연구 배경	5
2. 연구 필요성	7
제2 절 연구 범위와 연구 방법	11
1. 연구 범위	11
2. 연구 방법	11

제2장

온톨로지 개념 및 사례

제1 절 온톨로지 개념 및 구성요소	12
1. 온톨로지 정의	12
2. 온톨로지 구성요소	13
3. 온톨로지 작성	16
4. 온톨로지의 범위	18

CONTENTS

제2 절	국내외 온톨로지 사례	21
1.	해외 온톨로지 사례	21
2.	국내 온톨로지 사례	23

제3장

연구프레임워크

제1 절	연구 프로세스	29
제2 절	단계별 연구 방법	30
제3 절	질 의서 및 설문지 구성	31

제4장

결 과

제1 절	온톨로지 사용자 대상 워크숍	33
제2 절	온톨로지 개발자 대상 설문조사	40
제3 절	소결	47

제5장

결 론

1.	결론 및 시사점	48
2.	한계점 및 해결 방안	49
•	참고문헌	50
•	부록	54

표 목 차



〈표 1-1〉 「데이터 기반 미래예측·정책지원 모델연구」 이력	6
〈표 1-2〉 classification, taxonomy, ontology 구분	7
〈표 2-1〉 온톨로지 구성요소별 용어 정리	15
〈표 2-2〉 온톨로지 언어 종류	16
〈표 2-3〉 대표적인 온톨로지 언어	17
〈표 2-4〉 대표적인 상위 온톨로지	20
〈표 2-5〉 해외 온톨로지 사례	22
〈표 2-6〉 국내 온톨로지 사례	23
〈표 2-7〉 국내 온톨로지 구축 연구 사례	27
〈표 3-1〉 워크숍 진행 대상	30
〈표 3-2〉 설문조사 진행 대상	31
〈표 3-3〉 워크숍 질의서	31
〈표 3-4〉 설문조사지 구성	32
〈표 4-1〉 pykeen 패키지 구성	35
〈표 4-2〉 첫 번째 설문지 질의에 대한 전문가 의견의 공통점 및 차이점	40
〈표 4-3〉 두 번째 설문지 질의에 대한 전문가 의견의 공통점 및 차이점	41
〈표 4-4〉 세 번째 설문지 질의에 대한 전문가 의견의 공통점 및 차이점	41
〈표 4-5〉 네 번째 설문지 질의에 대한 전문가 의견	42
〈표 4-6〉 네 번째 설문지 질의에 대한 전문가 의견의 공통점 및 차이점	43
〈표 4-7〉 다섯 번째 설문지 질의에 대한 전문가 의견	44
〈표 4-8〉 일곱 번째 설문지 질의에 대한 전문가 의견	45
〈표 4-9〉 여덟 번째 설문지 질의에 대한 전문가 의견	46

그림목차



[그림 2-1] 온톨로지 예시	14
[그림 2-2] 순환 구조를 갖는 '피자' 온톨로지 예시	14
[그림 2-3] 온톨로지 범위에 따른 구분	19
[그림 3-1] 연구 프로세스	29
[그림 4-1] LoG conference website	34
[그림 4-2] paperswithcode 내 검색 결과	39
[그림 4-3] 온톨로지 개발 및 운영 주체 간 협력 흐름	42

국문요약

본 연구는 본래 정보통신정책연구원의 「ICT데이터플랫폼 구축 운영사업」을 통해 개발한 온톨로지 개발 과업이 시작이다. 예산상의 이유로 온톨로지 개발 과업이 중단되었지만, 「데이터 기반 미래예측·정책지원 모델연구」에서 유용하게 속행할 수 있는 잠재력이 있는 것으로 판단되어 이를 검토하는 목적으로 본 연구가 수행되었다.

2장에서는 문헌 연구를 통해 온톨로지의 기본 개념 및 구성요소에 대해 정리했다. 이를 통해 온톨로지 개발 방향 논의를 위한 배경지식을 확보할 수 있었다. 또한 사례 분석을 통해 국내외 최근 온톨로지 개발 동향을 확인했다.

3장에서는 연구 수행을 위한 프레임워크를 기술했다. 첫 번째 단계에서는 2장에서 정리한 문헌 연구를 통해 연구 방향을 설계했다. 이후 두 번째 단계인 온톨로지 사용자 전문가 그룹과의 인터뷰를 통해 (i) 사용자 입장에서 실효성 높은 온톨로지의 요건을 살펴보고, (ii) 온톨로지 개발자 그룹을 대상으로 하는 설문지를 작성하는데 필요한 정보를 확보하는 것을 목표로 한다. 세 번째 단계인 온톨로지 개발자 전문가 그룹과의 설문조사에서는 온톨로지 개발 방향 수립에 필요한 실질적인 정보를 얻고자 했다. 마지막 단계에서는 앞서 문헌 연구, 온톨로지 사용자, 온톨로지 개발자 그룹으로부터 수집한 정보를 바탕으로 온톨로지 개발 의사결정 및 방향 수립을 위한 정성 분석을 수행한다. 이와 같은 프레임워크 설계를 통해 체계적인 연구를 수행할 수 있도록 했다.

4장에서는 온톨로지 사용자 그룹과 온톨로지 개발자 그룹의 구성 및 워크숍과 설문조사 수행 결과를 기술한다. 두 가지 그룹 각각에 대한 질의는 달랐지만, 결과적으로 결이 같은 답변을 확인할 수 있었다.

마지막 5장에서는 2장과 4장의 정보를 바탕으로 온톨로지 개발 방향을 수립했다. 결론적으로는 공공연구기관과 연구개발 예산의 특성상 작은 목표들의 설정 및 이를 달성할 수 있는 특수목적형 온톨로지를 개발하고, 이를 통해 활용 및 성공 사례를 축적하는 방향이 선택되었다. 단, 이렇게 개발하는 온톨로지는 개방성과 확장성을 염두에 두기에 순차적인 확장을 통해 상위 온톨로지와 연계할 지향한다.



Abstract

This study is based on the ontology development task originally initiated through the “ICT Data Platform Construction and Operation Project” by the Korea Information Society Development Institute. Due to budget constraints, the ontology development task was suspended, but it was determined that it holds potential to be continued usefully in the “Data-driven Future Prediction and Policy Support Model Research.” Therefore, this study was conducted with the aim of reviewing its potential.

Chapter 2 organizes the basic concepts and components of ontology through a literature review. This provides the background knowledge necessary for discussing the direction of ontology development. Additionally, through case analysis, the study examines recent trends in ontology development both domestically and internationally.

Chapter 3 outlines the framework for conducting the research. In the first phase, the research direction was designed based on the literature review conducted in Chapter 2. In the second phase, interviews with the ontology user expert group were conducted to (i) examine the requirements for a practical ontology from the user’s perspective and (ii) gather necessary information for creating a questionnaire for the ontology developer group. The third phase involved a survey with the ontology developer expert group to gather practical information needed for establishing the direction of ontology development. The final phase involved qualitative analysis for decision-making and direction-setting of the ontology development, based on the information collected from the literature review, the user group, and the developer group. This structured framework was designed to ensure systematic research.

Chapter 4 details the composition of the ontology user and developer groups, as well as the results of the workshops and surveys. Although the questions posed to each group were different, the responses ultimately showed similar patterns.

Finally, Chapter 5 establishes the direction for ontology development based on the information from Chapters 2 and 4. In conclusion, due to the characteristics of public research institutions and research and development budgets, it was decided to set small, achievable goals and develop a specialized-purpose ontology that could accumulate successful use cases. However, this ontology will be developed with openness and scalability in mind, aiming for gradual expansion and integration with upper-level ontologies.

This report is the result of the fourth year of predictive model development research to support the anticipatory policy in the field of ICT and gender equality policy using text data stored in the NRC Data Information System and external data.

Chapter 2, A Study on the Improvement of the ICT Regional Employment Forecasting Model expands and enriches the contents of the previous study. This study aimed to evaluate and enhance the predictive power of the employment outlook, estimate and combine the employment in the ICT manufacturing industry, and expand the ICT employment index and discover implications. First, as a result of the various models, there was no model that showed an absolute advantage in the forecasts for ICT employment by region. However, we found that the predictive power of the models including exogenous information(country, region, and industry) or news data-based indices generally improved. Second, considering the employment of the ICT service(J) and the ICT manufacturing(C), the overall employment of the ICT industry was estimated.. Compared to the employment of ICT service(J) concentrated in the metropolitan area, that of the ICT manufacturing(C) was relatively widely distributed in the provinces. And in terms of outlook, the difference in employment between regions by ICT industry was identified. Finally, the implications of the ICT employment index were examined and we extended it to the weekly index.

정책제언

생성형 인공지능 기술이 주목받는 요즘, 지식 구조화를 가능하게 하는 온톨로지 개발의 필요성 또한 함께 증가하고 있다. 다만 국문으로 작성된 온톨로지는 그 수도 매우 적고 접근이 어려운 형편이다. 한글 사용자의 규모가 계속해서 증가하는 것과는 별개로, 사실상 글로벌 표준이 되는 거대 온톨로지 개발에서 국문 활용은 제한적이다. 본 연구에서는 이와 같은 상황에서 국내 디지털 역량 강화를 위한 데이터 자원인 온톨로지를 공공연구기관에서 개발할 때 어떤 실효성이 있는지를 확인하고자 했다.

온톨로지의 개발은 시간적·물질적 자원이 많이 소요되는 프로젝트로, 실제 유의미한 상위 온톨로지 개발을 위해서는 막대한 자원 투입이 요구된다. 도메인 전문가라고 하더라도 온톨로지 사용이 낮은 국내 상황에서는 이와 같은 막대한 자원 투입이 사실상 요원한 일이다. 본 연구에서 접촉한 전문가들은 이러한 상황 타개를 위해서 온톨로지 개발 프로젝트를 공공 지원으로 시작해서 민간에게 이전하고, 산업계의 활용까지 장려할 수 있는 순환 구조가 마련되어야 할 것이라고 이야기한다.

이를 위해서는 작은 수의 성공 또는 활용 사례가 꾸준히 축적되어야 하며, 처음부터 큰 규모의 온톨로지 구축에 초점을 맞추기보다 작은 규모더라도 확실한 활용처가 있는 온톨로지 개발이 적합할 것이라는 결론을 내릴 수 있었다.

제1장

연구 개요

제 1 절 연구의 배경과 목적

1. 연구 배경

본 연구는 정보통신정책연구원이 주관하는 미디어통계포털(stat.kisdi.re.kr)의 「ICT데이터플랫폼 구축 운영사업」에서 비롯한다. 2022년부터 2023년까지 2년간 미디어와 정보통신기술(ICT, Information and Communication Technology) 분야의 온톨로지 구축 연구를 선행적으로 진행한 바 있고, 이를 통해 정보통신정책연구원 보고서를 대상으로 하는 LOD(Linked Open Data) 기반의 통계 온톨로지를 개발했다. 당시 온톨로지 개발의 목적은 정보통신정책연구원의 출판물에서 사용된 통계 데이터의 연계로 (i) 미디어통계포털의 사용자 편의성을 증대하고, (ii) 이를 기반으로 ICT 영역의 텍스트 데이터 분석을 강화할 수 있는 데이터 인프라 제공이었다. 이는 미디어통계포털의 사용자 확대를 기본적으로 염두에 두었으나, 궁극적으로는 공공연구기관으로서 증거 기반 정책 수립에 이바지하기 위함이었다. 사업 예산 축소로 인해 중단되긴 했으나, 개발 과정 및 결과물은 인공지능 시대에서 유용한 지식 그래프(knowledge graph) 구축의 실마리를 제공하는 기회로 작용했다.

「ICT데이터플랫폼 구축 운영사업」과는 별개로 정보통신정책연구원에서는 2024년부터 「데이터 기반 미래예측·정책지원 모델연구」 또한 수행 중이다. 해당 연구는 정보통신정책연구원이 주최하는 신규 사업이지만, 연구의 연속성 관점에서는 2019년부터 2023년까지 경제인문사회연구회(NRC, National research council for economics, humanities and social sciences) 주최 및 정보통신정책연구원 주관으로 진행된 동명 사업의 후속 연구에 해당한다(〈표 1-1〉 참고). 사업의 주최권자가 경제인문사회연구회에서 정보통신정책연구원으로 넘겨진 모양새다. 「데이터 기반 미래예측·정책지원 모델연구」는 경제인문사회연구회 소관 연구기관을 대상으로 데이터 및 분석 환경을 제공하는 플랫폼인 NRC데이터정보시스템(nrcdata.re.kr)을 기반으로 영역별(e.g.,

정보통신, 법제, 여성정책 등) 데이터 분석 서비스를 통한 정책제언을 지향한다. 즉, 「데이터 기반 미래예측·정책지원 모델연구」는 (i) 연구용 데이터 축적 및 공유와 (ii) 데이터 분석 환경 제공을 위한 플랫폼을 운영·유지보수하는 동시에 (iv) 정책제언을 위한 데이터 분석 모델의 개발 및 서비스를 포괄하는 과제로 요약할 수 있다.

〈표 1-1〉 「데이터 기반 미래예측·정책지원 모델연구」 이력

구분	사업 수행 연도					
	2019년	2020년	2021년	2022년	2023년	2024년
예측모형 개발	1) ICT정책 지원을 위한 빅데이터 분석 및 예측 모형 개발	1) 국가사회 발전지수 및 ICT·보건복지 정책 지원을 위한 미래 예측 모형 개발	1) 데이터기반 미래예측·정책지원 모델연구Ⅱ	1) 데이터기반 미래예측·정책지원 모델연구Ⅲ	1) 데이터기반 미래예측·정책지원 모델연구Ⅳ	1) 데이터 기반 미래예측·정책지원 모델 연구
인프라 구축	1) 데이터 기반 미래예측·정책지원 시스템 구축 ISP 2) 경제·인문사회연구회 데이터 기반 미래예측·정책지원 사업 추진 전략 연구	1) 데이터 기반 미래예측·정책지원 사업 플랫폼 구축 부문 용역	1) 데이터 기반 미래 예측 정책지원 사업 데이터 플랫폼 구축 고도화 부문 용역			1) NDIS 서비스 및 시스템 이관·운영 용역
주최기관	경제인문사회연구회					정보통신정책연구원
주관기관	정보통신정책연구원					

자료: 저자 작성

NRC데이터플랫폼이 연구용 데이터의 축적 및 분석 환경을 제공하고, 온톨로지와 같은 데이터 자원이 향후 분석 모델 개발에 활용해 모델 성능을 제고할 수 있다는 점에서 「ICT데이터플랫폼 구축 운영사업」에서 중지한 온톨로지 개발은 「데이터 기반 미래예측·정책지원 모델연구」에서 지속할 수 있는 잠재 가치가 있을 것으로 판단된다. 다만, 「ICT데이터플랫폼 구축 운영사업」과 「데이터 기반 미래예측·정책지원 모델연구」 각각의 플랫폼 사용자가 다르고 지향하는 바가 다르다는 위험성 또한 내포하고 있기에, 이에 대한 면밀한 검토가 필요한 상황이다.

2. 연구 필요성

Van Rees(2003)는 분류(classification), 분류체계(taxonomy), 온톨로지(ontology) 각각의 용어 활용이 2000년대 초반까지도 명료하게 구분되지 않고 혼용되고 있는 문제점을 제기한 바 있다. 이러한 용어 활용의 모호함을 개선하기 위해 EBEW-conference 2001에서 위 세 용어를 활용한 논문을 분석했고, 비록 명확한 기술이 없더라도 각 용어의 정의와 차이점에 대해 <표 1-2>와 같이 구분할 수 있는 어떤 합의적 수렴이 이루어지고 있음을 주장했다.

<표 1-2> classification, taxonomy, ontology 구분

구분	개념
분류 (Classification)	“단순 분류”. 특정한 외부 기준에 의한 개체들의 묶음으로, 이러한 묶음은 특정한 관점에 의해만 들어지기 때문에 매우 자연스럽다. 분류는 기본적으로 레이블이 붙은 상자 세트로 볼 수 있고, 각 상자의 레이블에 따라 특정 개체들을 분류할 수 있는 것이다. 사용자 친화적인 관점에서 분류체계나 온톨로지를 편하게 부르기 위해 포괄적인 개념으로 사용되는 경향이 있다.
분류체계 (Taxonomy)	“분류체계” 또는 “단순한 온톨로지”. 분류체계에 내재된 데이터에 따른 계층적인 묶음을 의미한다. 단순한 온톨로지라는 의미로 사용될 경우, 분류체계의 계층구조(hierarchy)는 하위 클래스(sub class) 계층을 기반으로 구성한다.
온톨로지 (Ontology)	- 특정 도메인을 설명하는 잘 정의된 개념들의 집합. 이러한 개념들은 하위 클래스 계층을 통해 정의되며, 속성의 할당 및 정의, 개념 간의 관계 정의 등 다양한 방법으로 구성될 수 있다. ‘온톨로지’라는 용어는 명확한 설명이 필요한데, 단순한 온톨로지는 ‘분류체계’로 명명하는 것이 나을 수 있지만 고도화된 온톨로지는 그 기능을 명확히 하여 대다수의 단순한 온톨로지와의 구별이 필요하다. - 온톨로지의 목표는 소통에서 사용할 수 있는 공통의 참조 가능한 개념 집합을 제공하는 것으로, 서로 다른 도메인에 적합한 개념들을 제공하는 여러 온톨로지를 함께 사용하는 것이 일반적이다.

자료: Van Rees(2003)

Van Rees(2003)의 정의를 간략하게 요약하면, 결국 분류(classification)는 분류를 위한 기준 혹은 분류되는 개체 간 관계의 복잡성이나 체계성과 관계없이 포괄적으로 아울러 부를 수 있는 용어이며, 분류체계(taxonomy)는 계층적인 관계가 형성되는 분류를 의미한다. 온톨로지는 단순히 ‘구분하는’ 분류체계와 달리 더욱 고도화한 개념으로 크게는 외부 도메인까지 고려하고 개체 간 관계까지 포괄한다는 점에서 내포하고 있는 정보가 더욱 많다고 할 수 있다. 더욱 직접적인 예시로 ‘분류체계’는 [생물 - 동물 - 포유류 - 인간 - 남자/여자] 등을 통해 ‘철수’와 ‘영희’를 지구상의 생물군과 비교해서 나눌 수 있는 계층구조 그 자체라고 할 수 있다. 여기서 ‘온톨로지’는 ‘철수’와 ‘영희’의 구분에서 그치지 않고 남자-여자 간 관계가 부부 관계, 동료 관계 등의 속성을 지니고 있을 때 ‘철수’와 ‘영희’ 각각이 갖는 개체 정보를 통해 서로가 부부 관계인지 타인인지 등까지도 확인할 수 있는, 혹은 다른 속성 정보(e.g., 경제 공동체, 합의, 자녀 유무 등)를 통해 추론까지도 가능한 정보를 포함한다.

인간의 과학적 사고가 자연을 이해하기 위해 정교하게 다듬어진 분리 능력에서 기반하고 있음을 고려하면, 지식을 축적하고 정보를 활용하는데 정밀한 분류체계의 유무가 미치는 영향력은 매우 크다. 동시에 인간은 무리 지어 공동의 가치를 추구하기 때문에 복잡하고 진보한 현대 사회에 와서도 조직에 내재한 공동 지식의 관리를 위한 분류체계의 중요성은 더욱 증가하고 있는 경향이 있다. 그뿐만 아니라, 디지털 기술과 정보 시스템이 사회 전반에 녹아 들어가면서 인간의 지식을 기계에 전달하고 조직간 협업을 위한 명료한 분류체계에 대한 요구가 함께 증가하고 있다.

이러한 문맥에서 온톨로지의 필요성과 중요성에 대해서는 다양한 민간기업의 비전 및 목표 등에서도 잘 드러난다. 예를 들어, 세계에서 가장 널리 사용되며, 사용자 친화적인 그래픽 사용자 인터페이스를 제공하는 온톨로지 편집기 프로티지(Protégé)는 온톨로지 개발의 필요성에 대해 다음과 같이 네 가지 이유를 들고 있다(Noy & McGuinness, 2001). 첫 번째는 정보 구조에 대한 공통 이해의 공유이다. 특정 도메인 또는 학문 분야에 대한 여러 가지 종류의 데이터 세트 또는 웹사이트 등이 각각의 온톨로지를 공유하고 게시한다면, 특별한 표준화 작업 없이 단일 문맥(context)에서 정보 집계를 진행할 수 있게 된다. 즉, 표준화에 관한 내용이며, 이는 특히 다양한 데이터·정보 생산자가 나타나는 디지털 시대에 매우 중요한 사항이라고 볼 수 있다. 두 번째는 도메인 지식의 재사용을 가능하게 하기 위함이다. 만약 새로운 발견 또는 발전한 기술로 도메인 지식이 변경될 경우, 온톨로지를 기반으로 사용하지 않은 하드 코딩(hard coding)¹⁾ 결과물은 순식간에 효용성이 사라지고 변경 또한 쉽지 않다. 그러나 온톨로지를 활용한다면 변경된 도메인 지식을 온톨로지에 적용하는 것으로 기존 결과물의 재사용을 수월하게 한다. 세 번째는 도메인 지식을 운영 지식을 분리하는 것으로, 운영 지식의 타 도메인 활용이 가능하게 한다. 예를 들어, 특정 고객 요구 사양에 따라 제품을 구성하는 작업을 설명하는 서비스가 있을 때, 이를 컴퓨터 온톨로지나 엘리베이터 온톨로지에 각각 적용하면 우리는 컴퓨터 주문 서비스와 엘리베이터 주문 서비스를 모두 기대할 수 있다. 네 번째는 온톨로지 활용에 목적을 두지 않고, 도메인 지식의 분석과 구조화 자체를 위함이다. 주로 학술 연구 또는 공공의 목적을 위해 진행된다.

팔란티어 테크놀로지스(Palantir Technologies)는 주로 공공 정보 분석을 수행하는 미국의 빅데이터 프로세싱 기업으로, 특히 미국 국방부, 중앙정보국, 연방수사국 등에서 투자와 일감을 받는 등 공공 부문에서 입지가 튼튼하다. 팔란티어에서도 온톨로지 개발 및 유지·보수에 적극적인 투자를 수행하고 있는데, 이들이 설명하는 온톨로지 개발의 필요성은 크게 다섯 가지로 구분할 수 있다²⁾. 첫 번째는 온톨로지를 규모 확장성을 갖춘 연결성 확보 도구로 활용하는 것이다. 이는

1) 프로그램을 개발할 때 하드 코딩은 키 파라미터나 변수의 값을 고정하는 방식을, 소프트 코딩은 그 값을 가변적으로 변경할 수 있게 작성하는 것을 의미한다. 소프트 코딩이 파라미터나 변수의 가변성에 대한 강건성을 고려해야 하는 어려움은 있지만 환경이나 상황의 변화에 값을 바꾸는 것으로 유연하게 대처할 수 있다는 장점이 있다.

2) Palantir, "Why creat an Ontology?",

<https://www.palantir.com/docs/foundry/ontology/why-ontology/>(2024. 12. 10. 접속)

조직 전반의 의사결정을 위한 신뢰할 수 있는 단일 정보 원천을 온톨로지 형태로 제공함으로써 조직 내 사용 가능한 데이터를 쉽게 정리하고 이해할 수 있기 때문이다. 이는 개별적인 의사결정을 전체적인 맥락에서 볼 수 있게 되어 규모 확장성을 갖춘 연결성을 기대할 수 있게 한다. 예를 들어, 표준 데이터 레이크(data lake)를 운영할 때는 필연적으로 시간이 지날수록 데이터 세트, 대시보드, 애플리케이션의 증가로 복잡성이 증가한다. 이때 온톨로지를 활용하면 시간이 지나 데이터 자산의 규모가 성장하더라도 대규모 전환을 가능하게 하면서도 복잡성을 제어하고 데이터 관리의 어려움을 줄일 수 있다. 두 번째는 의사결정자의 정보 해석의 정확성 향상이다. 어떤 조직을 데이터 중심으로 운영할 때 가장 어려운 요소 중 하나는 데이터를 다양한 조직 내 의사 결정자에게 전달하는 것이며, 많은 의사결정자는 코드나 데이터 세트, 조인(join)과 같은 정보기술 개념에 익숙하지 않다. 온톨로지는 이러한 디지털 개념을 추상화하여 사용자들이 이상적으로 사용하는 표준 용어로 데이터를 다룰 수 있게 한다. 서로 다른 배경지식을 갖춘 사용자들에게 공유 언어를 제공함으로써 모든 사용자가 같은 정보를 보고 있다는 사실을 확인하기 위한 조정 과정을 피하고 협업할 수 있도록 해주는 것이다. 세 번째는 단일 재사용 데이터 자산의 활용으로 인한 규모의 경제 달성이다. 새로운 데이터 통합이 필요한 경우를 제외하고는 온톨로지를 기반으로 애플리케이션과 사용 사례를 구축할 수 있으며, 공유 데이터 자산은 애플리케이션 구축자가 데이터 정리에 집중하는 대신 조직 문제에 집중할 수 있게 해준다. 네 번째는 데이터 기록 및 지속적인 개선을 위한 관련 의사결정 기록이다. 온톨로지서 데이터 편집 및 확장에 관한 의사결정 결과를 기록함으로써, 한 사용자가 포착한 인사이트가 다른 사용자의 의사결정에 기여할 수 있어 데이터 자산의 가치를 시간과 함께 증가시킨다. 마지막은 기계학습 및 인공지능 알고리즘 활용의 보완을 위함이다. 일반적으로 인공지능 텔은 블랙박스이기 때문에 결과 도출의 논리에 대한 검증이 어려워, 의사결정에 사용하기 어렵다. 이때, 도메인 지식을 표현하는 온톨로지를 함께 활용한다면, 이에 대한 추론적 보완이 가능하다. 실제로 온톨로지는 설명 가능한 인공지능(XAI, eXplainable Artificial Intelligence) 기술에서 중심적인 임무를 수행할 수 있고, 인공지능 모형의 투명성을 강화하는 데 또한 필요하다.

이와 같은 온톨로지의 막대한 장점에도 불구하고, 실제 성공적인 온톨로지 개발 결과물을 확인하기는 쉽지 않다. 특히, 한국어 기반의 온톨로지 개발에 관한 연구는 다양한 학문 분야 및 도메인에서 쉽게 찾을 수 있지만, 활용도가 높은 대표적인 온톨로지 사례를 찾기는 어렵다. 이는 온톨로지 개발의 고질적인 단점 또는 높은 잠재적 위험에서 비롯되는데, 크게 다섯 가지로 요약할 수 있다. 첫 번째로 온톨로지의 복잡성에 따른 구현의 부담이다. 온톨로지를 개발하고 유지하는 것은 상당한 전문성과 자원을 요구하며, 복잡하고 노동 집약적이며 시간 소요가 크다. 이 때문에 온톨로지의 지속적인 유지 보수는 결국 상당한 비용 부담을 초래할 수 있다. 두 번째는 경직성으로 도메인 분야의 지식 변화가 잦거나 큰 변혁이 있을 때, 초기 구축과 유사한 수정 작업 및 부담이 요

구된다는 점이다. 세 번째는 잘못된 해석 가능성으로, 온톨로지 구축에 활용된 용어가 명확하게 정의되지 않거나 합의가 없으면 큰 불일치와 혼란을 불러일으킬 수 있다. 특히 새로운 용어의 등장이나 지식의 변형에 대해 다양한 관점이 존재할 수 있을 때 극대화될 수 있다. 네 번째는 표현력의 한계로 인해 일부 온톨로지가 지식 도메인의 전체적 복잡성을 담아내지 못한다면 오히려 투자 대비 활용성이 낮거나 중요한 뉘앙스가 생략될 수 있다는 점이다. 마지막으로 온톨로지가 구축되는 도메인에 따라 해당 도메인에 대한 깊은 지식과 이해를 마련하기 어렵거나 관련 지식의 확보가 쉽지 않을 수도 있다는 점이다.

결과적으로 온톨로지의 개발은, 특히 경제·인문 사회 영역의 넓은 도메인 지식 범위를 다루기 위해 다양한 데이터와 분석 애플리케이션을 담고자 하는 NRC데이터정보시스템에서는 장기적인 안목에서 그 실효성은 분명하다. 다만, 동시에 넓은 범위의 다양한 도메인 지식을 고려해야 하는 만큼 온톨로지 개발의 어려움 역시 크게 작용할 수 있어 고민이 이에 대한 고민이 필요하다. 더욱이 단년도 실적이 요구되고, 사업 예산의 안정성에 대한 담보가 되지 않는 연구개발 과제 특성상 온톨로지 개발의 여부 또는 방향에 대한 논의는 반드시 사전에 수행되어야 한다.

제 2 절 연구 범위와 연구 방법

1. 연구 범위

본 연구는 「데이터 기반 미래예측·정책지원 모델연구」 사업으로 온톨로지를 개발할 때의 기대 효과와 제약사항 그리고 개발 진행 여부를 결정하기 위한 것이다. 따라서 본 연구의 범위는 온톨로지 관련 배경지식 정리, 국내외 사례 조사, 온톨로지 개발자와 사용자 의견 청취로 한정된다. 이를 통해 온톨로지 전반에 대한 개념을 이해하고, 공공연구기관으로서 데이터 자원 개발의 득과 실을 확인한다.

2. 연구 방법

본 연구는 문헌조사와 더불어 설문조사를 활용한 정성 분석을 주된 연구 방법으로 한다. 설문조사를 활용함에도 정량이 아닌 정성 분석을 수행하는 까닭은 크게 두 가지다. 먼저 온톨로지를 개발하거나 사용하는 이는 대부분 전문가로 극히 한정적인 수에 불과해, 정량분석을 수행할 만큼의 대표성 있는 표본 추출 등이 불가능하다. 두 번째로 온톨로지 개발의 여부나 효용 및 제약사항의 파악을 위해서는 온톨로지 개발 경험 혹은 관련 배경지식을 바탕으로 하는 실무적 통찰력이 필요한데, 이는 정량분석보다 정성 분석을 통한 사례별 이해가 더욱 효과가 크다.³⁾

본 연구에서는 크게 온톨로지 개발자와 사용자를 구분해 연구를 수행한다. 온톨로지 개발자를 대상으로는 서술형의 설문조사를, 온톨로지 사용자를 대상으로는 워크숍을 통해 본 연구의 목적을 달성하고자 했다.

3) 온톨로지 개발의 동인, 활용 목적, 구조 등이 사례별로 각기 다르기 때문이다.

제2장

온톨로지 개념 및 사례

제 1 절 온톨로지 개념 및 구성요소

1. 온톨로지 정의

본래 온톨로지는 철학에서 시작된 개념으로 모든 실재의 본질을 다루었지만 1967년 G. H. Mealy에 의해 전산학 분야에 도입된 이후 최근에는 지식공학, 지식 표현, 정보 모델링, 정보관리, 지식 표현과 활용을 연구하는 AI 분야까지 다양한 분야에 적용되고 있다. 연구자 각각은 온톨로지에 대해 다양한 언어와 표현으로 정의를 내리고 있으며, 비록 각각의 관점은 다르더라도 ‘온톨로지’의 개념에 대한 보편적인 합의는 이루어져 있는 상황이다.

온톨로지(ontology)를 철학적 또는 어학적으로 접근하면 존재의 본질(the nature of existence)에 관해 연구하는 존재론(i.e., ontology)과 지식의 본질(the nature of knowledge)에 관해 연구하는 인식론(i.e., epistemology)의 비교를 기술하지 않을 수 없다(cf. Noh et al., 2018). 철학적으로는 다양한 관점이 논의되고 있지만, 쉽게 말하면 존재론과 인식론은 각각 “세상은 무엇으로 구성되어 있는가?”와 “세상에 대해 무엇을 어떻게 알 수 있는가?”에 대한 관심으로 축약할 수 있다. 이를 기반으로 하면, 통념적으로 온톨로지라고 칭하는 개념은 “어떤 특정한 것(e.g., 분야 또는 지식 도메인 등)의 구성요소 및 관계”를 정의하고 있을 것임을 유추할 수 있다.

기존 연구에서는 학자들 각각의 표현도 확인할 수 있다. Sowa(1992)는 온톨로지를 “어떤 특정 영역에 존재하거나 존재할 수 있는 대상들의 범주”로 정의하며, Gruber(1993)는 “형식적(formal)이고 명시적(explicit)인 공유된(shared) 개념”으로 정의하였다. 또한, 온톨로지는 추상적 개념들이 실세계의 대상이나 사물들과 어떻게 관련되는가에 대한 전산적 단위이자 자원으로 세상에 대한 지식을 제공하는 데이터베이스를 구성하는 것(윤진혁·유상봉·김인한, 2003)이라는 의견도 있다. 또는 어떠한 개념들(범주들)이 실제 세상에 어떤 도메인에 존재하고 있고, 어떠한 속

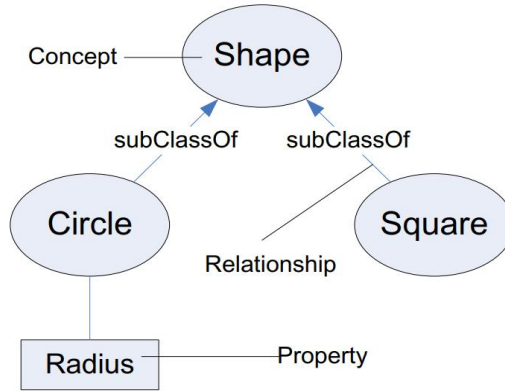
성(property)들을 지니고 있으며, 어떻게 서로 연결되어 있는가를 보여주는 지식체계로, 관련 연구들에서 온톨로지는 사람과 컴퓨터 또는 컴퓨터 간의 개념 및 개념 표현을 공유하기 위한 개념화의 명시적 규약으로 정의되기도 하며(황금하·신지애·최기선, 2008), 사람이 공유하는 개념을 컴퓨터가 이해할 수 있는 형식으로 정의 및 규정하는 것으로 표현되기도 한다(고영만, 2006).

즉, 온톨로지는 특정 영역의 개념과 용어 간의 관계를 정의하는 명세(노상규·박진수, 2007)라고 말할 수 있으므로, 기계가 이해할 수 있어야 하고(형식적), 개념에 대한 의미가 명확하게 정의되어야 하고(명시적), 온톨로지 체계에 대한 합의가 이루어져야 하며(공유된), 추상적 모델을 가져야(개념화) 한다. 정보과학 측면에서 온톨로지는 특정한 영역을 표현하는 데이터 모델로서 특정한 영역(Domain)에 속하는 개념과 개념 사이의 관계를 기술하는 형식적이고 공식적인 어휘의 집합으로 정의된다. 따라서, 온톨로지는 사람이 이해하고 있는 용어와 개념, 이들 간의 관계를 정의하여 프로그램과 인간 간의 지식 공유에 도움을 주는 것에 그 목적이 있는 것으로 이해할 수 있다(유기동·황현석, 2012).

2. 온톨로지 구성요소

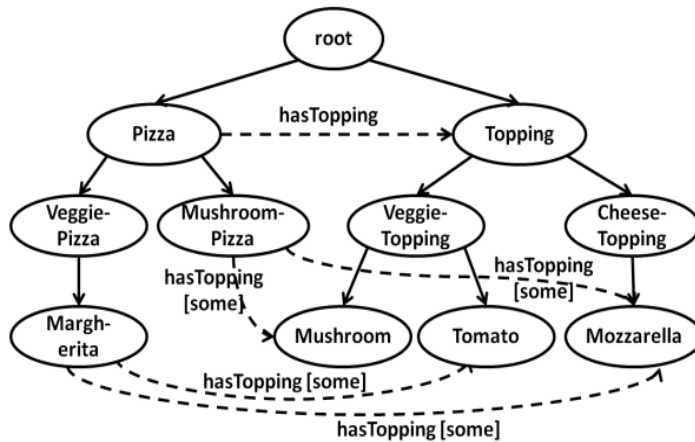
온톨로지의 구조 및 구성요소에 대해서 반드시 지켜야 하는 사항이 존재하지는 않는다. 개발하고자 하는 온톨로지를 구축하는 지식 도메인이나 온톨로지를 활용하고자 하는 목적에 따라 온톨로지의 구조나 구성요소는 얼마든지 바뀔 수 있다. 그러나 대부분의 온톨로지는 설명하고자 하는 어떤 ‘무엇’을 중심으로 위계적 트리 구조(tree structure)를 보여주는데, 간단한 예시는 [그림 2-1]과 같다. 그림을 보면 ‘형태(shape)’라는 개념(concept)을 설명하기 위한 하위 그룹(subclass)으로 ‘원(circle)’과 ‘네모(square)’라는 인스턴스(instance)가 있고, ‘원’의 속성(property)으로 ‘반경(radius)’을 확인할 수 있다. 개념-인스턴스-속성 간의 관계는 각 개체(entity)를 연결하는 선으로 묘사할 수 있는데 이 선은 각 개체 간의 관계를 의미한다.

[그림 2-1] 온톨로지 예시



자료: Timm & Gannod(2005)

[그림 2-2] 순환 구조를 갖는 ‘피자’ 온톨로지 예시



자료: Kim et al.(2012)

온톨로지는 개체 간 관계(relationship) 또한 포괄하기에 상위 개념을 설명하기 위한 복수의 하위 그룹은 대체로 트리 구조로 나타나지만, 관계의 설정에 따라 [그림 2-2]와 같이 순환 구조가 나타날 수도 있다.

Bergman(2018)에 의하면, 온톨로지를 구성하는 요소들은 연구자에 따라 또는 온톨로지 관련 분야에 따라 각기 다른 용어(terminology)로 표현되는데, 각각의 정의는 때에 따라 다를지언정 대체로 <표 2-1>과 같이 유사한 용어들을 묶을 수 있다.

〈표 2-1〉 온톨로지 구성요소별 용어 정리

구분	유사하게 활용되는 용어
개체 종류 (Entity type)	concept , kind, set, collection, type, class
개체 (Entity)	instance , object, exemplar, element, member, record, individual, dependent variable
개체 속성 (Attribute)	property , predicate, relationship, feature, facet, dimension, characteristics, field, header, independent variable

자료: Bergman(2018)

먼저 개체 종류(entity type)는 온톨로지를 구성하는 개체의 구분 기준으로, 그래프 관점에서는 어떤 구상개념에 관한 ‘노드(node)의 종류’를 의미한다. 예를 들어 어떤 온톨로지가 존재한다고 할 때, 최상위 노드에 해당하는 개념(concept)은 이름을 붙이거나 물질적이지 않을 수 있다. 구체적으로, 개념(concept)은 이에 해당하는 메타데이터를 가질 수는 있지만 속성(attribute)을 가지지는 않는다. 클래스(class)는 같은 문맥에서 설명할 수 있는 잠재적 관계나 속성을 공유하는 하나 이상의 멤버(member) 집합이며, 따라서 개체 종류(entity type)나 개념(concept)은 용어 사용에 따른 뉘앙스는 다르더라도 결국 클래스(class)에 해당한다. 두 번째로 개체(entity)는 온톨로지를 구성하는 핵심 요소 그 자체로, 그래프 관점에서는 노드(node) 자체를 의미한다. 온톨로지를 작성하는 환경(i.e., 언어)이나 사용자에 따라 다양한 명칭이 있지만, 주로 인스턴스(instance)나 요소(element), 개인(individual) 등이 많이 사용된다. [그림 2-2]를 통해 자세히 예를 들자면, ‘root’, ‘pizza’, ‘topping’ 등 원으로 그려진 것이 모두 개체(entity)에 해당한다. 그러나 ‘pizza’와 그 하위에 놓인 개체는 ‘메뉴판에서 구분되는 피자 종류’에 대한 개념 또는 클래스라면, ‘topping’과 그 하위에 있는 개체는 ‘피자의 특징을 결정짓거나 혹은 피자의 종류와 관계 없이 추가될 수 있는 부재료’에 대한 개념 또는 클래스에 해당한다. 마지막으로 개체 속성(attribute)은 어떤 개체(entity)에 내재한 특성(characteristic)이며, 이와 같은 속성은 개체를 정의하거나 더욱 자세히 설명할 수 있게 하는 요건에 해당한다. 온톨로지 내에서 모든 속성은 해당 속성에 해당하는 값(value)을 가지고 있으며, 같은 개체 종류(entity type)는 동일한 잠재 속성을 공유한다. 예를 들어, [그림 2-2]에서 ‘pizza’와 그 하위 개체는 ‘판매 가격’, ‘원가’, ‘맵기’ 등의 속성을 동일하게 가질 수 있고, ‘topping’과 그 하위 개체는 ‘식자재 구분(e.g., 채소, 육류, 해산물 등)’, ‘토픽 추가 가격’ 등의 속성을 공유할 수 있다.

지식 표현(knowledge representation)의 관점에서 벗어나 온톨로지의 실제 활용 측면에서는 데이터베이스 영역의 용어 또한 이해할 필요가 있다. 예를 들어, 공리(axiom), 관계(relationship), 제약조건(constraint), 이벤트(event), 액션(action) 등이 있다(최익규, 2013). 공리(axiom)는 특정 지식 도메인 하에서 무엇이 참인지를 명시하는 것을 의미한다. 공리는 클래스

스(class)나 속성(attribute), 주장(assertion), 주석(annotation) 등 다양한 형태로 온톨로지 내에 구성될 수 있다. 이와 같은 공리는 온톨로지에 새로운 개념이 추가되어 연역적 추론(deductive reasoning)이 필요하거나 온톨로지 갱신·검증을 수행할 때 활용한다. 관계(relationship)는 개체 간, 속성 간, 또는 개체-속성 간에 상위·하위의 위계적 관계뿐만 아니라 다양한 관계 설정이 필요할 때 정의한다. 제약조건(constraint)은 온톨로지 내에서 속성값(value), 개체 간 관계(relationship)의 방향 등에 대해 조건을 거는 것으로, 온톨로지의 구축·수정·활용 등에서 오류를 방지하기 위해 활용한다. 이벤트(event)는 온톨로지의 갱신 또는 수정시 개체·속성·관계·값 등의 변화를 의미한다. 액션(action)은 이벤트의 종류로, 예를 들어 ‘추가’, ‘변경’, ‘제거’, ‘생성’ 등을 포괄한다.

3. 온톨로지 작성

일반적으로 온톨로지는 특정한 문법(syntax) 규칙을 갖는 언어(language)로 작성하는데, Maniraj & Sivakumar(2010)는 다양한 온톨로지 언어를 크게 세 가지로 분류했다. 첫 번째는 논리형 언어(logical language)로 기호학 기반의 형식 논리(formal logic)를 구현하려는 목적으로 설계된 인공지능 언어를 의미한다. 논리형 언어는 다시 1차 술어 논리(first order predicate logic, FOPL)와 규칙 기반 논리(rule-based logic), 기술 논리(description logic, DL)과 같은 논리 체계에 따라 분리된다. 두 번째는 프레임 기반 언어(frame based language)로 인공지능과 지식 표현 분야에서 주로 사용되는데, 프레임(frame)이라고 부르는 전형화된 상황(stereotype situation) 구분에 따라 구조화된 데이터 표현 방식에 해당한다. 마지막은 그래프 기반 언어(graph based language)로 지식이나 데이터를 그래프 구조로 표현하는 방식에 기반한 언어 체

〈표 2-2〉 온톨로지 언어 종류

구분	논리형 언어	프레임 기반 언어	그래프 기반 언어
표현 방식	형식 논리 기반 표현 (FOPL, 규칙기반, DL)	객체·속성 간 구조화된 표현	노드·엣지를 활용한 그래프 구조 기반 표현
주요 특징	정밀한 표현 및 추론 가능	계층적, 속성 중심의 구조적 표현을 강조함	유연하고 확장 가능한 관계 표현
적합한 용도	복잡한 지식 표현 및 추론	도메인 모델링 및 속성 정의	대규모 연결 데이터 및 지식 그래프
장점	엄격한 표현, 강력한 추론	구조적이고 인간친화적인 표현	유연성, 시멘틱 웹 지원
단점	큰 계산비용	복잡한 논리 표현 어려움	논리적 제약 표현의 한계

자료: 저자 작성

계를 말하는데, 노드(node)와 엣지(edge)를 사용해 객체와 관계를 나타낸다. <표 2-2>는 이와 같은 온톨로지 언어 분류에 따른 차이를 기술한다.

최초로 개념적 지식을 ‘온톨로지’의 형식으로 표현하려는 대표적인 시도는 1970년대 후반 프레임 기반 언어인 KL-ONE으로 진행되었고, KL-ONE은 오늘날 사용되는 시맨틱 웹(semantic web) 기술의 이론적 기반을 제공한 여겨진다(Brachman & Schmolze, 1985). KL-ONE의 등장 이후 다양한 온톨로지 언어가 개발되고 발전해 왔으며, 오늘날에 와서는 오픈 소스(open source)를 지향하는 온톨로지 편집기(i.e., protégé 등)도 등장해서 다양하게 활용되고 있다. <표 2-3>은 각 언어 종류별 대표적인 언어를 5개씩 소개한다.

<표 2-3> 대표적인 온톨로지 언어

구분	대표적 사용 언어
논리형 언어	1. OWL(Web Ontology Language) - W3C(World Wide Web Consortium) 표준 온톨로지 언어로, 기술 논리(DL)에 기반해 설계되었지만, 범용적으로 사용 가능함 - 시맨틱 웹 및 온톨로지 설계에 활용, 복잡한 지식 표현 가능 2. Prolog - 선언형 프로그래밍 언어로, 규칙 기반 지식 표현과 논리적 추론이 가능함 - 전문가 시스템, 자연어 처리, 지식베이스 설계 등에 활용 3. KIF(Knowledge Interchange Format) - 1차 술어 논리(FOPL)를 기반으로 표준화된 지식 표현 언어 - 지식 교환 및 논리적 추론에 사용 4. CLIPS(C Language Intergrated Production System) - 전문가 시스템 구축을 위한 논리적 규칙 기반 언어 - 산업용 전문가 시스템 및 추론 엔진 개발 등에 활용 5. RIF(Rule Interchange Format) - W3C 표준 규칙 언어로 다양한 온톨로지 언어와 통합 가능하다는 특징이 있음 - 온톨로지의 규칙 기반 지식 표현 및 추론에 활용
프레임 기반 언어	1. F-Logic(Frame logic) - 프레임 기반 언어와 논리형 언어의 특징을 결합한 언어 - 객체 속성 정의 및 추론에 활용 2. FRL(Frame Representation Language) - 프레임 기반 언어의 대표적인 초기 언어 - 객체-속성 관계 표현에 집중 3. OKBC(Open Knowledge Base Connectivity) - 객체 중심 지식 표현을 지원하는 언어 - 지식 기반 시스템 설계에 활용 4. KL-ONE - 논리형 언어인 DL의 초기 형태 언어 - 개념 계층 및 관계 표현 가능 5. Ontolingua - 온톨로지 공유 및 재사용 언어 - 온톨로지 협업 및 상호 운용성이 뛰어남

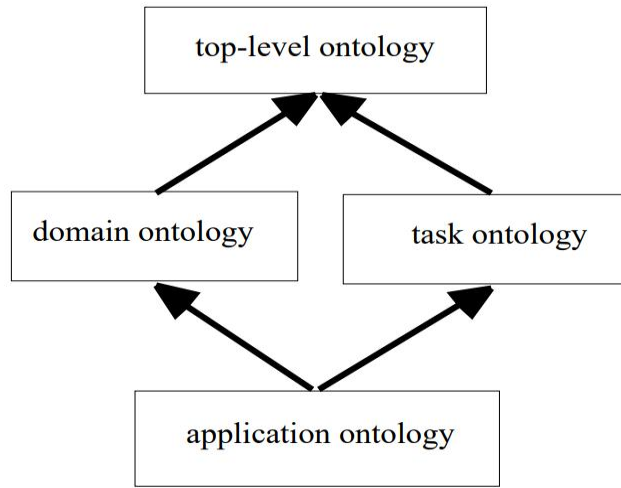
구분	대표적 사용 언어
그래프 기반 언어	1. RDF(Resource Description Framework) - 트리플(Subject-Predicate-Object / 주부-술부-목적어) 구조를 표현하기 위한 언어 - 시맨틱 웹 및 지식 그래프에 활용
	2. Turtle - RDF 표현을 간결하게 지원하는 언어로, 온톨로지 데이터 모델링에 활용함
	3. SPARQL(SPARQL Protocol And RDF Query Language) - RDF 데이터를 질의하기 위한 용도로 개발된 언어로, 시맨틱 데이터 검색에 특화
	4. Cypher - 그래프 데이터베이스용 언어로, 지식 그래프 설계 및 질의에 활용
	5. RDFa(RDF in Attributes) - HTML 문서에서 RDF를 표현하기 위한 언어로 웹 문서에 시맨틱 정보를 추가함

자료: 저자 작성

4. 온톨로지의 범위

시맨틱 웹 구축과 지식 표현의 관점에서 보면, 인류의 모든 지식을 한 번의 시도로 데이터베이스화하기란 사실상 불가능하기에 결국 온톨로지 간 연계와 협력을 염두에 둘 수 밖에 없다. 예를 들면, Guarino(1997)는 온톨로지 범위에 따라 각각 독립성이 다르며 이에 따른 네 종류의 온톨로지를 [그림 2-3]과 같이 구분한다. 독립성의 관점에서 응용 온톨로지(application ontology)는 도메인 온톨로지(domain ontology)와 과업 온톨로지(task ontology)에 종속되며, 다시 도메인 온톨로지와 과업 온톨로지는 상위 온톨로지(top-level ontology)에 의존하는 형식이다. 상위 온톨로지는 시공간, 사건, 행위, 물질 등 매우 보편적인 개념에 대한 것을 다루고 있기에, 특정한 도메인이나 문제와 독립적이다. 또한 이와 같은 특성 때문에, 일반적으로 매우 큰 사용자 집단을 가진다. 도메인 온톨로지와 과업 온톨로지는 각각 일반적인 도메인(e.g., 제약, 자동차 등)이나 일반적인 행동·과업(e.g., 진단, 판매 등)에 관련된 개념으로 구성되며, 최상위 온톨로지서 소개된 개념의 세분화 또는 맞춤화가 이루어진다. 마지막으로 응용 온톨로지는 특정한 도메인이나 과업에 종속된 하위 개념을 포괄한다. 예를 들어 특정 도메인 내 어떤 개체의 과업 수행을 위한 ‘역할’이나 ‘교체 품목’ 또는 ‘예비 부품’ 등이 이에 해당한다.

[그림 2-3] 온톨로지 범위에 따른 구분



자료: Guarino(1997)

Gangemi et al.(2001)은 상위 온톨로지를 설계하기 위한 체계적 방법론을 제시하고 있는데, 이를 통해 상위 온톨로지와 하위 온톨로지의 특징적 차이를 더욱 분명히 할 수 있다. 요약하자면 상위 온톨로지일수록 관계 속성에 대한 공리를 기반으로 하는 명확한 구분과 제약조건의 필요성이 커지며, 형식적 관계(formal relations)를 바탕으로 설계해야 한다. 여기서 형식적 관계란 도메인 중립적인 원시 관계(primitive relations)를 의미한다. 예를 들면, 부분-전체 관계(partthood), 속성의 상속(instantiation) 등이 이에 해당한다. 또한 온톨로지가 작성되는 언어의 영향을 크게 받기 때문에 언어 및 인지(cognition)와의 연결을 고려해야 한다.

Mascardi et al.(2007)은 상위 온톨로지(i.e., top-level ontology, upper ontology, or foundation ontology)에 대한 위키피디아의 정의⁴⁾를 인용하면서, 총 7개의 상위 온톨로지(i.e., BFO, Cyc, DOLCE, GFO, PROTON, Sowa's, SUMO)를 <표 2-4>와 같이 비교한 바 있다.

4) 모든 도메인에서 차별 없이 적용할 수 있는 매우 일반적인 개념들에 대한 설명으로, 각 상위 온톨로지의 목표는 해당 도메인과 연계할 수 있는 다수의 하위 온톨로지 확보를 통한 지식의 연결(an attempt to create an ontology which describes very general concepts that are the same across all domains. The aim is to have a large number on ontologies accessible under this upper ontology)

〈표 2-4〉 대표적인 상위 온톨로지

구분	대표적 사용 언어
BFO	▶ Basic Formal Ontology ▶ 1998년 Smith와 Grenon에 의해 BFO 작성 이론이 개발된 이후, 계속해서 개선 중에 있음 ▶ 시간에 따라 변화하는 개념(i.e., SNAP)과 시간에 걸쳐 확장되는 개념(i.e., SPAN)을 포괄하며, 각각을 모듈로 구성하는 모듈형 구조 ▶ OWL로 작성되었으며, 무료 라이선스를 제공함 ▶ “basic-formal-ontology.org/”에 의하면 물리학, 화학, 생물학 또는 그 외 특별한 영역에서 사용되는 용어들로 구성된 BFO 2020 버전이 최신 버전으로, BFO 1.0부터 현재까지 약 550개가 넘는 활용 및 하위 온톨로지 개발이 전세계에서 진행되었음 ▶ 주로 생물의학 분야에서 활용됨
Cyc	▶ Douglas Lenat이 1984년 발표한 후, 1995년부터 Cycorp에서 CEO를 맡고 계속해서 개발이 진행되고 있는 상위 온톨로지 ▶ 특정 도메인에 대한 지식을 담은 마이크로이론으로 구성된 방대한 지식 기반으로, “cyc.com”에 의하면 병원 운영, 공급망 관리와 같은 특정 영역뿐만 아니라 범용적인 기업 지식 레이어의 기반으로 활용할 수 있음 ▶ WordNet synsets와 매핑되어 있고, 모듈형 구조로 설계됨 ▶ CycL과 OWL을 사용하며, 상업용 버전(i.e., OpenCyc)과 무료 버전(i.e., ResearchCyc)을 제공함
DOLCE	▶ Descriptive Ontology for Linguistic and Cognitive Engineering ▶ Borgo et al.(2022)에 의하면 20여년 동안 안정적으로 유지되고 있는 온톨로지로, 현재는 사회과학, 제조, 금융, 문화 유산 등 다양한 도메인에서 활용되고 있음 ▶ 자연어와 일상적 사고를 반영하는 개념을 집중적으로 다루고 있고, 따라서 정보 검색 및 전자 학습 등에서 탁월한 성능을 보이고 있음 ▶ 자연어 기반으로 설계됨 ▶ FOPL, KIF, OWL을 사용하며, 무료로 제공됨
GFO	▶ General Formal Ontology ▶ 독일의 University of Leipzig의 의료정보연구소에서 Herre et al.(2006)에 의해 개발된 온톨로지인 만큼 생물의학에 초점을 맞추고 있지만, 정신적, 사회적 영역을 아우르는 통합적 개념 모델링에도 활용됨 ▶ 추상적 상위 수준, 코어 수준, 기본 수준으로 구성된 세 가지 계층에 기반 ▶ OWL, FOPL을 사용함
PROTON	▶ PROTo Ontology ▶ 2004년부터 2006년까지 유럽연합 6차 프레임워크 프로그램의 펀드를 받은 SEKT 프로젝트의 일환으로 개발됨 ▶ KIM(Knowledge and Information Management) 플랫폼에서 사용되는 KIMO(KIM Ontology)를 기반으로 개발된 만큼, 상위 온톨로지임에도 상대적으로 경량(light-weight)인 특징이 있음 - 예를 들어, Cyc(수십만개의 개념과 수백만개의 명제)나 SUMO(수천개의 클래스와 공리)에 비해 300개의 클래스와 100개의 속성으로만 구성됨 ▶ PROTON은 semantic annotation, 지식관리시스템, 미디어 연구 및 분석을 위한 연구용 온톨로지 작성, 연구 인텔리전스, 기업 데이터 온톨로지 등 다양한 분야에서 다양한 목적으로 활용되고 있음 ▶ OWL Lite를 사용함
Sowa's ontology	▶ Sowa(1999)에 의해 개발됨 ▶ 미리 정의한 계층 대신 다양한 구분을 기반으로 자동 생성되는 구조를 갖는 특징이 있음 ▶ 논리, 언어학, 철학을 기반으로 설계되었으며, 여러 상위 온톨로지 개발에 영감을 제공함 ▶ FOPL Modal Language, KIF 사용

구분	대표적 사용 언어
SUMO	▶ Suggested Upper Merged Ontology ▶ Teknowledge Corporation에서 시작되었으며 IEEE의 Standard Upper Ontology Working Group의 지원을 받아 Ontolingua 서버에서 제공하는 공개 온톨로지와 Sowa's ontology 등을 통합하여 개발함 ▶ 가장 큰 공용 온톨로지 중 하나로 검색, 언어학, 시맨틱 추론 등에 활용 ▶ WordNet 2.1 버전과 수작업으로 매핑됨 ▶ 공개 온톨로지 연구 및 산업 응용 분야에서 널리 사용되고 있음 ▶ SUO-KIF, OWL 사용

자료: Mascardi et al.(2007)의 내용을 기반으로 저자가 번역 및 보완

비교한 7개의 상위 온톨로지는 개발 동기와 개발 주체, 개발 시기와 작성 언어 등이 모두 달랐지만, 서로 다른 도메인 온톨로지와의 상호운영성(interoperability)을 지원하고 있기에 다양한 도메인, 과업, 응용 온톨로지 개발의 근간(foundation)으로 활용될 수 있었다.

한편 상위 온톨로지 관련 설명을 보면, WordNet이 중요한 특징으로 기술된다. 비록 WordNet이 온톨로지가 아닌 어휘 데이터베이스(lexical database)지만, WordNet과의 맵핑은 더욱 풍요로운 정보 제공과 논리적 추론을 가능케 하기에 핵심 데이터로 볼 수 있다. 프린스턴 대학의 인지 과학 연구소에서 1985년부터 WordNet을 개발했고(c.f., Miller et al., 1990), WordNet은 영어 단어를 'synset'이라고 하는 유의어 집단으로 분류하여 정의하고 어휘목록 간의 다양한 의미 관계를 규정하고 있다.

제 2 절 국내외 온톨로지 사례

1. 해외 온톨로지 사례

온톨로지가 본디 시맨틱 웹과 정보 추론을 지향하는 일종의 정보 체계이지만, 이를 면밀하게 구축하기 위해서는 막대한 시간적 물질적 자원이 소요된다. 특히 상위 온톨로지에 가까울수록 비용 증가는 필연적이기 때문에, 대다수의 실제 온톨로지 구축 사례들은 도메인 또는 응용 온톨로지에 해당한다. 특정 개념의 명확한 정의 구축과 구성요소 정리를 위한 사용자 서비스를 목적으로 할 때는 LOD 구조도 빈번하게 관찰된다.

〈표 2-5〉는 해외 온톨로지 사례 목록으로, 수천여 개의 온톨로지 관련 연구에서 연구 결과물로 소개하는 온톨로지가 아니라 실제 서비스 적용 또는 그 외 활용 사례가 있는 경우만 기술했다.

〈표 2-5〉 해외 온톨로지 사례

구분	온톨로지명	분야	목적	특징	적용결과 (사례)
1	Ontology for Media Resource	미디어	미디어와 관련된 서로 다른 설명을 연결하고 핵심 설명 속성의 집합을 제공하는 것	기존 메타데이터 형식의 집합 요소에 대한 매핑 정의, 미디어 리소스의 핵심 메타 데이터 속성 정의, RDF/OWL을 사용한 추상 온톨로지의 시맨틱 웹 호환 구현 제시	웹사이트 (www.w3.org/TR/mediaont-10/)
2	미국 OMC(Ontology for Media Creation)	미디어	클라우드 기반의 인간-기계 협업의 토대 마련을 통한 엔터테인먼트 산업 경쟁력 강화	소규모 온톨로지 구축에서 출발하여 확장해 나가는 형태로 진행하며 AWS 클라우드를 활용하여 반복적이고 기본적인 제작 과업의 자동화와 전반적인 창작 워크플로우에 있어 표준화 및 데이터화를 추구	할리우드 내 스튜디오에서의 영화제작 시 활용할 수 있는 미디어 리소스 제공
3	독일 Digital Media Ontology	미디어	여러 미디어 종사자들이 레가시 미디어와 OTT 등 다양한 채널을 넘나들며 유통되는 모든 데이터를 즉각 검색, 제작 및 유통 역량 강화에 활용할 수 있도록 추진	DMO, TripleStore, API, adaptors의 4개 레이어로 구성되는 LMDC(Linked Media Data Cloud)를 통해 다양한 앱을 개발하여 가치사슬 전반의 활동 데이터들이 자동으로 통합되고 상호운용성을 확보하도록 설계됨	독일 내 공영 방송사 및 스트리밍 서비스사 등이 활용할 수 있는 미디어 리소스 제공
4	EU 유로피아나 LOD	역사자료	유럽의 풍부한 문화 유산들을 토대로 다문화적, 다언어적 환경과 기술적 진보 및 새로운 비즈니스 모델을 연계하여 이를 온라인 환경에서 쉽고 흥미롭게 사용하도록 만드는 것	EDM(Europeana Data Model)을 이용하여 메타데이터를 저장하고, RDF로 변환하면서 이용자들은 SPARQL end-point나 HTTP URI를 통해 데이터에 접근할 수 있도록 함	유로피아나 전자도서관 사이트
5	아일랜드 Location LODer	지리학, 문화	기존에 제한적으로 접근할 수 있었던 logainm.ie 데이터를 Linked Data형식으로 구조화된 데이터 덤프에 바로 접근할 수 있도록 함	logainm.ie와 DBpedia, National Library of Ireland, Europeana의 데이터들을 결합하여 한 화면에서 여러 정보들을 동시에 볼 수 있도록 인터페이스를 구축하고, 구글 지도를 기반으로 하여 아일랜드 전역의 정보들을 지도를 통해 확인 가능함	아일랜드 국립도서관의 Location LODer
6	영국 BBC 온톨로지	미디어	BBC 스포츠, BBC 교육, BBC 뮤직, 뉴스 프로젝트 등 BBC에서 생산되고 소비자들에게 제공되는 콘텐츠에 대한 설명을 제공하기 위함	데이터 레이어와 서비스 레이어를 분리, 이러한 온톨로지 구축을 통해 Linked Data Platform 내에서 데이터 관리, 저장 및 공유가 가능하도록 하면서 데이터 생산과 재활용에 활용	구축된 온톨로지를 기반으로 BBC 디지털 콘텐츠 아카이브에 적용 및 LOD서비스를 통해 대중에게 제공

구분	온톨로지명	분야	목적	특징	적용결과 (사례)
7	미국 뉴욕타임스 Linked Data Cloud	언론	NYT의 디지털화와 함께 150년 동안의 기사와 기록을 Index로 정리한 것을 외부 정보와도 자유롭게 연동하여 쓸 수 있도록 제공	언론인 및 일반인들이 기사 등을 검색할 때 광범위한 자료를 유연하게 찾아볼 수 있도록 API 등을 제공함으로써 데이터가 다양하게 활용될 수 있도록 함	NYT Developers에서 다양한 분야의 API 검색 시스템 제공

자료: 저자 작성

쉽게 접근할 수 있게 대중에게 공개된 국외 온톨로지를 살펴보면, 대체로 미디어, 역사, 도서관 등 기존 정보를 온톨로지의 형태로 LOD와 연계하여 관련 정보를 찾는 사용자의 편의성을 높이려는 시도가 다수인 것으로 보인다. 이러한 접근은 기존 웹사이트 또는 레거시(legacy) 데이터베이스를 기반으로 큰 비용투자 없이 외부 협업을 기대할 수 있기 때문으로 판단된다.

2. 국내 온톨로지 사례

국내 사례 또한 해외사례와 마찬가지로 공개되어 서비스 중이거나 활용 사례가 있는 경우를 우선해서 탐색했다. 이 역시 LOD를 활용해 사용자 편의성을 증대하고자 하는 시도를 상대적으로 수월하게 찾아볼 수 있었다. 온톨로지 구축 사례로는 <표 2-6>과 같이 국립국어원의 말뭉치 구축, 외교부의 외교문서 LOD 데이터 발행을 위한 온톨로지 구축, 국립중앙도서관의 국가서지 LOD, 서울시의 개방데이터 온톨로지, 한국관광공사의 스마트 관광 정보 온톨로지 등이 있다.

<표 2-6> 국내 온톨로지 사례

구분	온톨로지명	분야	목적	특징	적용결과 (사례)
1	방송콘텐츠 온톨로지 구축	미디어	IPTV환경의 도래와 개인 맞춤형 서비스 제공을 위한 검색 시스템 고도화	방송정보를 의미정보로 분절화하고 해당 내용을 각 분류체계에 알맞게 태깅하는 과정을 거쳐 방송콘텐츠 정보를 메타태깅을 응용하여 내용 태깅 자동화 기반 마련	방송통신위원회 정책과제
2	국립국어원 말뭉치 구축	언어 (한국어)	국내의 언어 현실을 충실히 반영한 국어 말뭉치를 지속적이고 체계적으로 구축하여 다양한 분야에 사용할 기초자료를 마련하고 4차 산업혁명 대비와 인공지능 기술 개발 및 활용을 위한 준비를 위함	실제 이용되는 게시글들을 컴퓨터 기법을 활용해 수집 및 가공하여 웹 말뭉치 데이터를 구축	국립국어원 한국어 학습자 말뭉치 나눔터

구분	온톨로지명	분야	목적	특징	적용결과 (사례)
3	외교문서 LOD 데이터 발행을 위한 온톨로지 구축	외교	역사적 학술적 가치가 높은 외교 문서를 데이터 활용성이 높은 LOD 형태로 발행하여 활용도 제고	지역, 국가, 도시, 인물, 기관, 사건, 년도를 핵심 데이터(Core Ontology)로 구성하고 외교문서와 외교안보연구소 발간자료를 연결하여, 지역/ 국가/ 도시/ 인물/ 기관/ 사건/ 연도를 기준으로 관련된 외교문서 및 발간자료를 찾을 수 있도록 구성	외교부 OPEN DATA 사이트 내 '외교부 LOD 자료 탐색'
4	국립중앙 도서관 국가서지 LOD	문헌자료	국립중앙도서관이 보유하고 있는 정보자원을 도서관 밖 다양한 분야의 기관·개발자·이용자들 이 직접 활용하고, 다양한 웹 자원과 자유롭게 연계하여 도서관 및 웹 데이터를 더욱 풍부하게 만드는 기반을 조성하기 위함	기존의 (KOR)MARC 형태 또는 DBMS 형태의 데이터를 RDF 형식으로 변환하여 웹상에 Linked Open Data 형식으로 발행하고, 해당 메타데이터를 활용하여 다양한 검색 및 데이터 탐색 서비스 제공	국립중앙도서관 국가서지 LOD 서비스
5	서울시 개방데이터 온톨로지	지역 데이터 (인구통계, 통신, 카드 이용 등)	개발자 및 일반 사용자들이 보다 자유롭게 데이터를 활용할 수 있도록 제공함으로써 민간의 다양하고 새로운 비즈니스의 창출의 기회를 제공하고 IT를 바탕으로 한 콘텐츠 사업 육성 및 이를 통한 부가가치를 창출	고유 URI를 부여하여 LOD 형태의 데이터를 자유롭게 활용할 수 있고, SPARQL을 이용하여 데이터를 질의할 수 있으며, 내/외부 데이터 간 연결 정보를 제공하여 다양한 정보를 쉽게 획득할 수 있음	서울시 열린데이터 광장 LOD 라이브 서비스
6	한국관광공사 스마트 관광 정보 온톨로지	관광 데이터	관광산업의 디지털 전환과 맞춤형 스마트관광 서비스 제공	한국관광공사에 의해 만들어진 관광지, 숙박, 레스토랑, 축제, 관광분류체계 등의 메타데이터를 포함하고 있고, Creative Commons Attribution-ShareAlike 3.0 Unported License의 조건에 따라 구축됨. 데이터 포맷은 RDF를 이용해서 표현하고 있으며, 연결된 데이터를 바탕으로 차트나 그래프 제작 등의 기능 지원	스마트관광정보 Linked Open Data 서비스
7	인문지리정보 온톨로지	지리	언제 어디서나, 누구나 이용 가능한 공간정보와 시맨틱 웹 기반의 사용자 맞춤형 인문지리정보 통합 DB 구축 및 서비스	역사, 관광 등 12개 도메인을 기준으로 온톨로지 지식베이스를 구축하고, 시맨틱 웹 기반의 포털서비스를 구축. 이를 통해, 통합관리를 용이하게 하고 지리정보의 활용도 제고	국토지반정보 포털시스템

구분	온톨로지명	분야	목적	특징	적용결과 (사례)
8	역사용어 온톨로지	역사	대한민국 국민 누구나 우리 역사를 쉽게 접근하고 재미있게 배울 수 있도록 콘텐츠를 개발하고 한국사 외의 타 분야에서도 관련된 한국사 내용을 쉽게 연계하여 사용할 수 있도록 제공	국사편찬위원회 인물 온톨로지와 시소러스, 한국학중앙연구원 민족문화대백과사전, 문화재청 유물/유적데이터와 각 유관 기관의 역사 데이터들을 통합한 데이터 혹은 링크 형태로 융합하여 이를 시맨틱 서비스 프레임워크를 통해 LOD 기반의 한국사 연계 통합 DB 구축	한국사 LOD 서비스
9	법률 정보 온톨로지 모델	법	다양한 서비스와 자원 형태로 분산되어 있는 법률정보를 통합 서비스로 제공	대한민국, 일본, 독일, 프랑스, EU, 미국, 영국 등에 산재되어 있는 법률 서비스들과 법령, 판례, 논문, 민원, 행정문헌 등 다양한 정보 자원을 20개 클래스, 42개 데이터타입 속성, 22개 객체 속성 등으로 구성된 온톨로지 모델로 구축함	iLaw 시스템
10	생물정보 온톨로지	생물	국립수목원의 국가생물종 지식정보시스템. 분산된 생물정보를 통합하여 단일화된 창구를 통해 사용자에게 제공	생물종에 대한 기본적인 정보를 포함하여 생물종의 특성, 그리고 생물종들 사이의 관계와 지리적 정보, 역사, 문화, 음식 등 다양한 도메인과의 관계 정보를 구축, 2013년 사업에서 식물, 균류, 척추동물 3가지 분류 군에 대한 정보를 LOD로 구축하였고, 2014년 거미, 곤충, 지의류 등의 분류 군을 추가 구축하여 서비스	LOD생물정보통 합시스템
11	취업정보 온톨로지	취업	취업정보와 관련한 데이터들을 필요로 하는 국민들이 직무·훈련·직업 등의 원천 데이터 정보나 제공되는 데이터의 사전적 정보를 자유롭게 활용할 수 있도록 LOD 데이터를 제공	워크넷과 관련한 직무능력 · 직업 · 자격증 등의 데이터를 지식 기반으로 서비스 하기 위해 RDF 형태로 구축하여 제공, 국가직무능력표준(NCS) 데이터와 자격증, 직업훈련, 직업, 학과 관련 데이터가 수집되어 연결되어 있으며, 주기적으로 추가 수집되어 변경사항을 갱신	Worknet Linked Open Data
12	학술연구정보 온톨로지	학술	KERIS의 데이터가 필요한 분야에서 데이터 융합을 통한 새로운 서비스를 창출할 수 있는 기반을 제공	단행본/학위논문 서명의 고유키를 URI로 적용, Marc 전체를 RDF로 변환	RISS Linked Open Data

구분	온톨로지명	분야	목적	특징	적용결과 (사례)
13	시설물 안전정보 온톨로지	안전	시설물종합관리정보시스템에서 시설물의 안전관리에 관한 특별법에 의해 제공되고 있는 시설물에 대한 안전관리 현황 정보를 보다 자유롭게 데이터로써 활용할 수 있도록 LOD(Linked Open Data) 형태로 제공	시설물 온톨로지는 대한민국 국토안전관리원이 보유하고 있는 다양한 유형의 공공시설물 정보와 이에 대한 안전등급, 그리고 점검진단 실적 및 보수보강 실적을 기술하기 위한 다양한 용어를 포함	국토안전관리원 공공시서물 안전정보 LOD
14	부산 영화영상 관광정보 온톨로지	영화영상 관광	부산지역내 다양한 부서 및 직속 산하기관, 재단, 민간기관 등에 분산되어 있는 부산 영화·영상과 관광정보 LOD형태로 발행하여 민간기업의 활용도를 향상 시키고, 단편적인 정보검색에서 입체적으로 풍부한 영화·영상 및 연계정보 제공	영화정보(영화정보, 상영정보, 출연배우 등) 11,270건, 공연·전시정보(연극, 뮤지컬, 클래식, 오페라 등) 12,657건, 관광지정보(관광명소, 전시장, 갤러리, 극장 등) 4,590건, 기타 67,031건 데이터를 종합하여 온톨로지 모델 구축	부산영화 영상관광 Linked Open Data

자료: 저자 작성

본 연구가 국내 플랫폼을 위한 온톨로지 개발 방향의 검토인 만큼, 실제 활용 사례 외에 국내에서 진행된 온톨로지 관련 연구 결과 또한 추가로 탐색해보았다. 이러한 사례 탐색은 본 연구에서 설문을 위한 온톨로지 개발자 선정 및 연구관점에서의 온톨로지 개발 결과물을 어디까지 기대할 수 있는지를 확인하기 위함이다. 그 결과 <표 2-7>의 국내 관련 연구 사례 외에도 다양한 연구가 출판되고 있음을 확인했고, 기관 중심의 개발이 아니라 단일 연구팀의 구축 사례가 빈번하게 관찰되고 있음을 알 수 있었다. <표 2-7>에 기술한 국내 연구 사례로는 미디어 분야의 방송콘텐츠 온톨로지 구축, 식품 분야의 식품 안전 온톨로지 구축, 식품첨가물 정보 온톨로지 구축, 범죄 수사 분야의 침입, 절도 범죄 온톨로지 구축, 철학 분야의 철학 온톨로지 구축, 재난 분야의 재난 안전 정보 온톨로지, 문화 분야의 종교유적 건축물 정보 온톨로지 등이 있다.

〈표 2-7〉 국내 온톨로지 구축 연구 사례

구분	온톨로지명	분야	목적	특징	비고
1	방송콘텐츠 온톨로지 구축	미디어	IPTV환경의 도래와 개인 맞춤형 서비스 제공을 위한 검색 시스템 고도화	방송정보를 의미정보로 분절화하고 해당 내용을 각 분류체계에 알맞게 태깅하는 과정을 거쳐 방송콘텐츠 정보를 메타태깅을 응용하여 내용 태깅 자동화 기반 마련	김형주 외(2008)
2	식품안전 온톨로지 구축	식품	식품안전에 관한 소셜빅데이터를 분석하기 위한 틀로서, 식품안전 온톨로지를 개발	식품안전과 관련한 정부 정책과 연구 보고서 62편과 4개년간 의식품안전 사건을 다룬 언론기사(2011~2014년)를 분석해 식품안전 관련 용어를 추출, 용어들을 '예방요인', '생성요인', '결과요인', '사회관계요인'의 총 4가지 대분류로 분류	배지영· 조미숙(2016)
3	식품첨가물 정보 온톨로지	식품	식품의약품안전처와 식품안전정보포털에서 제공하는 식품첨가물 정보를 소비자에게 제공	식품첨가물을 Protégé를 이용하여 OWL기반의 지식으로 표현함. 식품첨가물명, 기원, 목적, 기본정보를 제공하기 위한 클래스, 속성, 관계를 정의	김은영· 김용기(2017)
4	침입 절도 범죄 온톨로지 구축	범죄/수사	대규모의 범죄 수사기록을 온톨로지로 구축하고 이를 통한 서비스를 구현	비정형 데이터인 범죄 수사기록 문서 중 침입 절도 분야로부터 추출한 정보를 통해 온톨로지를 구축	고건우 외(2019)
5	철학 온톨로지	철학	-	철학 학문의 지식을 구조화하는 실용적인 온톨로지를 구축하는 방법으로서 텍스트 내용 에 들어있는 지식을 온톨로지로 생성하는 토픽맵 기반의 텍스트 온톨로지 방법론을 제시	김정민 외(2005)
6	재난안전 정보 온톨로지	재난	재난안전정보의 효율적인 관리와 제공을 위해 LOD기반의 재난안전정보 서비스 모형을 제안	재난안전정보 6가지 영역을 토대로 표준용어집 및 모델링 도구를 활용하여 온톨로지 개념 모델을 정의하였으며, 이를 기반으로 클래스와 속성을 제시	김태영 외(2017)
7	종교유적 건축물 정보 온톨로지	문화	온톨로지를 이용한 문화유산정보의 콘텐츠화와 웹 서비스 제공	종교유적 건축물 메타데이터 항목에 근거하여 서울시 소재 72개 종교유적 건축물에 대하여 건축물 부분, 스토리텔링 부분, 참고자료 및 관련 자료의 세부분으로 이루어진 데이터베이스를 구축	정희선 외(2013)

자료: 저자 작성

연구 사례 결과물은 온톨로지 구축 방법론 또는 활용 데이터에 대한 상세한 정보가 제공되는 장점이 있지만, 구축한 온톨로지 결과물의 확인이 상대적으로 어렵다는 단점도 존재한다. 그러나, 온톨로지 구축에 필요한 자원과 설계 방법 및 예상 결과물을 짐작해볼 수 있다는 점에서 온톨로지 구축에 앞서 참고할 수 있는 자료로서의 가치가 있다.

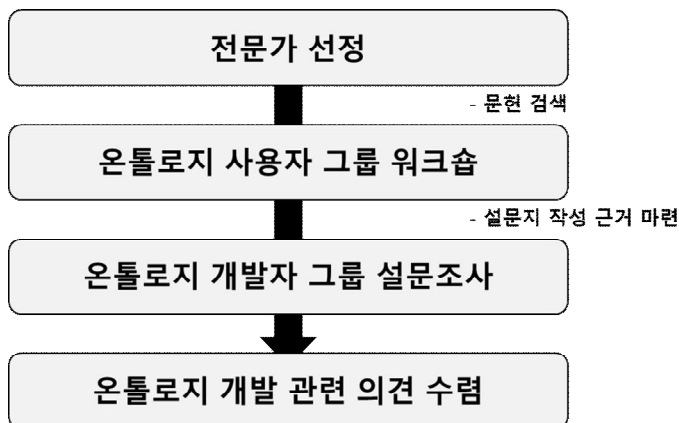
제3장

연구프레임워크

제 1 절 연구 프로세스

본 연구에서는 크게 두 차례에 걸쳐 각기 다른 전문가 그룹의 의견 수렴을 수행한다. 먼저 첫 번째 전문가 그룹은 ‘온톨로지 사용자’ 그룹으로, 온톨로지 연구 경험을 통해 느꼈던 실효성의 측면에서 온톨로지 개발에서 고려해야 할 부분에 대해 듣고자 했다. 특히, 온톨로지의 특성상 도메인을 설명하는 데 부족한 요소가 있을 때 그것이 연구 성과에 미치는 영향력에 대해 이해하고, 공공 연구기관 및 연구과제의 제약(e.g., 예산, 연속성 등)이 존재하는 상황에서 온톨로지 개발 실효성이 있을지를 확인하기 위함이었다. 두 번째 전문가 그룹은 ‘온톨로지 개발자’ 그룹으로, 온톨로지 개발 연구를 수행하면서 느꼈던 제약이나 한계점은 물론 이를 해결하는 과정에서 필요하다 느꼈던 니즈 등에 대해 듣고자 했다. 각 전문가 그룹을 대상으로 하는 연구 진행 순서는 [그림 3-1]과 같이 진행했다.

[그림 3-1] 연구 프로세스



자료: 저자 작성

첫 번째 단계는 전문가 선정으로, 온톨로지 관련 연구 논문 및 보고서를 기반으로 관련 전문가 풀을 먼저 만들었다. 이후, 전문가를 온톨로지 사용자와 개발자로 나눈 후, 우선적으로 선정했다. 두 번째는 사용자 그룹을 대상으로 진행한 대면 워크숍이다. 해당 워크숍에서는 사용자 관점에서 개발 요구 사항 등을 확인했고, 온톨로지 개발자 그룹에 대한 설문지 작성을 위한 기본정보를 확보했다. 세 번째는 온톨로지 개발자 그룹을 대상으로 하는 설문조사이며, 설문조사의 내용을 통해 네 번째 단계에서 개발 관련 의견을 수렴한다.

제 2 절 단계별 연구 방법

먼저 워크숍 및 설문조사를 위한 전문가 풀을 구성하기 위해, 최근 5년간 발표된 관련 온톨로지 관련 연구 논문 및 연구 보고서의 주저자와 교신저자 리스트를 작성했다. 온톨로지 개발·활용·방법론에 관한 논문의 수가 매우 많기에, 상위 또는 도메인 온톨로지의 개발 및 높은 활용도가 있는 경우만 포함했다. 다만, 온톨로지 활용에 관련된 대부분의 국내 연구 논문은 매우 제한적이며, 논문 제목에 활용이 있더라도 연구 내용은 온톨로지 개발 및 자체 활용에 가까운 경우가 많았다.

온톨로지 사용자 그룹은 대면 워크숍을 통해 실효성 관점에서 온톨로지의 요건에 대한 의견을 수렴하고, 공공연구기관의 온톨로지 개발 진행에 관한 생각을 청취했다. 대면 워크숍으로 진행된 만큼 반구조화된 인터뷰 질의(semi-structured interview questionnaires)를 준비해서 자유로운 토론을 진행했다. 온톨로지 사용자 그룹으로 초빙한 전문가는 총 3인으로 <표 3-1>에 기술한 바와 같다.

<표 3-1> 워크숍 진행 대상

구분	소속	기관구분	직위	비고
1	펜*****	학	Research Fellow	지식 그래프 전문가
2	한*****	연	책임연구원	온톨로지 개발 및 활용 과제 경험 有
3	한*****	연	선임연구원	

자료: 저자 작성

온톨로지 개발자 그룹은 비대면 설문조사를 통해 온톨로지 개발 전반에 대한 의견 수렴을 수행했다. 사용자 그룹 워크숍에서 나왔던 토론 내용 및 개발에 앞서 확인해야 하는 부분들을 근거로 설문지를 설계했으며, 설문지는 양적 분석보다 질적 분석에 초점을 맞췄기에 자세한 설명(narrative)을 서술하는 방식으로 작성했다. 이는 (i) 전문지식이 요구되며 학술적 논의가 필요한 온톨로지 개발의 특성상 개발 경험이 있는 전문가 수 확보가 어렵고, (ii) 온톨로지의 목적 및 개발

도메인이 상이하기에 표본의 대표성을 확보하기 어렵고, (iii) 개발 방향 설정을 위한 의견 청취는 전체 의견의 경향성보다 사례별 경험을 깊게 이해하는 것이 더욱 연구에 적합할 것으로 판단했기 때문이다. 설문지는 앞서 연구 논문 및 보고서를 통해 확보한 전문가 풀을 기반으로 콜드 메일(cold mail)을 보내는 형식을 취했으며, <표 3-2>와 같은 전문가 8인의 답신을 확보할 수 있었다.

<표 3-2> 설문조사 진행 대상

구분	소속	기관구분	포지션	비고
1	캠***	산	관리자	환경
2	서****	학	박사	공간
3	고****	학	교수	역사
4	청****	학	교수	일반
5	명****	학	교수	건축
6	이*****	학	교수	사회복지
7	조****	학	교수	전력
8	한*****	연/학	교수	언어

자료: 저자 작성

제 3 절 질의서 및 설문지 구성

먼저 워크숍 질의서는 <표 3-3>과 같이 준비했다. 그 외의 추가적인 질의나 새로운 아이디어 등을 기반으로 하는 논의는 워크숍을 진행하면서 자연스럽게 수행될 수 있게 진행했다.

<표 3-3> 워크숍 질의서

구분	질의
-	○ (사전 설명) 연구 배경 및 목적, 과제 상황 및 제약점 등 ○ (워크숍 참여자 소개) 온톨로지 관련 연구 및 과제 수행 경험 공유
1	○ 예산이나 과제 구조 등의 제약사항을 배제하고 판단할 때, 도메인 온톨로지 개발이 앞서 설명해 드린 NRC 데이터플랫폼에서 실효성이 있을 것으로 생각하십니까? - [조사] 온톨로지 개발 방향성 검토, 연구 적합도, 도메인 온톨로지 개발의 범위 등
2	○ 도메인 온톨로지 개발에 있어서, 참고할 만한 온톨로지 연구나 사례는 무엇인가요? - [조사] 도메인 온톨로지 개발 필요 사항(e.g., 예산, 시간 등) 확인, 참조 가능한 도메인 온톨로지 조사 등
3	○ 온톨로지를 활용하시면서 불편했던 점이나 아쉬웠던 점은 무엇이고, 이를 해결하기 위해 무엇을 하셨습니까? - [조사] 온톨로지 개발 목표 설정, 도메인 온톨로지 실효성 검토 등
4	○ 국문으로 작성된 도메인 온톨로지 개발이 지향해야 할 목표는 무엇이라고 생각하십니까? - [조사] 온톨로지 개발의 독창성, 온톨로지 개발 결과물의 활용성 등

구분	질의
5	○ 공공연구기관에서 온톨로지를 개발한다면 이로 인해 기대되는 장점과 우려되는 단점은 무엇이 있을까요? - [조사] 온톨로지 개발의 장단점, 사업 수행 적합도, 과제 실효성 등
6	○ 도메인 온톨로지를 개발한 이후, NRC데이터플랫폼에서 수집 중인 혹은 수집할 예정인 데이터나 서비스 등과의 연계가 어떤 방식으로 이루어질 수 있을 것으로 보십니까? - [조사] 온톨로지 개발 실효성, 중장기 활용안 등

자료: 저자 작성

설문조사지의 구성⁵⁾은 <표 3-4>와 같다. 최종적으로는 설문조사의 내용을 바탕으로 정성적 분석을 통해 본 연구의 목적을 달성하고자 했다.

<표 3-4> 설문조사지 구성

구분	질의
1	○ 온톨로지 개발의 의미 및 필요성 - 1.1. 개발 동인 - 1.2. 온톨로지 개발의 목표 - 1.3. 온톨로지 개발 및 운영 주체
2	○ 온톨로지 개발의 제약 및 한계점 - 2.1. 온톨로지 개발 난이도 - 2.2. 온톨로지 개발의 제약 - 2.3. 중소규모 단위의 온톨로지 개발 한계점
3	○ 온톨로지 개발의 향방 및 전망 - 3.1. 온톨로지의 향방 - 3.2. ICT온톨로지에 대한 기대 사항

자료: 저자 작성

5) 설문조사지는 [부록]을 참고하시오.

제4장 결 과

제 1 절 온톨로지 사용자 대상 워크숍

온톨로지 사용자를 대상으로 하는 워크숍은 대면으로 진행하는 만큼 (i) 과제 및 예산 등에 대한 최대한의 정보를 제공하고, (ii) 편안한 분위기에서 자유로운 의견 개진이 가능함을 알려 비록 정제된 정보나 지식이 아니더라도 필요한 정보를 최대한 획득하는 데 초점을 맞췄다. 또한 해당 워크숍을 통해 공유한 정보는 향후 수행할 온톨로지 개발 설문지 작성에 활용될 수 있음을 알렸다.

첫 번째 워크숍 질의는 NRC데이터플랫폼이라는 연구 중심 사용 환경에서의 도메인 온톨로지 개발 실효성에 관한 것으로, 사실상 본 과제에서 의사결정을 내리고자 하는 문제를 직접적으로 가져왔다. 이를 첫 번째 질문으로 한 이유는 (i) 본 연구의 배경과 워크숍의 목적을 분명히 하고, (ii) 참여자 간 의견 교환에 앞서 각 전문가가 온톨로지에 대해 평소에 견지하고 있는 생각을 짐작할 수 있을 것으로 판단하기 때문이다. 결과적으로 질의에 대한 의견은 온톨로지 사용자를 대상으로 한 워크숍인 만큼, 사용자가 만족할 만한 온톨로지를 개발한다면 그 실효성은 상당히 높을 것이라는 의견이 주를 이뤘다. 다만, 이러한 높은 실효성은 조건부로 달성할 수 있다는 점은 모두가 동의했는데, 이에 대한 조건은 반드시 사용자 집단이 형성될 수 있는 품질을 갖춰야 한다는 점이다. 다만 여기서 사용자 집단이란 불특정 다수의 집단뿐만 아니라 개발 집단 스스로가 사용자 그룹이어도 무방한 점을 분명히 한다. 따라서 만약 충분한 사용자 집단을 형성하기를 원한다면 (i) 온톨로지 자체의 범용성을 고려하거나⁶⁾, (ii) 온톨로지의 목적성을 분명히 하거나, (iii) 내부 활용을 염두에 둔 개발이 필요하다는 점이다. 다만, 관련 사업의 특성상 NRC데이터플랫폼 내에서 불특정 연구자를 대상으로 함을 지향해야 한다면, 개발자 스스로가 사용자가 될 수 있는 온톨로지를 개발해 성공 및 활용 사례를 만들고 확장하는 것을 추천했다. 이는 온톨로지를 사용하는 목적이 어떤 정

6) 범용성이 높은 온톨로지는 개발 비용이 동시에 증가하는 경향성이 있다. 이를 보완하기 위해서는 상위 온톨로지와 연계하는 ICT 도메인에 대한 하위 온톨로지 개발하는 방법도 고려할 수 있다.

교한 지식체계와의 연계에 있기에, 개발하면 사용자가 어딘가에 있을 것이라는 막연한 시도로는 사용자 만족이 쉽지 않을 것이기 때문이다.

[그림 4-1] LoG conference website



두 번째 워크숍 질의는 도메인 온톨로지 개발에 있어 참고할 만한 온톨로지 연구나 사례에 대한 추천을 받는 내용이었다. 이를 통해 도메인 온톨로지 개발 필요 사항을 확인하고 best practice를 찾고자 함이었다. 결론적으로 참여자 모두가 국내 연구나 사례를 공유하긴 했지만, 개별 사례의 우수성은 활용처나 개발 목적 또는 사용자의 니즈 등에 따라 판단의 기준이 모두 다르기에 ‘참고할 만한’ 사례에 대해서는 개발자(또는 연구자)가 선택할 사항이라는 동의했다. 다만, 한 참여자가 소개한 로그 컨퍼런스(LoG conference) 사이트와 pykeen이라는 파이썬 패키지지의 활용을 검토하면 좋은 사례나 아이디어를 얻을 수 있다고 추천했는데, 이에 관한 내용은 향후 온톨로지 관련 연구에서 충분히 참조할 수 있을 것으로 보인다. 먼저 [그림 4-1]에서 소개한 LoG conference는 그래프 학습(Learning on Graphs)에 관해 기계학습을 비롯한 광범위한 지식 그래프 학술 자료를 제공한다. 그래프에 대한 광범위한 주제를 다루는 만큼 온톨로지에 관련된 내용에만 집중하는 것은 아니지만 (i) 그래프 학습 관련 대가들이 다수 활동하고 있고, (ii) 지식 그래프가 온톨로지를 사용(포괄)한다는 점⁷⁾에서 특히 방법론이나 활용 사례를 발굴하기가 쉽다는 장점이 있다. 두 번째로 파이썬의 pykeen은 Python KnowlEdge EmbeddiNgs의 줄임말로 본래 지식 그래프 임베딩 모형을 훈련하고 평가하기 위해 설계된 패키지다(Ali et al., 2021). MIT

7) 지식 그래프와 온톨로지는 기술적으로 같은 맥락(e.g., 작성 언어, 개발 환경 등)에서 작성할 수 있지만, 지식 그래프는 일반적으로 온톨로지를 사용해 정의된 개념과 관련된 지식(e.g., 데이터, 모델, 확장 개념 등)을 모두 나타낸다. 온톨로지가 개념 구분에 초점을 맞춘다면, 지식 그래프는 개념 구분을 기반으로 더 자세한 정보를 덧붙인 체계라고 볼 수 있다.

라이선스 하에 공개적으로 제공되며, University of Bonn의 Smart Data Analytics, 프라운호퍼의 IAIS, 뮌헨 머신러닝 센터, 지멘스, 덴마크 공과대학교의 인지시스템 및 통계 데이터 분석 부문의 지원을 받아 유지·관리된다. 현재까지 과거의 기념비적인 지식 그래프 관련 자료가 모두 패키지에 수록되어 있는데, 총 37개의 그래프 데이터셋과 평가를 위한 5개의 추론용 데이터, 40개의 그래프 임베딩 모형이 이에 해당한다(〈표 4-1〉 참고).

〈표 4-1〉 pykeen 패키지 구성

구분	패키지 구성
지식 그래프 데이터 셋	○ 구분 명칭 [개체 수 / 관계 수 / 삼중항 수] - Aristo-v4 [42,016 / 1,593 / 279,425] - BioKG [105,524 / 17 / 2,067,997] - Clinical Knowledge Graph [7,617,419 / 11 / 266,691,525] - CN3I Family [3,206 / 42 / 21,777] - CoDEx(large) [77,951 / 69 / 612,437] - CoDEx(medium) [17,050 / 51 / 206,205] - CoDEx(small) [2,034 / 42 / 36,543] - ConceptNet [28,370,083 / 50 / 34,074,917] - Countries [271 / 2 / 1,158] - Commonsense Knowledge Graph [2,087,833 / 58 / 4,598,728] - DB100K [99,604 / 470 / 697,479] - DBpedia50 [24,624 / 351 / 34,421] - Drug Repositioning Knowledge Graph [97,238 / 107 / 5,874,257] - FB15k [14,951 / 1,345 / 592,213] - FB15k-237 [14,505 / 237 / 310,079] - Global Biotic Interactions [404,207 / 39 / 1,966,385] - Hationet [45,158 / 24 / 2,250,197] - Kinships [104 / 25 / 10,686] - Nations [14 / 55 / 1992] - NationsL [14 / 55 / 1992] - OGB BioKG [93,773 / 51 / 5,088,434] - OGB WikiKG2 [2,500,604 / 535 / 17,137,181] - OpenBioLink [180,992 / 28 / 4,563,407] - OpenBioLink LQ [480,876 / 32 / 27,320,889] - OpenEAFamily [15,000 / 248 / 38,265] - PharMeBInet [2,869,407 / 208 / 15,883,653] - PharmKG [188,296 / 39 / 1,093,236] - PharmKG8k [7,247 / 28 / 485,787] - PrimKG [129,375 / 30 / 8,100,498] - Unified Medical Language System [135 / 46 / 6,529] - WD50K(triples) [40,107 / 473 / 232,344] - Wikidata5M [4,594,149 / 822 / 20,624,239] - WK3I-120k Family [119,748 / 3,109 / 1,375,406] - WK3I-15k Family [15,126 / 1,841 / 209,041] - WordNet-18 [40,943 / 18 / 151,442] - WordNet-18(RR) [40,559 / 11 / 92,583] - YAGO3-10 [123,143 / 37 / 1,089,000]

구분	패키지 구성
추론 데이터 셋	○ 구분 명칭 - ILPC2022 Large - ILPC2022 Small - FB15k-237 - NELL - WordNet-18(RR)
임베딩 모델	○ 구분 명칭 - AutoSF - BoxE - Canonical Tensor Decomposition - CompGCN - ComplEx - ComplEx Literal - ConvE - ConvKB - CooccurrenceFiltered - CrossE - DistMA - DistMult - DistMult Literal - DistMult Literal(Gated) - ER-MLP - ER-MLP(E) - Fixed Model - HoIE - InductiveNodePiece - InductiveNodePieceGNN - KG2E - MuRE - NTN - NodePiece - PairRE - ProjE - QuatE - R-GCN - RESCAL - RotatE - Simple - Structured Embedding - ToursE - TransD - TransE - TransF - TransH - TransR - TuckER - Unstructured Model

자료: Github, pykeen, <https://github.com/pykeen/pykeen>(2024. 12. 15. 접속)

세 번째 워크숍 질의는 온톨로지 활용에서 불편하거나 아쉬웠던 점과 해결하고자 노력했던 방법에 관한 것이며, 온톨로지의 활용 수준, 활용한 온톨로지의 종류, 연구(과제)의 목적과 범위 등에 따라 사용자 각각이 갖는 기대나 불편의 정도가 달랐기 때문에 어떠한 결론을 내기는 힘들었다. 이와 같은 부분을 예상했음에도 이를 질문한 이유는 온톨로지 개발 목표를 설정할 때 무엇을 염두에 뒀야 하는지를 파악하고자 함이었다. 개인의 경험에 따라 다르긴 했지만, 어떤 이유에서건 온톨로지 활용에서 아쉬움을 느낀 개인이 당장 선택하는 방법은 크게 두 가지였다. 첫 번째는 온톨로지의 갱신 여부를 확인하거나 이마저 여의치 않다면 이와 유사하지만 다른 온톨로지를 찾고자 했다. 두 번째는 온톨로지 활용에서 얻을 수 있는 효용을 일부 포기하거나, 해당 온톨로지를 직접 보완하여 사용하는 방법이다. 다만, 온톨로지의 특성상 개인에게 맞춤화하기 위해 요구되는 노력의 양이 매우 크기에 당장 활용에 초점을 맞출 때는 그 시도가 쉽지 않다는 점에는 목소리를 함께 했다. 뒤집어 생각하면, 이는 결국 첫 번째 질의와 일맥상통하는 부분이 있다. 완전히 높은 범용성을 추구하거나 혹은 특수한 목적성을 추구하는 등의 목적을 분명히 할 때 온톨로지 활용의 만족도와 실효성이 증가한다는 점이다. 전자는 범용성을 고려해 사용자 스스로 넓은 범위의 활용을 추구한다면, 후자는 좁지만 분명한 니즈를 가진 사용자를 크게 만족시킬 수 있을 것이다.

네 번째 워크숍 질의는 국문으로 작성하는 온톨로지를 개발한다면 해당 온톨로지가 지향해야 할 목표의 설정에 관한 것이다. 참여자들은 입을 모아 국문으로 작성된 온톨로지도 잘 없고, 관련 연구가 있더라도 실제 데이터에 접근하기는 쉽지 않다고 이야기했다. 실제로 규모가 큰 온톨로지는 합의 표준 어휘를 기초로 하기에 일반적으로는 중앙 통제형 기간망 형식으로 관리하는데, 국제적으로 저명하고 활용도가 높은 온톨로지는 글로벌 커뮤니티 주도로 개발 및 관리되는 것이 일반적이다. 다만 안타까운 점은 이러한 거대 학술 커뮤니티에서 한국어가 차지하는 중요성이 높지 않다는 점이다. 특히, 참여자들은 국문 작성 온톨로지에 관한 연구를 찾아보면 특정 학문 도메인을 연구하는 학자들의 컴퓨터 기반 온톨로지에 관한 이해 부족으로 도메인 온톨로지 개발은 물론, 이미 구축된 글로벌 온톨로지의 활용 연구조차 찾기가 어려운 형편이라고 입을 모아 이야기했다. 대부분의 국내 온톨로지 개발 관련 연구는 해당 도메인 전문가에 의해 주도되기보다 온톨로지 작성 관련 역량을 갖춘 컴퓨터 공학 전문가가 특정 도메인에 초점을 맞춰 진행하는 경우가 많으며, 온톨로지 활용 관련 연구는 소규모 프로젝트로 국제 저명 온톨로지를 기계 번역한 후 사용하거나 표준 호환 언어 그대로 사용하는 경우가 잦다. 워크숍 참가자들은 이러한 상황에서 국문으로 온톨로지 개발을 염두에 둔다면, 관련 글로벌 온톨로지와의 호환성을 고려함은 물론 기계 번역을 통한 온톨로지 활용보다 매력적인 특징을 반드시 포함해야 할 것이라는 데 동의했다.

다섯 번째 워크숍 질의는 공공연구기관에서 온톨로지를 개발할 때 기대되는 장점과 우려되는 단점을 묻는 내용이다. 공공연구기관은 (i) 공공기관의 특성상 직접적인 경제적 부가가치 창출보

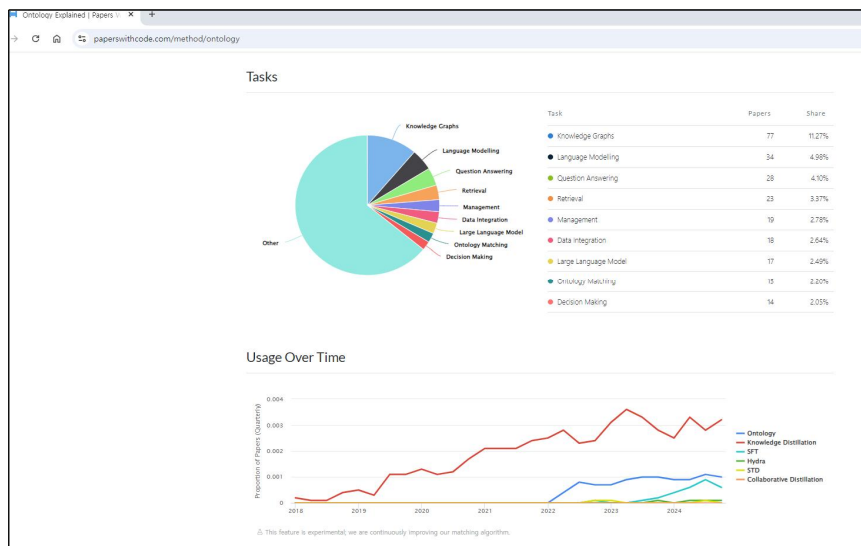
다 공공의 편익 또는 정책적 의사결정의 지원에 무게를 두고, (ii) 연구기관의 특성상 연구개발(R&D, Research and Development) 예산을 기반으로 학술적 결과물 창출을 지향한다. 따라서 온톨로지와 같이 이를 기반으로 하는 연구개발 혹은 부가가치 창출의 근간이 될 수 있는 데이터 자원을 구축하는 과업은 공공연구기관의 역할을 고려하면 적합하지 않은 것은 아니다. 다만, 워크숍 참가자들은 다음과 같은 세 가지의 현실적 제약으로 인해 공공연구기관에서 구축하는 온톨로지 개발 및 관리가 쉽지 않을 것이라고 이야기했다. 먼저 인력 구성의 문제가 있다. 연구개발 과제와 예산의 특성상 당해연도의 실적을 요구하며, 온톨로지 개발이 종료된다고 하더라도 이를 유지·보수 및 관리하는 일은 연구개발 인력이 상시 투입될 필요성이 줄어든다. 또한 온톨로지 개발에는 방대한 인적 노동력이 요구되는데 고도로 훈련된 인력으로 구성된 공공연구기관에서는 단순 개체 입력 및 검토 등에 투입할 인력 구성이 쉽지 않을 수 있다는 점이다. 두 번째는 예산의 문제로 온톨로지의 규모에 따라 다르긴 하지만 앞선 인력의 문제와 연계되어 상당한 규모의 재원이 온톨로지 개발에 요구될 수 있다는 점이다. 그러나 막대한 예산과 인력을 요구하는 것과 달리, 온톨로지라는 성과물은 직접적으로 정책의 지원 및 공공 편익 증진을 위한 의사결정에 즉각 기여하기 어렵다. 이는 연구개발 예산을 관리하는 중앙부처의 입장을 고려하면 우선순위와 필요성이 낮게 비칠 수 있다. 마지막으로 앞서 두 가지 한계점으로 인해 범용성 높은 대규모의 온톨로지 개발이 공공연구기관에서는 실질적으로 어려울 수 있기에, 사실상 상대적으로 작은 규모의 온톨로지 개발이 강제된다는 점이다. 세 번째 질의에서 논의한 내용과 국문 개발을 고려하면, 어떤 명확한 니즈를 정의하고 이를 만족시키기 위한 온톨로지 개발 정도가 현실적인 선택으로 보인다. 즉, 온톨로지 자체가 어떤 하나의 데이터 분석 모형의 하위 구성요소로 들어갈 때 가장 실효성이 높을 수 있다는 점이다.

마지막 워크숍 질의는 ICT 도메인 온톨로지 개발 후, NRC데이터플랫폼의 데이터나 서비스 등과의 연계 방식에 관한 내용이다. 이를 위해 앞선 사전 설명 단계에서 향후 NRC데이터플랫폼에서 포괄하고자 계획 중인 데이터로 특허, 국내 논문 등이 있음을 공유했다. 워크숍 참여자들은 온톨로지와 연계할 수 있는 데이터베이스의 증가는 이에 따라 늘어나는 노동 및 비용적인 측면을 제외하면 사용자 그룹의 확대와 온톨로지 실효성에 긍정적인 영향을 미칠 수 있을 것으로 평가했다. 그러나 다섯 번째 질문에서 나왔던 의견과 일맥상통한 의견 또한 있었는데, 처음부터 대규모의 연계를 기획하기보다는 작은 규모로 개발하고 활용 실적을 쌓아갈 때 기관의 특성에 따른 예산 구조, 사용자 그룹의 확산 등에서 전략적인 효과를 거둘 수 있을 것이라는 점이다.

그 외 본 연구의 배경 및 온톨로지 개발 방향에 대해 워크숍에 참여한 전문가들과 나누고 서로 동의한 의견은 다음과 같다. 첫 번째로 본 연구에 앞서 수행된 연구 흐름에 대해서는, KISDI STAT 운영 과제에서 미디어 온톨로지로 시작해서 ICT 통계 온톨로지로 변모한 방향성은 결국

온톨로지의 범위를 좁히고 실효성 있는 정의를 내리는 과정이지 않았냐는 의견이 있었다. 비록 KISDI STAT 운영 과제에서 수행한 온톨로지 개발 연구의 방향이 절차적 기획으로 변화한 흐름은 아니었지만, 워크숍에서 논의한 바와 같이 공공연구기관의 특성에 따른 예산과 인력 문제, 단기 성과 필요성 등으로 활용처가 분명한 소규모 온톨로지 개발이 권장되는 바와 일맥상통하는 부분이 있다는 것이다. 다만, NRC데이터플랫폼은 KISDI STAT 웹사이트와는 전혀 별개의 기능과 지향점을 가지고 있기에, 기존 온톨로지 결과물의 이전(migration) 및 개선은 특별한 효용성이 있기 힘들 것이라는 의견 또한 모두 동의하는 바였다. 두 번째로, 온톨로지 개발에 있어 데이터 분석 방법론과 딥러닝 기반의 언어 모형을 접목한다면 비용적인 측면을 어느 정도 해소하면서, 방법론적인 측면에서 연구개발 성격이 접목될 수 있다는 의견이다. 예를 들어 지식 그래프의 경우 paperswithcode.com에서 ontology나 graph를 키워드로 검색하면 이미 잘 만들어진 방법론과 구현 코드도 확보할 수 있고(그림 4-2) 참고), 이를 활용할 수 있다는 것이다. 실제로 앞서 개발한 LOD 형태의 ICT 통계 온톨로지 역시 이러한 방법을 염두에 두고 진행했다.⁸⁾

[그림 4-2] paperswithcode 내 검색 결과



자료: Paperswithcode, <https://paperswithcode.com/>(2024. 12. 15. 접속)

8) 단 통계 온톨로지라는 특수한 성격상 적극적인 언어 모형 활용에 제약이 있었다.

제 2 절 온톨로지 개발자 대상 설문조사

온톨로지 개발자를 대상으로 하는 설문조사는 서면으로 진행하는 만큼 워크숍 진행과 비교해 다량의 정보를 직접 전달하는 데 한계가 있었다. 따라서 워크숍과 달리 많은 정보를 주는 방향으로 온톨로지 개발의 취지와 방향만 설명하고 일반적인 의견을 묻는 방향으로 설문지를 설계했다. 이를 통해 얻은 전문가 의견은 비록 과제 상황에 대한 정보가 없더라도 들을 수 있는 보편적인 전문지식의 취합이라고 간주하고, 해석할 때 과제 상황과 연계해 의미 있는 해석을 수행하고자 한다.

첫 번째 질문은 개발자 각각의 온톨로지 개발 동기(motivation)이나 동인에 관한 것으로, 온톨로지 설문 시작과 함께 (i) 각각의 경험을 반추할 수 있게 하고 (ii) 이어질 모든 답변에서 각 전문가가 어떤 문맥에서 답하는지 이해하기 위해 사전 확인하고자 했다. 개발자 전문가 풀을 구성할 때는 도메인 온톨로지의 분야를 한정 짓지 않았기에 각각의 도메인 선정 이유 및 온톨로지 활용처는 달랐다. 예를 들면, 인문 데이터의 빠른 공유, 중복 기록 문제 해결을 위한 역사자료 표준 개발, 이종(異種)의 과학 데이터 연계 등 매우 다양함을 확인했다. 그러나, 도메인이 다르더라도 크게 세 가지의 공통점과 두 가지의 차이점을 <표 4-2>와 같이 확인할 수 있었다.

<표 4-2> 첫 번째 설문지 질의에 대한 전문가 의견의 공통점 및 차이점

구분	내용
공통점	▶ 데이터 통합 및 효율적 관리 지향 ▶ 특정 도메인 지식의 구조적 표현과 공유를 위해 온톨로지 선택 ▶ 시맨틱 웹, 인공지능, 기계학습 등과의 연계 강조
차이점	▶ 도메인에 따라 온톨로지 개발 동기가 다양 ▶ 각 분야의 특수한 문제 및 상황에 따른 접근법(방법론) 존재

자료: 저자 작성

공통점에서 확인할 수 있는 온톨로지의 효용은 크게 세 가지다. 첫 번째는 온톨로지가 데이터 통합 및 효율적 관리를 위한 도구로 사용될 수 있다는 점이다. 두 번째는 온톨로지의 표준화된 스킴(scheme)이 지식의 구조적 표현과 공유의 원활함을 기대할 수 있게 한다는 점이다. 마지막으로, 지식의 구조적 표현이 일정한 규칙에 기반을 두어 구축된 만큼, 인공지능이나 기계학습 알고리즘의 투입 데이터로 사용해 더 큰 가치의 연구 결과물을 기대할 수 있다는 점이다.

두 번째 질문은 온톨로지가 어떤 방향성과 지향점을 가지고 개발되어야 하는지에 대한 것으로, 이를 통해 온톨로지 개발 방향에 대한 전문가들의 의견을 수렴할 수 있을 것으로 기대했다. 전문가 각각이 중요하게 여기는 요인의 순서는 달랐지만, (i) 온톨로지 통합 및 확장 고려, (ii) 전문가

협력 등에 대해서는 모두가 필요하다는 견해였다. 크게 세 가지의 공통점과 두 가지의 차이점을 <표 4-3>과 같이 확인할 수 있었다.

<표 4-3> 두 번째 설문지 질의에 대한 전문가 의견의 공통점 및 차이점

구분	내용
공통점	▶ (재사용성과 호환성 강조) 온톨로지의 확장 가능성과 타 도메인과의 통합 필요성 언급 ▶ (전문가 협업) 다양한 분야의 전문가 협력을 통한 온톨로지 품질 강화 ▶ (지속적 관리) 온톨로지의 지속적인 최적화와 업데이트 중요성
차이점	▶ (기술 활용) 인공지능과 딥러닝 등 최신 기술과의 융합 필요성을 언급한 전문가 그룹과 전통적인 구축 방법론에 더 무게를 두는 전문가 그룹이 공존 ▶ (도메인 접근) 특정 도메인에 최적화된 온톨로지의 중요성을 강조하는 전문가 그룹과 범용적인 온톨로지를 지향할 때의 구축 방향이 차이가 있음을 확인

자료: 저자 작성

공통점에서 확인할 수 있는 온톨로지 개발에서 고려해야 하는 요건은 크게 세 가지다. 첫 번째는 도메인 온톨로지를 개발하더라도 확장성을 염두에 두고 타 도메인과의 통합을 고려한 설계가 필요하다. 두 번째는 온톨로지를 개발하는 과정에 있어서 목표 도메인 전문가뿐만 아니라 컴퓨터 공학, 지식 표현 등 다양한 전문가와의 협력을 바탕으로 온톨로지 품질의 시작점을 높일 수 있어야 한다는 점이다. 마지막은 온톨로지의 특성상 단순히 개발에서 그치지 않고 지속적인 최적화와 업데이트가 있을수록 실효성과 사용자 그룹의 확대를 기대할 수 있기에, 지속가능성 또한 고려해야 한다는 점이다.

<표 4-4> 세 번째 설문지 질의에 대한 전문가 의견의 공통점 및 차이점

구분	내용
공통점	● 공공성, 재사용성, 데이터 표준화 등의 온톨로지 요건을 만족하기 위해서는… ▶ 다양한 이해관계자 간의 협력이 필요하며, 특히 정부와 공공기관의 역할이 중요함 ▶ 학계의 전문성과 산업체의 실질적 응용 능력의 결합이 필요함
차이점	▶ (중심 주체) 정부·공공기관 중심을 주장하는 전문가 그룹과 학계 중심을 주장하는 전문가 그룹의 견해 차이 발견 ▶ (단계적 접근) 학회를 중심으로 시작해 민간 커뮤니티로의 확장이 필요하다는 개방성에 중심을 둔 전문가 견해가 있음을 확인 ▶ (산업 실무자 중심 개발) 산업체 참여의 비중을 강조하는 전문가 의견이 존재함을 확인

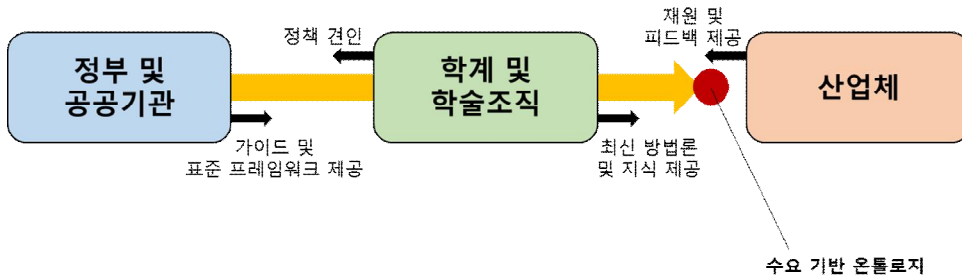
자료: 저자 작성

세 번째 질문은 온톨로지 개발 및 운영 주체는 누가 되는 것이 바람직하고 또 어떻게 관리해야 하는지에 대한 것으로, 이를 통해 공공연구기관이 개발하는 온톨로지에 대한 전문가 의견을 엿보고자 했다. 이상적으로 여기는 개발 및 운영 주체는 전문가에 따라 달랐지만, 다양한 이해관계자 간 협력이 필요하다는 점에서는 의견을 함께했다. 크게 두 가지의 공통점과 차이점을 <표 4-4>와

같이 확인할 수 있었다.

온톨로지 개발 주체에 대해 의견이 수렴하는 부분은 크게 두 가지다. 먼저, 개발 주체의 중심이 온톨로지의 공공성, 재사용성, 데이터 표준화 등을 제고하기 위해서 다양한 이해관계자 간의 협력이 중요하다는 점이다. 단, 온톨로지 영역에서 전략적인 투자와 실효성 증대를 위해서는 정부와 공공기관의 주도가 필요하다는 의견이었다. 두 번째는 학계의 전문성을 중심으로 개발을 진행하되, 산업체의 실질적 응용 능력이 뒷받침될 때 온톨로지의 효과가 극대화된다는 점이다. 이를 정리하면 [그림 4-3]과 같은데, 정부 혹은 공공기관의 가이드 및 이니셔티브를 바탕으로 학계의 전문성을 녹이고 이렇게 탄생한 온톨로지가 산업계에서 쓰일 수 있도록 전후방 공급사슬이 기획되어야 한다는 것이다.

[그림 4-3] 온톨로지 개발 및 운영 주체 간 협력 흐름



자료: 저자 작성

네 번째 질문은 온톨로지 개발 프로세스를 “기획 및 구조 설계 → 데이터 수집 → 구축 → 유지·보수·갱신”으로 구분할 때의 단계별 난이도에 관한 것으로, 온톨로지 개발을 기획할 때 기간 및 비용 산정에 참고하기 위함이다. 각 개발 단계별 평균 난이도와 그 이유는 <표 4-5>와 같다.

<표 4-5> 네 번째 설문지 질의에 대한 전문가 의견

단계	내용
기획 및 구조 설계 단계	○ 평균 난이도: 4.125 ○ 난이도 산정 이유 ▶ 온톨로지의 목적과 범위를 정의하고, 도메인 지식을 체계적으로 정리하는 작업은 상당히 어려운 의사결정 작업에 해당함 ▶ 도메인 전문가 및 온톨로지 전문가 간의 협업이 필수적이며, 초기 설계 오류가 전체 과정에 영향을 미칠 수 있어 신중한 작업 필요
데이터 수집 단계	○ 평균 난이도: 2.75 ○ 난이도 산정 이유 ▶ 크롤링 기술, 공공데이터 포털 등 데이터 확보 도구의 발전과 outlet 채널의 증가로 비교적 용이

단계	내용
데이터 수집 단계	▶ 이미 정의된 데이터 스키마에 따라 데이터 수집이 이루어지면 작업이 단순화됨 ◦ 난이도가 상승할 수 있는 경우 ▶ 전처리, 클렌징, 중복 제거 등 고품질 데이터 확보 과정이 요구되는 경우
구축단계	◦ 평균 난이도: 4.25 ◦ 난이도 산정 이유 ▶ 지식 추출, 개념화, 위계화, 공식화 등 다단계 작업이 요구되며 도메인 전문성과 반복적 검토가 필수적 ▶ 대규모 데이터 처리와 성능 최적화가 필요 ▶ 주관적 판단을 배제하고 일관성을 유지하기 위한 노력이 필요
유지·보수·갱신 단계	◦ 평균 난이도: 3.625 ◦ 난이도 산정 이유 ▶ 빠르게 변화하는 환경 속에서 지속적인 업데이트와 새로운 데이터 반영이 필요 ▶ 온톨로지의 구조적 재구성을 요구할 수 있는 상황 발생 가능 ◦ 난이도가 감소할 수 있는 경우 ▶ 온톨로지가 충분히 성숙하고 안정화 단계에 이르렀을 경우, 정형화된 프로세스를 따라 기존 온톨로지를 업데이트는 하는 작업이 비교적 단순 과업으로 변화할 수 있음

주: 난이도는 1점 이상 5점 이하 중 자유롭게 평가함

자료: 저자 작성

전문가에 따라 난이도의 순위는 조금씩 변동이 있지만, 평균 의견으로 난이도를 고려하면 구축 단계(3단계)가 가장 어려움이 크고 미세한 차이로 기획 및 구조 설계 단계(1단계)가 그 뒤를 잇는다. 세 번째 순위의 난이도는 유지·보수·갱신(4단계) 단계인 것으로 나타났고, 가장 낮은 난이도는 데이터 수집 단계(2단계)인 것으로 확인되었다. 전문가 의견의 공통점과 차이점을 정리한 내용은 <표 4-6>에 기술했다.

<표 4-6> 네 번째 설문지 질의에 대한 전문가 의견의 공통점 및 차이점

구분	내용
공통점	▶ 기획 및 구조 설계와 구축 단계가 상대적으로 높은 난이도로 평가받음 ▶ 데이터 수집은 도구의 발전으로 대부분 전문가가 낮은 난이도로 평가함 ▶ 유지·보수·갱신 단계는 중간 정도의 난이도로 평가되지만, 온톨로지의 성숙도에 따라 혹은 외부 니즈와 같은 변화에 대한 대응 능력에 따라 난이도가 변화할 수 있는 가능성을 내포함
차이점	▶ 기획 및 구조 설계와 유지·보수·갱신 단계에 대한 난이도는 전문가 별로 의견 차이가 나타나는 것을 확인할 수 있었음 - 예를 들어 기획 및 구조 설계 단계에 대해 어떤 전문가는 기획 단계가 다음 차례 단계에 미치는 영향이 크다는 이유로 높은 난이도로 평가한 반면, 어떤 전문가는 기존 온톨로지 구조와 도구 활용 가능성이 있기에 과업의 난이도는 낮다고 평가하기도 함

자료: 저자 작성

평균 난이도 점수가 아니라 개별 전문가 의견을 상세하게 살펴보면, 단계별 난이도는 도메인 지식, 기술적 숙련도, 협력 체계, 기존 온톨로지 활용 여부, 온톨로지 성숙도 등에 따라 변화할 수

있음을 확인할 수 있었다.

다섯 번째 질의는 온톨로지 개발을 어렵게 했던 제약과 해당 제약을 최소화하거나 해결했던 방법에 관한 내용이었다. 이를 통해 향후 온톨로지를 개발할 때 고려해야 할 제약 사항을 미리 파악함으로써, 온톨로지 개발 방향 수립에 적용하고자 했다. 전문가들이 이야기한 주요 제약과 해결 방안은 <표 4-7>에 기술된 바와 같다.

<표 4-7> 다섯 번째 설문지 질의에 대한 전문가 의견

구분	내용
주요 제약	● 인력 문제 ▶ 해당 분야의 심도 있는 이해와 작업 수행의 융합이 필요한데, 도메인 및 온톨로지 전문가가 부족함 ▶ 혹은 전문 인력이 있더라도, 지속적 확보가 어려움(e.g., 학생 중심의 연구팀) ● 예산과 범위 ▶ 한정적인 예산 ▶ 인건비 확보의 어려움과 고급 전문가 동원 비용 ● 데이터 관련 문제 ▶ 방대한 데이터의 처리와 품질 확보 ▶ 특정 도메인의 경우 데이터의 접근 및 수집이 어려움 ▶ 전자화 및 데이터 정제 작업의 시간과 비용 부담 ● 기술적·교육적 제약 ▶ 최신 온톨로지 자료 부족 ▶ 기술 교육 및 훈련 환경 미비
해결 방안	● 인력 강화 ▶ 도메인 전문가와 온톨로지 전문가 간 교류 기회 확대 - 융합 전문가 활용 또는 협업 네트워크 구축 ▶ 교육 과정 강화 및 실습 중심 프로그램 도입 ● 기술 도입 ▶ 최신 언어 모형 적용으로 일부 작업 반자동화 ▶ 국제적 온톨로지 호환 체계 유지 ● 데이터 수집 및 처리 개선 ▶ 협력 네트워크와 유관 기관 협조 확보 ▶ 크롤링 및 데이터 전처리 기술 활용 ▶ 공개 데이터 적극 활용 ● 체계적 관리 ▶ 정부와 공공기관 주도의 도메인별 관리 체계 구축 ▶ 프로젝트 범위의 명확한 조율과 우선순위 설정

자료: 저자 작성

온톨로지 개발의 주요 제약은 (i) 전문 인력의 지속적 확보 어려움, (ii) 한정적인 예산, (iii) 데이터의 접근 및 수집 또는 방대한 데이터의 처리, (iv) 기술 자료 부족 및 교육 환경 미비와 같이 앞서 워크숍에서도 논의되었던 사항과 유사한 것으로 보인다. 이를 해결하기 위해서는 (i) 적극적인

인 협업, (ii) 최신 기술 도입을 통한 비용 절감, (iii) 관리 체계 강화가 공통적으로 언급되었다.

여섯 번째 질의는 중소기업 온톨로지를 개발할 때 가장 큰 한계점은 무엇인지에 관한 것이다. 앞선 질의와 달리 상위 온톨로지와 같은 대규모가 아니라 상대적으로 더 적은 중소기업의 온톨로지 개발의 한계점을 물었고, 차이가 있는지 확인하고자 했다. 먼저 중소기업의 온톨로지 개발은 한정적인 예산과 맞닿아 있기에, 고급인력의 지속적 확보 및 협업 기회의 상실로 인한 품질 저하 가능성이 언급되었다. 두 번째로, 중소기업 온톨로지는 특성상 매우 좁은 특정 사례에 맞춰 개발되기 때문에 표준화와 확장성이 부족할 수 있어, 다른 시스템이나 조직과의 데이터 통합 및 상호 운용성의 저하로 이어질 수 있다는 의견도 있었다. 세 번째로는 유지 관리에 필요한 자원(i.e., 인력, 예산, 기술)의 부족으로 장기적 관점의 운영 지속성 보장이 어렵다는 의견이 있었다. 이는 앞서 <표 4-7>에서 정리한 온톨로지 개발의 문제점이 한층 더 심화하는 것으로, 특별히 중소기업의 온톨로지 개발의 문제점이라기보단 ‘중소기업의 온톨로지를 개발할 수밖에 없는 상황’이 개발 자원의 부족이라고 판단하기 때문으로 보인다.

일곱 번째 질의는 온톨로지라는 데이터 자원의 향방에 대한 견해를 묻는 것으로, 특히 생성형 인공지능 기술의 등장 이후 온톨로지의 역할과 기능 등에 관한 내용을 담고자 했다. 전문가 의견을 요약 및 정리한 내용은 <표 4-8>에 기술했다.

<표 4-8> 일곱 번째 설문지 질의에 대한 전문가 의견

구분	핵심 논점
LLM과의 관계	▶ (보완적 관계 강조) 온톨로지는 거대언어모형의 신뢰성 부족, 도메인 지식 한계, 할루시네이션 문제를 해결하는데 기여 가능 ▶ (결합 가능성) GraphRAG와 같은 기술을 통해 거대언어모형과 온톨로지가 통합되면 더욱 신뢰성 높고 구조화된 결과 생성 가능
온톨로지의 역할	▶ (도메인 전문성) 특정 도메인(e.g., 과학기술, 전쟁사 등)에서 정확하고 신뢰성 높은 지식 제공 가능 ▶ (데이터 구조화) 소셜 빅데이터, 텍스트 마이닝 등에서 데이터 관계를 구조화하여 패턴을 파악하고 인사이트도 풀 가능
온톨로지의 강점	▶ 복잡한 관계와 개념 정의, 지식 표준화 및 데이터베이스 구축에 뛰어남 ▶ 자동화 학습이 관계 확장에서 거대언어모형과 상호보완적 소임을 수행은
향후 방향	▶ (다층 온톨로지) 단순한 의미 연결을 넘어 다양한 의미 연계를 반영한 고급 온톨로지 필요 ▶ (융합 연구) 기존 온톨로지와 인공지능 기술을 결합해 새로운 형태의 지식체계 개발 가능

자료: 저자 작성

전문가들은 온톨로지가 인공지능 기술과 거대언어모형의 발전 속에서도 여전히 중요한 데이터 자원으로 남을 것이라는 데는 동의했다. 특히, 전문성과 신뢰성이 요구되는 분야에서 온톨로지는 거대언어모형의 한계를 보완하고, 복잡한 지식 구조화와 데이터 분석에 유용한 도구로 활용될 가

능성이 클 것으로 기대했다.

마지막 여덟 번째 질의는 향후 ICT 도메인 온톨로지가 구축된다면 가장 크게 이바지할 수 있는 영역이나 문제는 무엇인지에 대한 것으로, 기존에 개발했던 미디어 온톨로지의 활용 방안 또한 함께 물었다. 이에 대한 전문가 답변은 <표 4-9>와 같았다.

<표 4-9> 여덟 번째 설문지 질의에 대한 전문가 의견

구분	내용
ICT 온톨로지의 주요 활용 방향	● 지식체계 정립 및 통합 ▶ ICT 분야의 용어 및 개념 간 관계를 체계적으로 정리하여 기술 문서와 관련 데이터베이스 활용성을 높일 수 있음 ▶ 스마트시티, IoT, 의료 데이터 등의 넓은 분야에서 정보 통합과 문제 해결 촉진 가능 ● 정보검색과 학습 지원 ▶ 의미 기반 검색을 통해 검색 정확성과 효율성 강화 ▶ ICT 부문의 전문 용어 표준화로 언어장벽 극복 및 글로벌 지식 공유 촉진 ▶ 초중고 교육 및 ICT 전문 교육에서 학습자 맞춤형 자료제공 ● 인공지능과의 융합 ▶ 급변하는 ICT 분야에서 인공지능 시스템의 신뢰성과 정확성을 보완 가능 ▶ 거대언어모형과의 결합을 통한 자동 학습 및 관계 확장으로 ICT 도메인의 지식체계 고도화 ● 정책 수립 및 기술 발전 ▶ ICT 정책 수립과 기술 간 연관관계 분석 지원 가능 ▶ 연구자와 기업의 ICT 기술 경쟁력 확보를 위한 지식 인프라 제공
미디어 온톨로지의 활용 방안	▶ (다중 데이터 형식 통합) 다양한 포맷의 미디어 데이터(e.g., 텍스트, 이미지, 비디오 등) 형식을 통합적으로 관리하여 풍부한 정보 제공 ▶ (맞춤형 콘텐츠 제공) 사용자 선호도에 따른 미디어 콘텐츠 추천 및 학습 자료 제공 ▶ (디지털 저널리즘 강화) 사실 검증 및 정보의 관계성을 명확히 하는 데 기여 ▶ (검색 및 분석 고도화) 연관성 높은 미디어 콘텐츠 검색과 데이터 분석 효율화

자료: 저자 작성

ICT 온톨로지는 ICT 영역의 지식체계의 정립과 통합, 정보검색과 학습 지원, 정책 수립 및 기술 발전 등 다양한 영역에서 핵심적인 역할을 수행할 잠재력을 갖춘 것으로 판단된다. 또한 미디어 온톨로지는 다중 데이터 형식 통합과 콘텐츠 맞춤화 등을 통해 미디어 산업의 혁신을 지원할 수 있을 것이라는 의견을 들었다. 특히, 두 가지 온톨로지 모두 인공지능 기술과의 융합을 통해 지능화된 ICT 및 미디어 서비스와 미래 지식 기반 시스템의 중심으로 자리잡을 것으로 기대된다는 의견을 확인할 수 있었다.

제 3 절 소결

본 연구에서는 온톨로지 사용자 그룹을 대상으로 한 워크숍과 온톨로지 개발자 그룹을 대상으로 한 설문조사를 수행했다. 먼저 워크숍을 수행해 사용자 관점에서 (i) 공공연구기관에서 국문 온톨로지를 개발할 때 어떤 요인을 고려해야 실효성이 높을 수 있을지에 대한 의견을 청취하고, (ii) 개발자 그룹을 대상으로 하는 설문지 항목 설계를 위한 기타 논의 등을 수행했다. 이어 설문조사에서는 (i) 온톨로지 개발의 방향성, (ii) 개발 주체, (iii) 개발상의 제약과 해결 방안, (iv) ICT 도메인 온톨로지의 활용처 등에 대한 의견을 물었다.

먼저 워크숍에서는 과제 상황은 물론 예산 및 인적 자원의 한계점 등을 공유한 후 논의를 수행했기에 사용자 관점에서는 “비록 사용자 규모는 작더라도 그 목적성이 분명하기에 ICT 도메인에서 특정 분석 목표를 가진 경우 그 필요성이 매우 높을 수 있는 온톨로지” 개발에 대한 선호도가 높았다. 이는 온톨로지라는 거대한 데이터 자원 개발에 드는 물적·시간적 투입 요소가 막대하기에, 오히려 거대한 정책적 지원이 없는 상황에서 큰 목표 대비 매년 작은 진척도로 온톨로지 개발을 수행하면 개발 욕구의 감소 및 성과의 부진으로 이어지는 것을 우려하기 때문으로 보였다. 특히 워크숍 참석 전문가들은 국문 온톨로지가 저명한 글로벌 온톨로지에서는 찾기 힘들기에, 온톨로지를 적용한 작은 성과를 창출해 가면 그것만으로도 국내에서는 특징점이 있을 것이라고 이야기했다. 다만, 온톨로지의 확장성과 개방성은 매우 중요한 요인이기에 이러한 확장성을 고려한 개발은 염두에 두고 있어야 함을 동시에 주장했다.

설문조사에서는 비록 전문가 각각의 의견 차이는 관찰되고 있지만, 온톨로지 개발의 전반적인 방향이나 흐름에 대해서는 같은 시선을 유지하고 있음이 확인되었다. 먼저 온톨로지 개발의 방향성에 대해서는 온톨로지의 재사용성과 호환성 및 체계적이고 지속적인 협업 필요성을 강조했다. 개발 주체 관점에서는 다양한 이해관계자의 협업을 중시하면서도 정부 및 공공기관의 역할 또한 중요하다고 했는데, 단순히 온톨로지의 개발 및 활용에서 그치지 않고 학계와 민간 산업계 등의 연쇄적 활용을 통한 부가가치 창출이 중요함을 주장했다. 개발상의 주요 제약으로는 인력, 예산, 데이터 품질 확보, 기술·교육적 어려움이 언급됐는데, 이를 해결하기 위해서는 적극적인 협업, 최신 기술 도입을 통한 비용 절감, 관리 체계 강화가 필요하다고 했다. ICT 도메인 온톨로지의 활용처에 대해서는 다양한 방면에서 효과가 있을 것이라는 평이 있었다.

제5장 결론

1. 결론 및 시사점

본 정보통신정책연구원이 주관하는 「ICT데이터플랫폼 구축 운영사업」의 예산 감축으로 인해 중지된 온톨로지 개발 연구를 「데이터 기반 미래예측·정책지원 모델연구」에서의 수행 여부 의사 결정을 위해 문헌 및 사례 분석, 워크숍과 설문조사를 통한 전문가 의견 수렴 등을 진행했다.

제2장에서는 먼저 온톨로지의 개념, 구성요소, 작성 언어, 범위에 대한 정의 및 구분을 통해 온톨로지에 대한 기본 이해를 돕고자 했다. 이를 통해 온톨로지 개발 방향을 논의하기에 앞서 온톨로지에 관한 배경지식을 정리했다. 이어서 국내외 온톨로지 사례를 정리했는데, 특정 도메인에 대한 온톨로지 활용 사례는 있었지만, ICT 도메인을 포괄하는 온톨로지는 확인하지 못했다.

제3장에서는 연구 목적 달성을 위해서 어떤 순서와 절차에 따라 방법론을 활용했는지, 워크숍과 설문조사를 수행하면서 정성 분석을 위해 어떤 전문가를 대상으로 연구를 수행하고자 했는지를 정리했다. 이러한 연구 프레임워크 정리를 통해 본 연구의 방법론적 논리를 탄탄히 하고, 과학적인 분석 절차를 소개했다.

제4장에서는 실제 워크숍 및 설문조사 수행 결과를 정리했다. 먼저 워크숍은 온톨로지 사용자 그룹을 대상으로 진행했으며, “비록 사용자 규모는 작더라도 그 목적성이 분명하기에 ICT 도메인에서 특정 분석 목표를 가진 경우 그 필요성이 매우 높을 수 있는 온톨로지” 개발에 대한 선호도가 높고 동시에 온톨로지의 확장성과 개방성 또한 고려해야 한다는 것을 알 수 있었다. 워크숍 때 논의한 의견을 바탕으로 설문조사를 설계한 후에는 온톨로지 개발자 그룹을 대상으로 설문조사를 수행했다. 재미있게도 워크숍과 설문조사 내용을 정리하면, ICT 도메인 온톨로지를 개발하되, 인력 및 예산상의 제약이 크고 당해연도 성과를 내기 어려운 큰 규모의 온톨로지가 아니라 작더라도 활용처가 분명한 도메인 온톨로지 개발을 지향해야 한다는 결과가 나온다, 단 그 규모가 작더라도 온톨로지의 중요한 성공 요인인 개방성과 확장성을 반드시 고려해야 한다는 점은 눈여겨보아야 할

것으로 보인다.

본 연구의 시사점은 총 두 가지로 정리할 수 있다. 먼저, 국문으로 작성된 ICT 분야의 도메인 온톨로지를 개발한다면 반드시 작은 목적성이라도 분명히 해서 착수하되 중장기적인 기획 속에서 지속적인 발전 및 유지·보수를 수행해야 한다는 점이다. 이는 연구개발을 위한 연구개발 목표로 온톨로지 개발을 착수하면 오히려 지속가능성과 활용 사례 도출이 어렵기에 중장기적인 개선이 어려워질 수 있기 때문이다. 두 번째로 국문으로 작성된 ICT 온톨로지는 사용자 입장에서 일정 수준 이상의 품질을 확보할 수 있다면, 그 활용처가 매우 넓다는 점이다. 따라서 이러한 데이터 자원을 마련하는 연구에 공공연구기관이 뛰어들다는 점에 있어서는 고급인력을 지속적으로 유지할 수 있고, 글로벌 협력 네트워크 구축의 허들이 낮다는 점에서 그 실효성이 높을 것으로 기대된다.

2. 한계점 및 해결 방안

본 연구는 온톨로지 개발의 방향에 대한 검토를 위해 다양한 전문가 의견을 확인하며 진행했다. 본 연구를 기반으로 하면 향후 비록 다른 도메인 온톨로지라고 하더라도 개발을 위한 기본 방향 및 주의할 점을 유의할 수 있지만, 실질적으로 ICT 온톨로지 개발에 직접적인 진입을 위한 연구 구성은 아니었다. 공공연구기관이라고 하더라도 예산의 변동성에 항상 노출되어 있을 뿐만 아니라, 작은 규모의 성공사례라도 쌓기 위해서는 실제 어떤 방식으로 온톨로지를 활용할 것인지가 논의되어야 하며 이를 위해서는 온톨로지의 개방성과 확장성을 고려하면서도 특수한 온톨로지 활용처를 먼저 계획하는 것이 필요하다. 따라서 NRC데이터플랫폼을 운영하면서 온톨로지 개발에 대한 것은 긍정적인 잠재력이 있다는 정도로 두고, ICT 관련 데이터 분석 모형 개발에 어떤 온톨로지가 필요할지를 고민해야 하는 시간적 여유가 요구된다.

참고문헌

■ 국내문헌

- 고건우·김선우·박성진·노윤주·최성필(2019), “대규모 범죄 수사기록을 활용한 온톨로지 기반 서비스 구현-침입 절도 범죄 분야를 중심으로”, 《한국문헌정보학회지》, 53(1), pp.57-81.
- 고영만(2006), “시소러스 기반 온톨로지에 관한 연구”, 《정보관리》, 5, pp.5-22.
- 김은경·김용기(2017), “온톨로지를 이용한 식품첨가물 정보 지식의 구축”, 《한국지능시스템학회 논문지》, 27(1), pp.42-49.
- 김정민·최병일·김형주(2005), “텍스트 내용 지식 기반의 철학 온톨로지 구축”, 《정보과학회논문지: 컴퓨팅의 실제 및 레터》, 11(3), pp.275-283.
- 김태영·강주연·김혜영·김용(2017), “LOD 기반의 재난안전 정보서비스 확장에 관한 연구”, 《한국문헌정보학회지》, 51(3), pp.163-188.
- 김형주·김종덕·이중식·정준희·조나현·오지연·홍상표·한창진(2008), 『IPTV 환경에서의 방송콘텐츠 온톨로지 구축』, 방송통신위원회 자유 2008-06, 방송통신위원회.
- 노상규·박진수(2007), 『온톨로지(ontology)』, 가즈토이.
- 배지영·조미숙(2016), “소셜빅데이터 분석을 위한 식품안전 온톨로지 구축”, 《보건정보통계학회지》, 41(3), pp.278-289.
- 유기동·황현석(2012), “온톨로지를 이용한 프로세스 기반 지식지도 구축”, 《한국산업정보학회논문지》, 17(4), pp.87-97.
- 윤진혁·유상봉·김인한(2003), “건축 용어의 온톨로지 구축과 활용”, 《한국전자거래학회지》, 8(2), pp.49-62.
- 정희선·김희순·송현숙·이명희(2013), “종교유적 건축물 정보의 메타데이터 구성과 온톨로지 구축”, 《한국도서관·정보학회지》, 44(1), pp.5-26.
- 최익규(2013), “온톨로지 자동구축을 위한 지능형 계층 관계 추출 시스템 연구”, 아주대학교 박사학

위 논문.

황금하·신지애·최기선(2008), “개념 및 관계 분류를 통한 분야 온톨로지 구축”, 《정보과학회논문지 : 소프트웨어 및 응용》, 35(9), pp.562-571.

■ 해외문헌

- Ali, M., Berrendorf, M., Hoyt, C. T., Vermue, L., Sharifzadeh, S., Tresp, V., & Lehmann, J. (2021). “PyKEEN 1.0: a python library for training and evaluating knowledge graph embeddings.” *Journal of Machine Learning Research*, 22(82), pp.1-6.
- Bergman, M., K. (2018). *A Knowledge Representation Practionary: Guidelines based on Charles Sanders Peirce*. Springer Charm.
- Borgo, S., Ferrario, R., Gangemi, A., Guarino, N., Masolo, C., Porello, D., Sanfilippo, E. & Vieu, L. (2022). “DOLCE: A descriptive ontology for linguistic and cognitive engineering.” *Applied Ontology*, 17(1), pp.45-69.
- Brachman, R. J., & Schmolze, J. G. (1985). “An Overview of the KL-ONE Knowledge Representation System.” *Cognitive Science*, 9(2), pp.171-216.
- Gangemi, A., Guarino, N., Masolo, C., & Oltramari, A. (2001). “Understanding top-level ontological distinctions.” In *Proceedings of the Intenational Journal of Computer Applications Workshop on Ontologies and information sharing*.
- Gruber, T. (1993). “What is an Ontology.”
- Guarino, N. (1997). “Semantic matching: Formal ontological distinctions for information organization, extraction, and integration.” In *Information Extraction A Multidisciplinary Approach to an Emerging Information Technology: International Summer School, SCIE-97 Frascati, Italy, July 14-18, 1997*, pp. 139-170, Springer Berlin Heidelberg.
- Herre, H., Heller, B., Burek, P., Hoehndorf, R., Loebe, F., & Michalek, H. (2006). “General formal ontology (GFO)-part I: Basic principles.” *Onto-Med Report*, 8.
- Kim, M., Cobb, J., Harrold, M. J., Kurc, T., Orso, A., Saltz, J., Post, A., Malhotra, K., & Navathe, S. B. (2012, July). “Efficient regression testing of ontology-driven systems.” In *Proceedings of the 2012 international symposium on software testing and analysis*, pp.320-330.
- Maniraj, V., & Sivakumar, R. (2010). “Ontology languages-a review.” *International Journal of Computer Theory and Engineering*, 2(6), pp.887.-891.

- Mascardi, V., Cordi, V., & Rosso, P. (2007). "A Comparison of Upper Ontologies." In *Workshop from Objects to Agents*, pp.55-64.
- Mealy, G. H. (1967). "Another Look at Data." In *Proceedings of the November 14-16, 1967, fall joint computer conference*, pp.525-534.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). "Introduction to WordNet: An on-line lexical database." *International Journal of Lexicography*, 3(4), pp.235-244.
- Noh, H., Seo, J. H., Yoo, H. S., & Lee, S. (2018). "How to improve a technology evaluation model: A data-driven approach." *Technovation*, 72, pp.1-12.
- Noy, N. F., & McGuinness, D. L. (2001). "Ontology development 101: A guide to creating your first ontology.",
https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html(2024. 12. 10. 접속).
- Sowa, J. F. (1999). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co.
- Sowa, J. F. (1992). "Semantic networks." *Encyclopedia of artificial intelligence*, 2, pp. 1493-1511.
- Timm, J. T., & Gannod, G. C. (2005, July). "A model-driven approach for specifying semantic web services." In *IEEE International Conference on Web Services (ICWS'05)*, pp. 313-320.
- Van Rees, R. (2003). "Clarity in the usage of the terms ontology, taxonomy and classification." *Cib Report*, 284(432), pp.1-8.

■ 홈페이지

- Palantir, "Why creat an Ontology?"
<https://www.palantir.com/docs/foundry/ontology/why-ontology/>(2024. 12. 10. 접속).
- LoG conference website, <https://logconference.org/>(2024. 12. 15. 접속).
- Github, pykeen, <https://github.com/pykeen/pykeen>(2024. 12. 15. 접속).
- Paperswithcode, <https://paperswithcode.com/>(2024. 12. 15. 접속).
- LOD 국내 서비스 현황(2024. 5. 17. 기준) .
- 국립중앙도서관 국가서지LOD, <https://lod.nl.go.kr/home/>.
 - 국토안전관리원 공공시서물 안전정보 LOD, <https://lod.fms.or.kr/main/>.

- 부산 영화 영상 관광 Linked Open Data, <https://data.busan.go.kr/lod/main.do>.
- 서울 열린데이터 광장, <http://lod.seoul.go.kr/home/intro/openLod.jsp>.
- 외교부 OPEN DATA, <https://opendata.mofa.go.kr/lod/ontologyModel.do>.
- 청년일자리 LOD서비스(기획재정부), <https://opendata.alio.go.kr/lod/browser>.
- 한국관광공사, https://data.visitkorea.or.kr/linked_open_data.
- 한국사 LOD(국사편찬위원회), <http://lod.koreanhistory.or.kr/lodIntro.do>.
- LOD생물정보통합시스템, <http://lod.nature.go.kr/service/wildlife/intro.do>.
- Ontology for Media Resource, <https://www.w3.org/TR/mediaont-10/>.
- RISS Linked Open Data, <https://data.riss.kr/serviceHome.do>.
- Worknet Linked Open Data, <https://lod.work.go.kr/>.

부 록

온톨로지 개발 방향 연구(온톨로지 개발 경험자 대상)

바쁘신 가운데 시간을 전문가 조사에 시간을 내어 주셔서 감사드립니다.
정보통신정책연구원 ICT통계정보연구실 연구진은 정보통신정책연구원의 과제인
"데이터정보시스템을 위한 온톨로지 개발 방향 연구"를 수행하고 있습니다.

정보통신기술 분야에서 기계학습 등 인공지능 기술이 중심 기술로서 도입됨에 따라
관련 용어 간의 관련성을 통합적으로 관리, 활용할 수 있는 종합적인 원시자료 데이터
자원이 필요하게 되었습니다.

이 연구는 ICT 관련 온톨로지 개발의 실효성과 활용도를 제고하고자 합니다.

ICT 온톨로지 개발의 의미, 한계점, 향후 전망 및 방향 등 관련 주제에 대한
전문가분들의 의견을 얻고자 합니다.

해당 주제가 최신성이 있고 주제의 복잡성이 있어 최대한 포괄적인 질문 형태로
문항을 구성했습니다. 답변이 가능하신 문항에 대해 의견을 주시면 연구에 큰 도움이
되겠습니다.

노희용 연구위원

※ 조사관련 문의: heeyongnoh@kisdi.re.kr

온톨로지(ontology) 관련 참고내용

○ 온톨로지(ontology)에 대한 기본개념

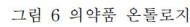
- 존재하는 사물과 사물 간의 관계 및 여러 개념을 컴퓨터가 처리할 수 있는 형태로 표현하는 것을 의미한다. 스탠포드 대학의 그루버(Thomas A. Gruber) 교수는 온톨로지를 ‘일반적으로 개념화된 것을 형식적으로 명백하게 기술하는 명세(an explicit formal specification of a shared conceptualization)’라고 정의하였다(1992). 어떤 일정 범위에서 사용되는 단어들의 개념, 특성, 연관 관계 등을 표현하여 단어에 대한 일반적 지식이 명시적으로 드러나고, 단어 간 관계 정의를 통해 문장의 의미를 파악할 수 있다(정보통신용어사전).
- 실세계(도메인)에 존재하는 모든 개념들과 개념들의 속성, 그리고 개념들이 상호간 의미적으로 어떻게 연결되어 있는가에 대한 정보를 가지고 있는 지식베이스로서 관심 영역(도메인)에 대한 형식적이고 명시적인 명세화이다.
- 특정한 분야(domain)에 속하는 개념과 개념 사이의 관계를 기술하는 정형(formal)어휘의 집합을 의미한다.

○ 온톨로지 구축을 위한 말뭉치(Corpus,코퍼스) 구축 필요

- 말뭉치(Corpus, 코퍼스): 말뭉치란 컴퓨터가 읽을 수 있는 형태로 입력하고 분석한 대규모 언어 빅데이터로, 말뭉치는 쓰임에 따라 종류가 다양하다. 특히, 원시 말뭉치는 다른 정보 없이 단순히 입력된 자료로 모든 분석 말뭉치를 만들기 위해 필요한 자료를 의미한다. 온톨로지 구축을 위해서는 관련 분야의 말뭉치(코퍼스)가 우선적으로 확보되어야 한다.

○ 온톨로지 구축 관련 연구 및 사례 예시

- 손종수,김도형,정인정(2010) 온톨로지를 이용한 의약품 정보 지식의 구축 사례
 - ✓ 의약품 정보들의 관계를 표현하여 의약품 정보의 효율적인 검색 및 광범위한 활용이 가능하게 하기위한 연구를 진행함
 - ✓ 본 논문에서는 한국식품의약품안전청에서 제공하고 있는 7개의 웹사이트에서 제공되는 23종의 의약품 관련 정보를 수집 분석하여 온톨로지 기술을 활용하여 여러 웹사이트에 흩어져 있는 의약품 정보의 연결성을 분석하고 표현함
 - ✓ 온톨로지 범위 설정 → 핵심 개념 확립 → 클래스와 속성 관계 정립 → 온톨로지 생성 및 보완 단계를 거침



- 나민영, 양경용(2009) 6단계 상향식 국방 온톨로지 추출
 - ✓ 국방 훈련분야 자료를 대상으로 핵심용어를 추출한 후 범용 온톨로지 매핑 프로그램인 Protege를 사용하여 국방 온톨로지를 추출
 - ✓ (1단계) 온톨로지 도메인/범위 결정 →(2단계) 핵심 용어 추출 →(3단계) 중복용어/의미모호 용어 정리 및 동의어/유사어 정리 →(4단계) 클래스 정의 →(5단계) 계층구조 형성 →(6단계) 계층구조 세부 정리



Fig. 4. National Defense Educational Training(NDET) Ontology

- 김희순 외(2013). 문화유산정보의 온톨로지 기반 코퍼스 생성을 통한 디지털 콘텐츠화 방안
 - ✓ 문화유산의 콘텐츠화를 위하여 각종 유산정보를 수집하고 가공하여 디지털 정보로 전환 하는 방안을 제시
 - ✓ 문화유산 가운데 구 한성부 경역에 현존하는 종교건축유산을 정보 수집 대상으로 하였으며, 대규모 데이터를 통합하여 정보 간의 의미 관계를 활용하는 방법으로서 온톨로지(ontology)를 적용
 - ✓ 종교건축유산정보를 기술하는 텍스트를 분석하여 개체명 및 이벤트명을 분류하고 태그를 부착

1. 온톨로지 개발의 의미 및 필요성

1-1. (온톨로지 개발 동기) 온톨로지를 개발하게 된 동인은 무엇입니까?

1-2. 온톨로지는 어떤 방향성과 지향점을 가지고 개발되어야 한다고 생각하십니까?

1-3. 온톨로지 개발 및 운영 주체(예를 들어 정부, 공공기관, 학계, 산업체, 시민사회 등)는 누가 되는 것이 바람직하며 어떤 방식으로 관리되어야 한다고 생각하십니까? 또한 그렇게 생각하시는 이유는 무엇입니까?

〈개발 운영 주체 및 방식〉

〈이유〉

II. 온톨로지 개발의 제약 및 한계점

2-1. 온톨로지 개발 프로세스를 “기획 및 구조 설계 → 데이터 수집 → 구축 → 유지·보수 (버전 업)”으로 구분한다면

단계별 난이도에 대해 어떻게 생각하십니까? 그 이유는 무엇입니까?

〈단계별 난이도〉

개발 단계	기획 및 구조 설계	데이터 수집	구축	유지보수(버전업)
난이도				

※ 난이도 표기는 1부터 5까지 숫자로 표시해주시기 바랍니다.
 ※ 단계별 난이도가 같다면 중복 값 표기도 가능합니다.

〈이유〉

2-2. 온톨로지 개발을 어렵게 했던 제약(e.g., 데이터 수집, 노동 집약적 과업, 예산(인건비 포함) 등은 무엇이었습니다?

또한 해당 제약을 최소화 또는 해결했던 방법은 무엇이었습니다?

〈제약 사항〉

〈최소화 또는 해결 방안〉

2-3. 중소기업의 온톨로지 개발이 수행될 때 나타날 수 있는 가장 큰 한계점은 무엇이 있을까요?

III. 온톨로지 개발의 향방 및 전망

3-1. 국제무대에서 여전히 온톨로지 개발 및 활용에 대한 투자(e.g., 미국의 disease ontology 등)가 진행되는 사례도 있지만, 인공지능 기술의 발달과 거대언어모형(large language model)의 등장으로 온톨로지의 향후 필요성에 대한 의구심 가득한 의견이 나오고 있는 것도 사실입니다. 온톨로지라는 데이터 자원의 향방에 대해 어떤 견해를 가졌는지 서술 부탁드립니다.

3-2. 향후 ICT 온톨로지가 구축된다면 가장 큰 도움을 줄 수 있는 영역 혹은 해결에 도움을 줄 수 있는 문제 등은 어떤 것이 있을까요? 향후 미디어 온톨로지의 활용방안에 대해 자유롭게 의견 주시기 바랍니다.
(e.g., 지능화된 E-Biz분야에 활용, 검색 고도화, 공통사전, 정보 이용 공통방법 제공, DB의 주석과 기술적 문서 이해 등)

정책자료 24-05-03

NRC데이터정보시스템을 위한 온톨로지 개발 방향 연구

인 쇄: 2024년 12월

발 행: 2024년 12월

발행인: 김 정 언

주 소: 27872 충청북도 진천군 덕산읍 정통로 18

전 화: 043-531-4114

팩 스: 043-535-4695~6

I S B N: 979-11-7000-413-4 / 94320