



환경·생태 지식베이스의 국제적 조화를 위한 한국형 표준 식물 온톨로지 개발

주은정¹ , 이헌주^{2*}

¹서울교육대학교 과학교육과, ²캠아이넷(주)

Development of Standardized Korean Plant Ontology for International Harmonization of Environmental and Ecological Knowledge Bases

Eunjeong Ju¹ and Hunjoo Lee^{2*}

¹Department of Science Education, Seoul National University of Education, ²CHEM. I. NET, Ltd.

ABSTRACT

Background: To describe domain knowledge consistently and precisely, the establishment of a controlled vocabulary, a so-called ontology, is essential. Internationally, the plant ontology (PO) in the ecology field has been developed for the anatomy and developmental stages of plants in English, Spanish, and Japanese, but there is no Korean version of the PO due to a lack of knowledge on standardization for Korean plants.

Objectives: We aimed to establish a Korean plant ontology with core PO architectures.

Methods: The latest ontology web language (OWL)-formatted raw version of the PO was collected from the PO consortium site. A formal workflow process and OWL file-handling tools for efficient Korean content development were conducted and executed.

Results: The macro- and micro-perspective frameworks of the PO were presented by analyzing the upper model and the internal OWL-leveled physical structure, respectively. We developed and validated Korean knowledge content for a total of 1,957 classes included in the PO and transplanted them into an ontology modeling system.

Conclusions: A Korean plant ontology was established for international harmonization through improved compatibility and data exchangeability with multilingual environmental and ecological knowledge bases.

Key words: Biological ontologies, knowledge bases, plants, semantics, vocabulary

Received June 29, 2023

Revised July 24, 2023

Accepted July 25, 2023

Highlights:

- The architectural frameworks of the plant ontology (PO), a global ontology, were projected.
- A semantic-based Korean knowledge bases of plants harmonized with the PO was developed.
- The developed Korean plant ontology will support database construction with high compatibility.

*Corresponding author:

CHEM. I. NET, Ltd., 43 Mokdong-jungang-ro, Yangcheon-gu, Seoul 07964, Republic of Korea

Tel: +82-2-2647-4930

Fax: +82-2-2647-4932

E-mail: adstar@cheminet.kr

I. 서 론

환경과 보건 분야에 다양한 종류의 대용량 데이터(예, 환경, 생태, 노출, 건강영향, 인구특성 등)를 융합하여, 인류 건강의 유지·증진을 위한 새로운 통찰력을 만들어내는 환경보건학의 특성상, 공신력 있는 데이터소스(예, 학술문헌, 데이터베이스 등)의 확보 및 이종 데이터소스 간의 통합 과정은 매우 필수적이다.¹⁻³⁾ 그러나, 데이터소스가 서로 다른 목적, 주체, 수준 및 형식으로 생산되다 보니, 서로 다른 어휘로 표기된 데이터소스

내의 콘텐츠를 검색하여 확보하는 단계는 물론 데이터 정제, 가공 및 합성 등 통합 과정에 이르기까지 데이터 호환성 확보를 위한 수작업 부하가 상당히 많이 발생하고 있다.^{4,5)} 이와 같은 문제를 해결하는 수단으로써, 디지털 정보에 관한 인간과 컴퓨터가 모두 이용 가능한 형식의 지식 표현 모델 및 의미(semantic) 기반의 검색 기술이 다양하게 제안되어 왔다. 이중 대표적이지 가장 진보된 지식표현 모형이 온톨로지(ontology)이다.⁶⁾ 온톨로지는 학문 분야 별(예, 법률, 의학, 생물학 등)로 사용되는 어휘들을 대상으로 구축되며, 단순한 어휘집이 아닌 어휘를



의미와 관련된 클래스, 속성 및 어휘간 관계망으로써 지식을 구조화하여 표현하는 방식이다.⁷⁻⁹⁾

온톨로지는 합의 표준 어휘의 개념을 기초로 하고 있어, 중앙 통제형 기간망(backbone) 형식으로 관리되고 있다. 국제적으로 활용도가 매우 높은 온톨로지는 민간 커뮤니티 주도의 이니셔티브(initiative)인 개방형 생물의학 온톨로지(Open Biological and Biomedical Ontologies, OBO)이다. OBO는 OBO 파운드리에서 개발 및 운영되고 있으며, 다양한 분야의 온톨로지가 색인화되어 있다. 환경보건 분야는 환경 온톨로지(Environmental Ontology, ENVO), 식물 온톨로지(Plant Ontology, PO), 동물 온톨로지(Uber Anatomical Ontology, UBERON), 질병 온톨로지(Disease Ontology, DOID), 미국생물공학센터 텍소노미(NCBI Taxonomy, NCBITaxon), 유전자 온톨로지(Gene Ontology, GO), 공통 해부학 참조 온톨로지(Common Anatomy Reference Ontology, CARO) 등이 있다.¹⁰⁾ 이중 PO는 환경·생태 분야에서 매우 중요하게 취급되는 식물에 관한 온톨로지로서, 다양한 데이터 기반 생태학 연구에 활용되고 있다.¹¹⁾ 초기에는 중요한 모델 식물 종인 *Arabidopsis thaliana*, *Zea mays* 및 *Oryza sativa*에 대한 내용을 위주로 개발되었으나 최근 식물 과학의 발전으로 인해 녹조류부터 속씨 식물에 이르는 모든 녹색 식물(veridiplantae) 관련 어휘를 포괄하고 있다.¹²⁾

PO는 영어를 대표 어휘로 하여 구축되어 있고, 스페인 및 일본이 참여하여 자국의 식물 어휘 콘텐츠를 개발하여 연계하고

있다. 그러나, 아직 국내는 한글 식물 온톨로지 콘텐츠 표준 개발의 부재로 연계망이 구축되어 있지 않은 상황이다. 또한, 활용의 주체인 환경·생태학자들의 컴퓨터 기반 온톨로지에 관한 이해 부족으로 국외와는 달리 이미 구축된 PO 활용 연구가 전무한 실정이다. 따라서, 본 연구에서는 국제적인 온톨로지인 PO를 심층 분석하여 핵심 프레임워크를 기술하고, PO와의 호환성을 가진 한국형 식물 어휘에 관한 지식 온톨로지 콘텐츠 구축 결과를 제시하고자 한다.

II. 재료 및 방법

1. PO 프레임워크 심층 분석 및 한국형 식물 지식 온톨로지 구축 대상 결정

PO의 구조적 분석을 위해 국제적인 통합 온톨로지 허브인 OBO의 국제화된 자원 식별자(International Resource Identifier, IRI)를 기반으로 한 온톨로지 탐색 도구인 유럽분자생물학연구소-유럽생물정보학 연구소(The European Molecular Biology Laboratory - the European Bioinformatics Institute, EMBL-EBI) 온톨로지 검색 서비스(Ontology Lookup Service, OLS) 및 미국 국립 생체의학 온톨로지 센터(The National Center for Biomedical Ontology, NCBO) BioPortal을 활용하였다. EMBL-EBI OLS 및 NCBO BioPortal 서비스를 통하여 PO 온톨로지 세트를 조회한 후, 온톨로지 내 수록된 개별 식물 어휘에 대하여

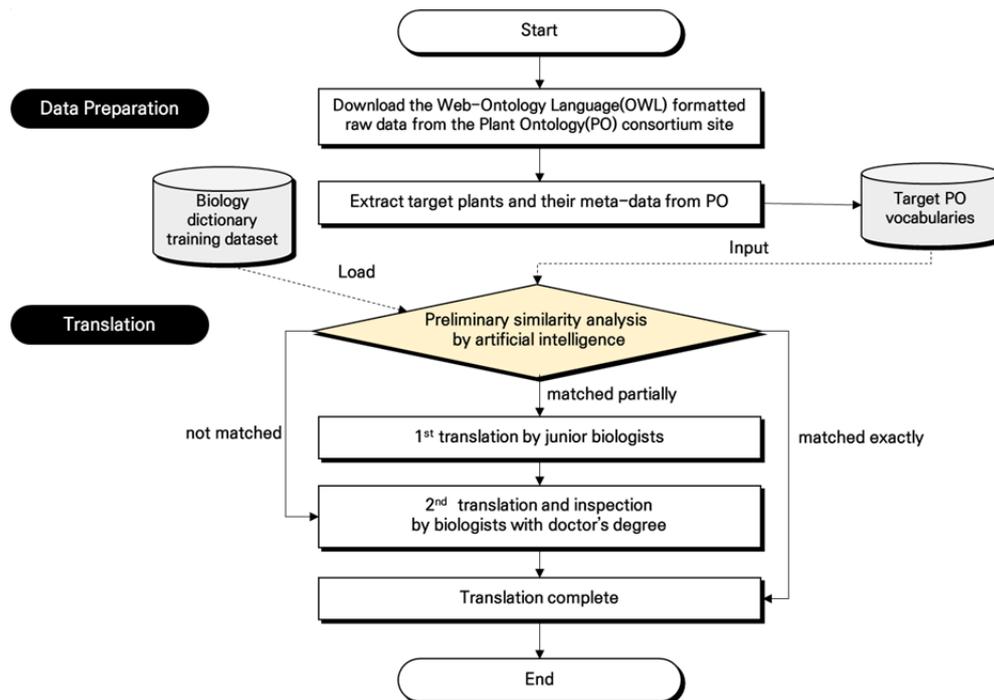


Fig. 1. Process flow for Korean contents construction of the plant ontology

지식 표현 모형 측면에서 프레임워크를 분석하고, 그 결과를 제시하였다.^{13,14)} 특히, 전체적인 PO가 수록된 상위 온톨로지인 OBO 프레임워크의 견지에서 PO에 수록된 클래스들의 거시적 구조로부터 세부적인 구성요소인 개별 클래스의 구조 및 클래스들간 적용된 의미 관계망 등 미시적 구조까지 단계적으로 분석한 후 한국형 식물 지식 온톨로지의 구축 대상을 결정하였다.

2. 한국형 식물 지식 온톨로지 콘텐츠 구축

PO에 대한 한글 지식 콘텐츠 구축 절차는 크게 온톨로지 콘텐츠의 준비와 번역과정으로 나누어 진행하였으며, 세부 절차는 아래 그림과 같다(Fig. 1).

2.1. 대상 PO 콘텐츠 준비

PO 컨소시엄 공식 웹사이트(web-site)로부터 월드 와이드 웹(World Wide Web Consortium, W3C)의 온톨로지 웹 언어(Ontology Web Language, OWL) 형식의 최신 온톨로지 버전인 PO 2.1을 다운로드하여, 전체 PO가 수록된 원시자료를 수집하

였다.^{15,16)} 그러나, OWL 파일형식은 컴퓨터 프로그래밍용 기계어로 코딩이 되어 있어, 식물학 연구자의 접근이 용이하지 않았다. 따라서, OWL표준의 문법 규칙으로 쓰여진 구문을 해석하여 일반 텍스트(text)로 변환하는 프로그램인 파서(parser) 도구가 요구되었다. 초기 한국형 식물 지식 온톨로지 콘텐츠 개발의 지원을 위하여 OWL형식의 원시자료로부터 구축 대상 및 범위에 해당하는 핵심 메타데이터만을 추출한 후, 가독성이 좋은 스프레드시트 형식으로 변환·저장할 수 있는 파서 도구를 개발하였다. 또한, 해당 도구를 구동하여 한글 콘텐츠 개발을 위해 식물학자들에게 제공할 대상 온톨로지 목록 1,788건을 준비하였다.

2.2. 한글 PO 지식 콘텐츠 개발 절차

한글 PO 지식 콘텐츠는 인공지능(Artificial Intelligence, AI)을 활용한 어휘 유사도 분석, 생물학 전공자 21인에 의한 1차 번역, 생물학 박사학위 소지자 3인에 의한 2차 번역 및 감수의 과정으로 진행되었다. AI 활용 어휘 유사도 분석을 위한 레퍼런스는 생물과학협회에서 발간된 생물학어휘집 제3판(생물과

Table 1. Example of established Korean plant ontology knowledge content

ID	Term	Term in dictionary	Similarity	Draft	Revision	Reference	Remark
obo:PO_0000001	Plant embryo proper	Embryo (Baea), Plant (Sikmul)	0.33	Sikmul Goyu Baea	Sikmul Goyu Baea	Papers and professional books	[Report] Expectations and problems of blood manipulated from human pluripotent stem cells
obo:PO_0000002	Anther wall	Anther wall (Yakbyeok, Kkotbapbyeok)	1	Yakbyeok, Kkotbapbyeok	Kkotbapbyeok	Biological terms (KA OBS)	
obo:PO_0000003	Whole plant	Plant (Sikmul)	0.5	Cheonchae Sikmul	Cheon Sikmulchae	Biological terms (KA OBS)	
obo:PO_0000004	In vitro plant structure	In vitro (Siheomgwannae), Plant (Sikmul), Structure (Gujo, Guseong)	0.5	Siheomgwannae Gujo	Siheomgwannae Sikmul Gujo	Biological terms (KA OBS)	
obo:PO_0000005	Cultured plant cell	Cultured cell (Baeyangsepo), Plant (Sikmul)	0.66	Sikmul Baeyangsepo	Baeyang Sikmulsepo	Biological terms (KA OBS)	
obo:PO_0000006	Plant protoplast	Plant (Sikmul), Protoplast (Wonhyeongilche)	0.5	Sikmul Wonhyeongilche	Sikmul Wonhyeongilche	Biological terms (KA OBS)	
obo:PO_0000007	Leaf-derived cultured plant cell	Leaf (Ip, yeop), cultured cell (Baeyangsepo), Plant (Sikmul)	0.5	Ip Yurae Baeyang Sikmulsepo	Ip Yurae Baeyang Sikmulsepo	Biological terms (KA OBS)	

톨로지 그룹을 크게 1) 시간관계성, 2) 생물개체수준을 축으로 세분화된 계층(Fig. 2)으로 구획화하고, 온톨로지내 어휘 클래스들을 배치하고 있다. PO내 수록된 본질적인 식물 어휘 클래스들은 1) 식물세포(plant cell), 2) 식물해부(plant anatomy), 3) 식물발생단계(plant development stage)의 3가지 온톨로지 서브 그룹으로 모델링 되어 있다. OBO BFO 프레임워크 관점에서 살펴보면 시간관계성 측면에서 지속체(continuant)-독립체(independent)로써, 식물세포 및 식물해부 그룹을 발생체(occurrent)로써 식물발생단계 그룹을 배치하였다. 또한, PO구성에 필요한 NCBITaxon, CARO, GO 어휘 클래스를 상호 참조 개념으로 연결하는 구조로 설계되어 있다.

1.2. PO의 미시적 프레임워크

PO는 미시적으로 각각 어휘, 관계망을 설명하는 클래스와 속성으로 구성되어 있다. 클래스는 OBO 온톨로지 허브 사이트의 하위의 IRI를 통하여 고유하게 웹상에서 식별된다. 속성은 개체, 데이터 및 주석 카테고리로 세분화되고, 계층적으로 구성되어 있다. 클래스는 대표 어휘 명칭, 식별번호 및 정의로 표현하고, 속성은 동등(has_exact_synonym), 상위/하위(is_a), 소유(has_part) 관계 등으로 표현하며, 클래스와 속성 간의 조합을 통해 관계망이 형성되도록 구조화되어 있다.

1.3. 한국형 식물 지식 온톨로지 구축 대상·범위 도출

이상과 같은 PO 거시적 및 미시적 프레임워크 분석을 토대로, 한국형 식물 지식 온톨로지 구축을 위한 구축 대상 및 범위를 다음과 같이 도출하였다. 첫째, 구축 대상은 PO에는 PO 외에 CARO, GO 등 다양한 참조된 온톨로지들이 포함되어 있었다. 그러나, 온톨로지 별 독립성 및 전문가 풀 구성의 용이성을 이유로 본 연구에서는 PO의 본질적 클래스들만으로 정하였다. 둘째, 구축 범위는 국제적 PO와의 연결성 확보를 위한 IRI 식별자, 대표 영문명칭(label), 정의(definition) 및 비고(comment) 속성을 입력 변수로 추출하고, PO의 한글 지식 콘텐츠 연계 변수로써 동등관계(has_exact_synonym) 속성을 선정하였다.

2. 한국형 식물 지식 온톨로지 콘텐츠 구축

2.1. 전체 식물 온톨로지 콘텐츠 구축 결과

II. 2.1.에 제시한 절차에 따라, PO에 등재된 식물 어휘 전체에 해당하는 총 1,957건의 한국형 식물 지식 온톨로지 콘텐츠를 구축하였으며, 어휘 카테고리별 구축현황은 아래와 같다 (Table 4). 가장 상위 수준의 클래스인 지속체와 발생체를 중심으로, PO에 포함된 지속체는 모두 독립적 지속체로서 생물학적 실제(biological entity)와 물질적 실제(material entity)로 구분되어 있으며, 가장 많은 어휘가 포함된 것은 총 1,567건의 식물 구조 관련 어휘였다. 발생체 클래스는 상대적으로 어휘의 수가

Table 4. Construction status of Korean knowledge-base for classes in the plant ontology

Class	Number
Continuant (BFO)	
Independent continuant (BFO)	
Biological entity (CARO)	
Anatomical entity (CARO)	10
Archegoniophore	0
Archegonium head	0
Archegonium megagametophyte	0
Coma	0
Gametophyte perianth	0
Inflorescence	26
Infructescence	1
Second order inflorescence	0
Second order infructescence	0
Shoot axis tegument layer	1
Material entity (BFO)	1
Plant anatomical entity	3
Plant anatomical space	35
Plant structure	1,567
Portion of plant substance	15
Organism or virus or viroid (CARO)	
Occurrent (BFO)	
Process (BFO)	1
Biological_process (GO)	
Molecular_function (GO)	
Plant structure development stage	5
Collective plant organ structure development stage	2
Collective phyllome structure development stage	30
Shoot system development stage	11
Multi-tissue plant structure development stage	3
Fruit development stage	8
Plant organ development stage	51
Seed development stage	10
Plant tissue development stage	1
Vascular tissue development stage	10
Trichome development stage	2
Leaf trichome development stage	4
Seed trichome development stage	4
Whole plant development stage	3
Gametophyte development stage	36
Life of whole plant stage	0
Sporophyte development stage	117
Total	1,957

적은 편이고, 전초 발생 단계(whole plant development stage) 중 포자체 발생단계(sporophyte development stage)에 해당하는 어휘가 117건으로 가장 많은 비중을 차지했다. 최종적으로 개발된 한글 식물 지식 온톨로지 콘텐츠를 W3C OWL 형식의 PO 원시데이터 규격에 맞게 변환하고, Web-Protégé 시스템내에 이식하였다(Fig. 3).

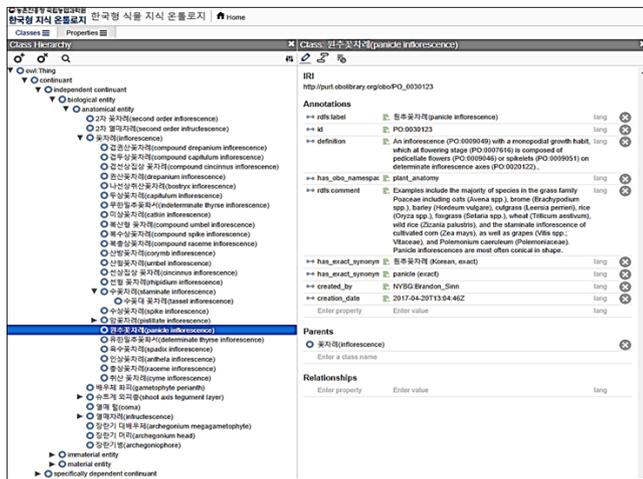


Fig. 3. The screen that transplanted Korean plant ontology into the Web-protégé system

2.2. 개별 온톨로지 지식 콘텐츠 구축 예시 및 활용

기후변화 생물지표(Climate-sensitive Biological Indicator Species, CBIS) 중 한 식물 종인 검노린재(*Symplocos tanakana* Nakai)를 일례로, 클래스 수준에서 온톨로지 콘텐츠 결과 및 활용방안을 제시하고자 한다. Fig. 4는 검노린재의 식물해부학적 특징과 관련한 어휘간 의미관계망 그래프로 나타난 것이다. 기후변화 지표종은 기후 온난화에 의해 북쪽으로 서식지를 넓혀가거나, 우리나라에서 사라질 것으로 예상되거나, 지구상에서 멸종이 우려되는 대표적인 생물종으로 식물은 44종이 포함된 바 있다.¹⁷⁾ 그 중 검노린재는 낙엽 떨기나무로 6월에 햇가지 끝에서 원추꽃차례(panicle inflorescence)로 흰색 꽃이 핀다. 개발된 한글 식물온톨로지 지식체계에서 원추꽃차례와 관련 어휘간 의미관계망 그래프로 생성하면 검노린재 꽃차례의 해부학적, 발생학적 특징을 구조화할 수 있다. 원추꽃차례는 슈트계(shoot sytem) 중 생식 슈트계(reproductive shoot system)에 해당하는 꽃(flower)의 배열 방식, 즉 꽃차례(inflorescence) 중 한 종류이다. 꽃차례는 꽃차례 발달기(inflorescence development stage) 동안 발생하게 되며, 이는 크게 생식 슈트계 발달기(reproductive shoot system development stage)에 포함되며, 더 상위 개념으로는 슈트계 발달기(shoot system development stage)에 포함된다고 할 수 있다. 이와 같은 방식으로 구축된 모든 한글 식물 지식 온톨로지 콘텐츠는 고유 IRI식별자와 함께 글로벌 PO와 연결되고, 식물 어휘에 적용되는 고유한 관계(예, is a, developed from 등)를 정의하여, 어휘간 의미 관계망을 표현하

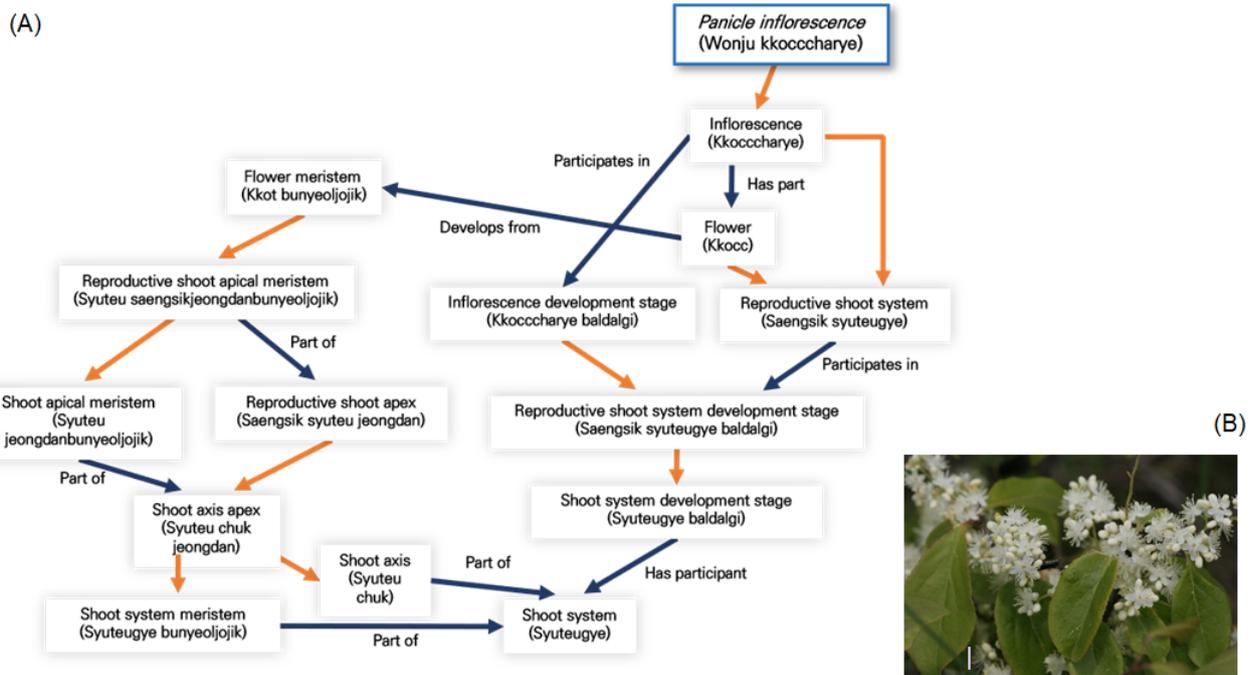


Fig. 4. (A) An ontology graph of selected term, panicle inflorescence which part of *Symplocos tanakana*. (B) Inflorescence of *Symplocos tanakana*.

고 있다.

IV. 고 찰

독성, 측정, 노출, 건강 영향 등 광범위한 학문 영역이 공존하는 환경보건학 분야에서, 지식베이스의 구축 주체 혹은 목적에 따라 서로 다른 표기 방식(예, 언어, 이명 등)과 수준으로 표기된 어휘 체계는 데이터 검색의 불완전성 및 이중 자료원간 호환성 확보를 위한 수작업 가공 등의 문제점을 야기하고 있다.^{18,19)} 이를 해결하기 위한 수단으로써 온톨로지는 환경·생태 분야의 다양한 생물종에 대한 독성시험 빅데이터 정보의 자동화된 비교 및 분류 처리에 활용되고 있다.²⁰⁾ 또한, 독성과 노출 정보를 건강영향 정보와 연계를 통한 인과관계 규명, 노출 정보 부족(information gap)을 파악하고자 노출 인자(stressor)-발생(event)-영향(outcome) 중심으로 구축된 노출 온톨로지(exposure ontology), 실내 공기질의 모니터링 정보 표준 마련 및 제어를 위한 온톨로지의 표준화 및 활용 연구들이 환경보건 분야에서 전방위적으로 수행되고 있다.²¹⁻²⁵⁾ 2022년 유럽식품안전청(The European Food Safety Authority)은 인공지능을 활용한 화학물질의 위해성평가를 위한 증거 관리 체계 구축을 위한 최우선 인프라로서 온톨로지의 구축·활용을 권고하고 있다.^{26,27)} 원 헬스(one health)의 생태계(ecosystem)적 견지에서, OBO에 집적된 환경, 농업, 식품, 보건 등 다양한 분야의 온톨로지들을 유기적으로 융합하여 지식을 통합·활용하고 있다.²⁸⁾

본 연구는 환경·생태 분야 중 식물에 대하여 국내 최초로 수행된 한글 온톨로지 콘텐츠 구축 연구이다. PO는 영어로 구축된 국제적인 식물 온톨로지이며, 언어권이 다른 국가들은 다국적 정보 통합을 위하여 자국어로 된 콘텐츠를 마련하고 PO에 관계망을 통하여 등재함으로써 국제적인 조화를 이루려는 노력을 하고 있다. 구축된 식물 어휘를 양적인 측면에서 살펴보면, 2023년 6월 20일 기준으로 PO에 수록된 전체 건수는 1,957건이고, 스페인어와 일본어는 각각 1,360건, 1,350건이 구축되어 있다. 반면, 한국어 식물 어휘는 본 연구를 통하여 1,957건 모두를 구축하였다. 이에 따른 효과는 특정 언어(예, 한국어)로 된 검색어를 활용하여 다른 언어(예, 일본어, 스페인어, 영어)로 작성된 전세계 디지털 정보를 연계 및 확장하여 완전하고 정밀하게 검색할 수 있다. 또한, 특정 식물 매체의 화학물질의 모니터링 자료를 생산 시 표준으로 활용할 수 있다. 가령, 앞서 결과에서 제시한 원추꽃차례 식물의 고유식별 코드는 'PO:0030123'으로 식물 매체 데이터베이스에 구축 시 부여한 후, 해당 식물의 IRI인 'http://purl.obolibrary.org/obo/PO_0030123'을 웹 브라우저(web-browser) 주소창에 입력하면, 해당 식물의 계통, 정의, 특성 및 연관된 어휘(예, 그 식물이 원료로 활용된 식품) 등 추가적인 정보를 획득할 수 있다. 즉, 환경·생태 분야 연구 데이터 표준코드로서 적용함으로써 국내

의 국가 단위 연구데이터의 신속한 통합은 물론 국제적인 연구데이터의 수집, 통합 및 분류에 활용될 수 있을 것으로 생각한다.

한글 식물 어휘의 공신력 확보를 위하여 생물학어휘집을 수집하여 적용하고, 식물학자로 구성된 작업반을 구성하여 검토 및 검증을 실시하였다. 그러나, 생물학어휘집에 수록되지 않은 어휘들은 국내 전문가 그룹 차원에서의 합의 등 대표성 측면에서 한계점을 가지고 있으며, 향후 공개 검증 사이트를 개설하여 공신력을 확보할 필요가 있다. 또한, 본 연구는 PO 내 어휘를 연구 범위로 한글 콘텐츠를 개발하였다. 그러나, 연관된 온톨로지인 BFO, GO 및 CARO 온톨로지에 대한 추가적인 콘텐츠 확대가 시급하며, 장기적으로 환경보건과 관련된 ENVO, UBERON, DOID 및 NCBITaxon 온톨로지로 전면적으로 확대해 나갈 필요가 있다.

구축된 식물 지식 온톨로지는 일종의 한국어 식물 어휘 표준으로 직접적인 학문 분야인 식물학은 물론 생태학, 농업, 유전체학, 표현체학, 식품과학 및 영양학 등 다양한 분야의 학술문헌 및 보고서 작성, 연구데이터 생산, 공공 및 민간 데이터베이스 구축에 일관성 있는 디지털 정보 생산에 광범위하게 활용될 수 있다.^{29,30)} 또한, 지식 표현 측면에서 단순한 어휘사전이 아닌 개체, 클래스, 관계망으로 구성된 최상위 수준의 온톨로지 모형으로 구현되어, 특정 식물에 대하여 해부학적, 형태학적 및 발생 단계의 정의, 특성, 계통, 부위 등 다양한 의미 관계망까지 정교하면서도 일관성 있게 기술할 수 있다.

디지털 지능정보 처리 측면에서, W3C OWL 표준 등 시멘틱 웹기술이 적용된 본 식물 지식 온톨로지는 유의어, 연관어 등 관계망을 활용하여 컴퓨터가 정보를 읽고 이해하며, 의미 기반의 논리적 추론 검색에 직접적으로 활용되고 있으며, 자연어 처리 분야의 인공지능 학습에도 적용되고 있다.³¹⁻³³⁾ 또한, 글로벌 PO를 기반으로 만들어진 본 한국형 식물 온톨로지는 GO, NCBITaxon 등 유전체 관련 온톨로지와의 상호 결합을 통하여 유전자, 돌연변이 등의 정보를 제공하는 데이터베이스와 연계하여 확장 가능하다.³⁴⁾

마지막으로 본 연구에 통해 확립된 한국형 식물 온톨로지는 국가 단위 생태 모니터링 데이터 생산 시 식물 매체나 생물지표에 대한 표준 코딩 체계로 적용함으로써, 국가 단위 식물 매체 중 화학물질의 오염도 데이터베이스 구축, 식물 종 혹은 부위 간의 상호작용, 생태학적 기능 및 생물 다양성 연구 등에 활용할 수 있다. 또한, 서로 다른 목적으로 생산된 이중 데이터소스들에 대하여 호환성 확보를 위한 수작업 없이 연계 및 통합함으로써, 신속하고 정확한 환경보건 정책 의사결정 지원에 기여할 것으로 생각된다.

V. 결 론

본 연구는 환경 독성시험 및 화학물질 모니터링의 대상이 되는 식물에 대한 글로벌 지식 통합을 위한 기반 연구로써, 식물의 통제된 어휘, 구체적으로 식물 해부학, 형태학 및 발생 단계 어휘들로 구성된 PO 프레임워크를 심층 분석하여 제시하였다. 또한, 식물 콘텐츠를 구획하고 공신력 있는 어휘사전을 기초로 한글 식물 콘텐츠를 개발하였으며, 어휘 사전에 존재하지 않는 어휘들의 번역 및 전체적인 콘텐츠 감수는 국내 생물학 전문가로 조직된 작업반을 구성하여 실시하였다. 최종적으로, 개발된 식물 지식 콘텐츠는 종래의 PO 지식베이스에 온톨로지 체계에 맞추어 물리적으로 이식하였다. 또한, 이 과정에서 요구되었던 OWL 파일의 전처리 및 이식 프로세스를 자동으로 수행할 수 있는 컴퓨팅 알고리즘을 개발하였다.

개발된 한국형 식물 지식 온톨로지는 환경보건 분야에서 다음과 같이 활용될 수 있다. 첫째, 환경·생태 독성시험 및 모니터링 결과 데이터베이스 구축 시, 대상 식물에 대한 표준화된 코드로 적용함으로써, 서로 다른 연구 주체에 의해 생산된 정보의 연계 및 확장에 별도의 수작업 없이 활용할 수 있다. 둘째, 한글 식물 어휘만으로도 국경을 넘어 다국어(예, 일본어 등)로 생산된 글로벌 환경·생태 분야 디지털 정보에 대한 의미 기반의 지능적 검색, 수집 및 통합을 효과적으로 수행할 수 있다. 셋째, 글로벌 OBO에 수록된 환경(예, ENVO), 유전체(예, GO, NCBITaxon), 보건(예, DOID) 등 다른 온톨로지와의 연계를 통하여, 환경보건학적 관점에서 요구되는 다양한 지식확장의 효율적인 도구로써 활용이 가능할 것으로 생각된다.

따라서, 향후 본 연구를 통해 체계적으로 구축된 한국형 식물 온톨로지를 국내 식물학 전문가 그룹의 공식적 합의 절차를 통하여 국가 단위 표준으로 발전시키고, 신종 식물 어휘 등재 및 고도화 등 지속적인 업데이트에 관련된 국가적 차원에서의 중장기적인 지원 정책을 마련할 필요가 있다.

감사의 글

본 성과물은 농촌진흥청 연구사업(과제번호:PJ01704701)의 지원에 의해 이루어졌으며, 이에 감사드립니다.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

1. Parkes M, Panelli R, Weinstein P. Converging paradigms for envi-

- ronmental health theory and practice. *Environ Health Perspect.* 2003; 111(5): 669-675.
2. Charleston AE, Wilson HR, Edwards PO, David F, Dewitt S. Environmental public health tracking: driving environmental health information. *J Public Health Manag Pract.* 2015; 21(Suppl 2): S4-S11.
 3. Yang W. Risk assessment in environmental health. *J Korean Soc Environ Eng.* 2007; 29(5): 489-495.
 4. Kwon N, Suh J, Lee H. Data cleaning and integration of multi-year dietary survey in the Korea National Health and Nutrition Examination Survey (KNHANES) using database normalization theory. *J Environ Health Sci.* 2017; 43(4): 298-306.
 5. Lee J, Yang S, Lee H. Knowledge modeling and database construction for human biomonitoring data. *J Food Hyg Saf.* 2020; 35(6): 607-617.
 6. Moreira A, Alvarenga L, de Paiva Oliveira A. "Thesaurus" and "ontology:" a study of the definitions found in the computer and information science literature, by means of an analytical-synthetic method. *Knowl Organ.* 2004; 31(4): 231-244.
 7. Li GKJ, Trappey CV, Trappey AJC, Li AAS. Ontology-based knowledge representation and semantic topic modeling for intelligent trademark legal precedent research. *World Pat Inf.* 2022; 68: 102098.
 8. Moon K, Park S. Oriental medical ontology for personalized diagnostic services. *J Korea Soc Comput Inf.* 2010; 15(1): 23-30.
 9. Hoehndorf R, Slater L, Schofield PN, Gkoutos GV. Aber-OWL: a framework for ontology-based data access in biology. *BMC Bioinformatics.* 2015; 16: 26.
 10. OBO Foundry. Open Biological and Biomedical Ontology Foundry. Available: <http://obofoundry.org/> [accessed 20 June 2023].
 11. Plant Ontology Consortium. The Plant Ontology Consortium and plant ontologies. *Comp Funct Genomics.* 2002; 3(2): 137-142.
 12. Walls RL, Cooper L, Elser J, Gandolfo MA, Mungall CJ, Smith B, et al. The plant ontology facilitates comparisons of plant development stages across species. *Front Plant Sci.* 2019; 10: 631.
 13. European Molecular Biology Laboratory. Ontology Lookup Service. Available: <https://www.ebi.ac.uk/ols/index> [accessed 20 June 2023].
 14. National Center for Biomedical Ontology. BioPortal. Available: <https://bioportal.bioontology.org> [accessed 20 June 2023].
 15. Plant Ontology. Plant Ontology Database. Available: <https://www.plantontology.org/> [accessed 20 June 2023].
 16. World Wide Web Consortium. Web ontology language. Available: <https://www.w3.org/OWL/> [accessed 20 June 2023].
 17. National Institute of Biological Resources. Biodiversity of the Korean Peninsula. Available: <https://species.nibr.go.kr/> [accessed 20 June 2023].
 18. Song H, Shin Y, Jang BY, Kang MS, Kim HR, Lee H. Development of Korean agro-food thesaurus knowledge representation model. *J East Asian Soc Diet Life.* 2023; 33(3): 241-250.
 19. Buttigieg PL, Pafilis E, Lewis SE, Schildhauer MP, Walls RL, Mungall CJ. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *J Biomed Semantics.* 2016; 7(1): 57.
 20. Boyles RR, Thessen AE, Waldrop A, Haendel MA. Ontology-based

- data integration for advancing toxicological knowledge. *Curr Opin Toxicol.* 2019; 16: 67-74.
21. Chan LE, Thessen AE, Duncan WD, Matentzoglou N, Schmitt C, Grondin CJ, et al. The environmental conditions, treatments, and exposures ontology (ECTO): connecting toxicology and exposure to human health and beyond. *J Biomed Semantics.* 2023; 14(1): 3.
 22. Mattingly CJ, McKone TE, Callahan MA, Blake JA, Hubal EA. Providing the missing link: the exposure science ontology ExO. *Environ Sci Technol.* 2012; 46(6): 3046-3053.
 23. Donkers A, Yang D, de Vries B, Baken N. Semantic web technologies for indoor environmental quality: a review and ontology design. *Buildings.* 2022; 12(10): 1522.
 24. Adeleke JA, Moodley D. An ontology for proactive indoor environmental quality monitoring and control. Paper presented at: the 2015 Annual Research Conference of the South African Institute of Computer Scientists and Information (SAICSIT '15); 2015 Sep 28-30; Stellenbosch, South Africa. New York: Association for Computing Machinery, 2015. p. 1-10.
 25. Fuertes A, Casals M, Gangoellis M, Forcada N, Roca X. An ontology for environmental and health and safety risks' evaluation for construction. In: Borgo S, Lesmo L. editors. Vol. 174, Formal ontologies meet industry. Amsterdam: IOS Press; 2008. p.100-108.
 26. PwC EU Services & Intellera Consulting, Bersani C, Codagnone J, David L, Foiniotis A, Galasso G, et al. Roadmap for actions on artificial intelligence for evidence management in risk assessment. *EFSA Support Publ.* 2022; 19(5): 7339E.
 27. Wittwehr C, Blomstedt P, Gosling JP, Peltola T, Raffael B, Richarz AN, et al. Artificial intelligence for chemical risk assessment. *Comput Toxicol.* 2020; 13: 100114.
 28. Dooley DM, Griffiths EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Lange MC, et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *NPJ Sci Food.* 2018; 2: 23.
 29. Ibrahim S, Fathalla S, Lehmann J, Jabeen H. Toward the multilingual semantic web: multilingual ontology matching and assessment. *IEEE Access.* 2023; 11: 8581-8599.
 30. Costa SD, Barcellos MP, de Almeida Falbo R. Ontologies in human-computer interaction: a systematic literature review. *Appl Ontol.* 2021; 16(4): 421-452.
 31. Mehta S, Tiwari S, Siarry P, Jabbar MA. Tools, languages, methodologies for representing semantics on the web of things. Hoboken: Wiley; 2022. p.211-240.
 32. Ahmed I, Yadav PK. Ontology-based classification method using statistical and symbolic approaches for plant diseases detection in agriculture. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4367154 [accessed 20 June 2023].
 33. Chen H, Luo X. An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing. *Adv Eng Inf.* 2019; 42: 100959.
 34. Smaili FZ, Gao X, Hoehndorf R. Formal axioms in biomedical ontologies improve analysis and interpretation of associated data. *Bioinformatics.* 2020; 36(7): 2229-2236.

〈저자정보〉

주은정(교수), 이현주(대표이사)