ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

ontocommons.eu
twitter.com/ontocommons
linkedin.com/company/ontocommons

Report D4.5
"Ontology ecosystem knowledge graph, first version"

Grant Agreement: 958371

OntoCommons - Ontology-driven data documentation for Industry Commons, has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 958371.

Project Title	Ontology-driven data documentation for Industry Commons
Project Acronym	OntoCommons
Project Number	958371
Type of project	CSA - Coordination and support action
Topics	DT-NMBP-39-2020 - Towards Standardised Documentation of Data through taxonomies and ontologies (CSA)
Starting date of Project	01 November 2020
Duration of the project	36 months
Website	www.ontocommons.eu

Report D4.5 "Ontology ecosystem knowledge graph, first version"

Work Package	WP4 | Ontology Commons Ecosystem Toolkit
Task	T4.6 – Knowledge base construction
Lead author	John Breslin (NUIG), Lan Yang (NUIG)
Contributors	Florina Piori (TU WIEN-ISE), Umutcan Serles (UIBK), María Poveda-Villalón (UPM)
Peer reviewers	María Poveda-Villalón (UPM)
Version	Final
Date	30/11/2022

Versioning History

Revision	Date	Editors	Comments
0.1	25/04/2022	Lan Yang	Creation of base version / outline
0.2	14/06/2022	Lan Yang	Creation of first version
0.3	22/06/2022	John Breslin, Lan Yang	Update version
0.4	28/06/2022	María Poveda	Include internal comments
0.5	29/06/2022	Lan Yang	Including some additional content
0.6	02/11/2022	Hedi Karray	Include internal comments
1.0	14/11/2022	Lan Yang	Update version
Final	29/11/2022	Nadja Adamovic	Final approval, submission

Glossary of terms

Item	Description
OWL	Web Ontology Language
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform Resource Identifier

Keywords
Ontology; OCEANS; Knowledge base; Knowledge graph

Disclaimer

Executive Summary

This document describes the knowledge base of the OntoCommons project, which is developed using a designated platform named **O**ntoCommons **C**ollaborative **E**diting and **A**uthoring k**N**owledge graph**S** (OCEANS), used for recording the knowledge generated in the project in a machine-readable fashion, and visualised through knowledge graphs. The OCEANS platform is developed on the basis of an open-source collaborative environment supporting RDF/OWL named WebProtégé. While retaining most of the features of WebProtégé, the OCEANS develops a new feature on top of it and customises the user interface to better suit the needs of OntoCommons. The aim of creating the knowledge base is to ensure that all the information generated within Work Package 4 (WP4) and across Work Packages (WPs) is encoded into machine-readable knowledge graphs that describe the ontology commons ecosystem toolkit and beyond.

In the first version of the ontology ecosystem knowledge graph, we primarily report the following outcomes: (1) the definition of the ontology ecosystem knowledge base and knowledge graph, (2) the content model of the ontology ecosystem knowledge base, (3) the requirement analysis for the collaborative editing and authoring platform, (4) the architecture of OCEANS, (5) the first version of the ontology ecosystem knowledge base and knowledge graphs, and (6) the future directions for the ontology ecosystem knowledge graph, second version (D4.8).

Table of Contents

List of Figures

1. Introduction

The OntoCommons project aims to develop an ontology ecosystem consisting of ontologies, methodologies, tools, references, demonstration cases and additional sources related to ontologies and their application, which can be effectively used as foundation for data documentation, annotation, and standardisation in the industrial domain. Over the course of the project, there are enormous information, data and knowledge collected and generated. It is the duty of T4.6, to ensure all of them is encoded into a knowledge base that organises machine-readable data describing the ontology ecosystem at the meta-level. A knowledge graph is a knowledge base that uses a graph-structured data model or topology to represent data. Through T4.6, two versions of knowledge graph of the ontology ecosystem will be developed. This report describes the first version of the ontology ecosystem knowledge graph (D4.5) that is generated in the first 12 months of the task duration, focusing on encoding the specification of the ontology development phases and the descriptions of tools that implement them.

As a result of the focus, this deliverable cross-refers the ontology development phases defined in D4.2 – Methodological Framework for Ontology Management [Fernández-Izquierdo et al., 2021] and the survey results from D4.3 – Landscape Analysis of Ontology Engineering Tools [Skjæveland et al., 2021]. D4.2 introduces the Linked Open Terms (LOT) methodology that is the methodological framework for OntoCommons which describes the activities that should be performed in the ontology development phases. D4.3 provides a landscape analysis of tools for ontology engineering, detailing the information about each tool and categorising them by their features. Considering both, D4.5 formalise the meta-level knowledge about the methodological framework along with the relevant tools in a machine-readable, fully interoperable, and standards-compliant knowledge graph.

In order to develop the knowledge graph, a platform named OCEANS, which is short for OntoCommons Collaborative Editing and Authoring kNowledge graphS, is established. OCEANS is a customised application of a generic web-based RDF/OWL editor, WebProtégé [Tania et al., 2013]. The use of OCEANS ensures the knowledge graph uses RDF/OWL as a representation language and enables the collaborative authoring of the ontology ecosystem content. One of the main features of OCEANS is that it reuses the functionalities of WebProtégé and provides extensive support for collaboration in the development of knowledge graph content, including change tracking, discussion thread, access control, and visualisation.

This report is a supplementary document to the knowledge base file. It focuses on the aspects that the knowledge base file does not specify. The remaining of the report proceeds as follows. Section 2 provides a task overview of the task objectives, methodology, and deliverables. Section 3 defines the meaning of ontology ecosystem knowledge base and knowledge graph. Section 4 discusses the content model for the knowledge base and the OCEANS platform for developing the knowledge graph. The links to the OWL file of the ontology ecosystem knowledge graph first version are provided in Section 5, while Section 6 overlooks several directions for its second version. Section 7 concludes the report.

2. Task Overview

T4.6 is linked to two deliverables: D4.5 – Ontology Ecosystem Knowledge Graph, First Version (in M20) and D4.8 – Ontology Ecosystem Knowledge Graph, Second Version (in M32).

D4.5 presents a knowledge graph that formally encodes and represents the specification of ontology development phases and guidelines and maintains formal links with descriptions of tools that implement them. D4.8 deals with any updates to the data schema or any amendments to the reference implementation, as well as provides a report accompanying this deliverable.

T4.6 is towards a knowledge base construction for the ontology ecosystem. It starts from M12 and continue until the end of the project. T4.6 is primarily divided into two phases: Phase 1 is from M12 to M20, while Phase 2 is from M21 to M32. The following sections describe the task objectives, methodology and deliverables for the entire task duration, while drawing emphases on the implementation in the first phase in detail.

2.1 Task Objectives

The overall task aim is to construct a knowledge base for capturing the knowledge about the ontology ecosystem collected and generated over the course of OntoCommons project. To be specific, the task aims to achieve the following targets:

- Describe the LOT methodology of the ontology development phases in the form of knowledge graph.
- Include details of existing ontology engineering tools collected from landscape analysis into the knowledge base, making them indexable and readable both by humans and machines.
- Encode the top reference ontology and industrial domain level ontologies of the ontology ecosystem based on their specifications and classifications.
- Link the demonstration cases with the reference implementation in the knowledge graph.
- Establish a mechanism to ensure the content of the knowledge graph can be continuously updated with the progress of the project.
- Introduce a tool in the knowledge base construction to support configuration management between various incremental revisions.
- Create a tunnel for all project participants to collaboratively contribute to the editing and authoring the knowledge graph.
- Display the knowledge base in the form of a visualised graph.
- Publish the knowledge graph so that the public can view and understand the meta-level description of the ontology ecosystem.

The first phase (M12 – M20) of T4.6 primarily places the emphasis on creating the content of the methodological framework and the reference implementation in the knowledge graph. The updates from the set of ontologies in the ontology ecosystem, the amendments from the demonstration's feedback loop, and any other revisions will be added in the second phase (M21 – M32). The first phase also focuses on providing the platform that fulfil the aforementioned requirements for a collaborative and exchangeable environment for drafting the knowledge graph.

2.2 Task Methodology

As part of T4.6, a collaborative authoring process was adopted. Different workflows were supported for the different phases of development. Upon the foundation of the OntoCommons organisational structure, the workflow for collaboratively constructing the ontology ecosystem knowledge graph is built. A managing editor leads an advisory group that is assigned to at least one branch of the ontology ecosystem, e.g., top reference ontology, domain level ontology, ontology ecosystem toolkit, etc. By identifying tags in the knowledge graph, one can locate these branches. A few examples of tags are cooperation, TRO, DLO, toolkit, demo, etc. A group of advisory members has been tasked with entering the knowledge base content in their respective expertise domains, such as LOT methodology, ontology engineering tools, domain-level ontologies, etc. Once the knowledge graph for a branch has all the relevant content entered, the managing editor will send it to at least three external reviewers for review. Managing editors will modify the knowledge base once reviewers submit their comments. Once no more changes are received in comments, the branch will be considered 'approved'.

3. Introduction to Ontology Ecosystem Knowledge Graph

In an information system, the knowledge base is a large collection of structured or unstructured data which can be used to share information between stakeholders. Knowledge bases vary in type such as tabular data, relational data, hierarchical data, etc. For T4.6, a knowledge graph is proposed that collects, organises, stores and shares knowledge that is collected or generated during the course of the project about the descriptions of each component in the ecosystem.

Important entities consist of a taxonomy supplying a hierarchical structure to key concepts that are relevant to the ontology commons ecosystem. The taxonomy along with triples pay more attention to describing the semantics of the ecosystem.

For a knowledge base, there are different representation models. Object models such as the knowledge graph are used to represent the knowledge base. Therefore, it is a specification of the knowledge base. The idea of a knowledge graph is that it is a directed labelled graph in which the labels have well-defined meanings. Nodes, edges, and labels make up a directed labelled graph. Each node can be an entity or class. Each edge represents a relationship between a pair of nodes. Each label describes the relationship.

RDF schema constructs (e.g., class, property, type, subClassOf, subPropertyOf, domain, range) are used to develop the knowledge graph. An overarching goal in the use of RDF is to be able to automatically merge useful information from multiple sources to form a larger collection that is still coherent and useful. As a starting point for this merging, all the information is conveyed in the same simple style, subject-predicate-object triples. Each node in the knowledge graph represents an entity of the ontology ecosystem. Each edge in the knowledge graph represents a class hierarchy or entity relationship. In hierarchical relationships (subClassof), the labels are left blank by default. In other cases, the labels are defined by properties.

4. Content and Tool Design for Ontology Ecosystem Knowledge Graph

To achieve the goals effectively, the development process of the knowledge graph needs support from sophisticated models and tools. First, the authoring process will take place on an internet-based platform that will be available throughout the project life cycle, and that supports collaboration among the experts from different tasks, work packages, and domains. For the initial phase of the development, a web-based tool for collaborative authoring of D4.5 ontology ecosystem knowledge graph, first version is needed, which is presented in this report. Second, the underlying model of the knowledge base will be formalised in a machine-readable form. The current version of the knowledge graph uses RDF as its representation language and allows the verification of the classification consistency, as well as easier merging information. Third, the content of the ontology ecosystem knowledge base will also be curated and peer-viewed on the online authoring platform and eventually the knowledge graph will be published to public and viewed by interested parties.

4.1 Ontology Ecosystem Knowledge Base Content Model

The knowledge base of the ontology ecosystem follows the RDF schema, so taxonomy is a major component of the content model, where ontology ecosystem entities are organised in a hierarchy based on a parent-child relationship: more generic entities appear at the top of the hierarchy and have more specific entities as children. An IRI is associated with each entity, along with other information.

Ontology ecosystem knowledge base uses RDF schema constructs to capture content and semantics of entities, as well as linkages between aspects of the ecosystem.

4.2 Design of OCEANS for Collaborative Editing and Authoring

To assist the realisation of the knowledge graph, a platform is developed based on the requirements. The OntoCommons Collaborative Editing and Authoring kNowledge graphS (OCEANS) is a customisation of WebProtégé that is designed for the knowledge graph authoring process. WebProtégé is a web-based and lightweight RDF browser and editor that supports collaboration. It offers a very flexible and extensible web framework that can be tailored into different user interface layouts to suit OntoCommons' needs. A key feature of WebProtégé is the ability to collaborate. The knowledge base can be browsed and edited simultaneously by users. They may discuss the content of an entity (e.g., a class, a property, or an individual). The system tracks all changes made to the knowledge base. User interface settings of WebProtégé are configured declaratively via an XML file. The XML configuration file can be easily modified to create a new user interface quickly. The new interface is often just a combination of pre-existing components that are put together like puzzle pieces.

The knowledge graph collaborative authoring tool – OCEANS – is a customisation of WebProtégé. Taking into consideration that most of the knowledge graph editors are ontologists, who have background in formal knowledge representation, the main objective in designing the platform is to reuse the functionalities provided by WebProtégé while add a new knowledge graph visualisation.

The login interface of OCEANS is shown in Figure 1.

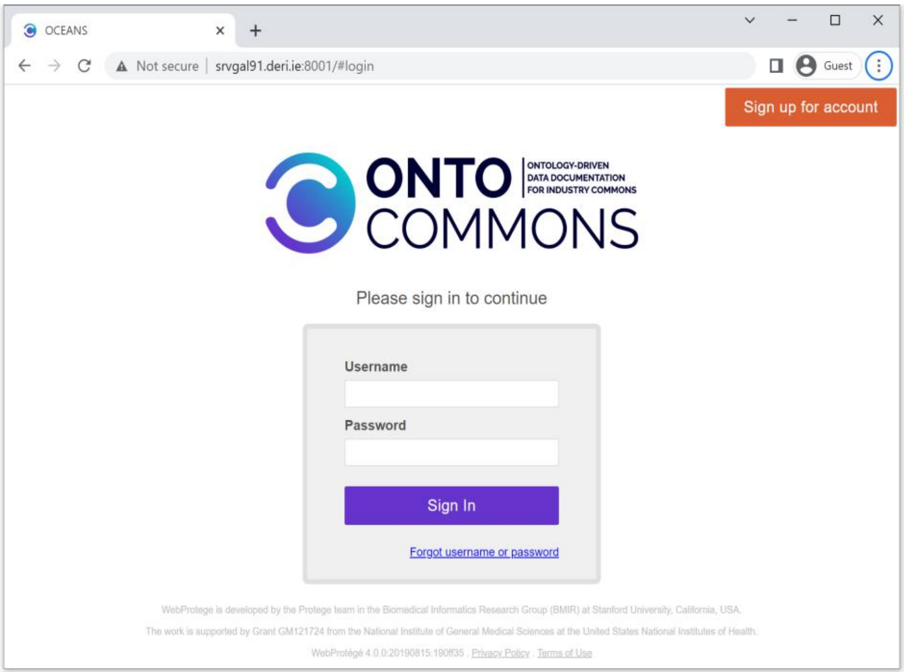

Figure 1 - Login page of OCEANS

The main user interface is shown in Figure 2 - Main user interface of OCEANS. The user interface is organised in a series of tabs, such as Classes, Properties, Individuals, Comments, History, and so on. Each tab presents a certain piece of functionality to the user.

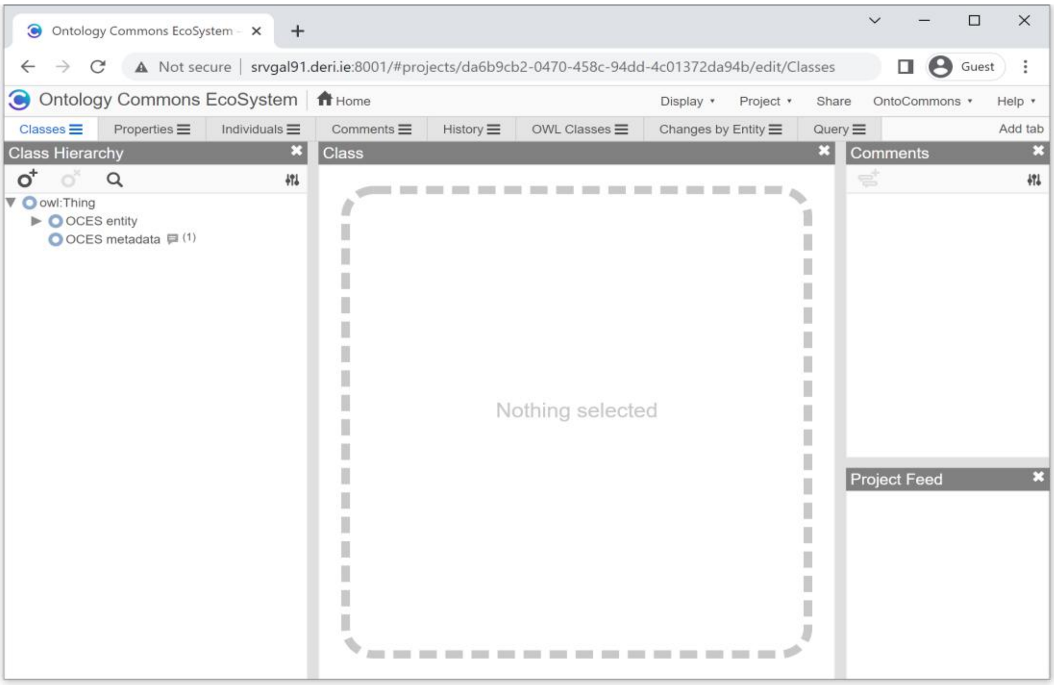

Figure 2 - Main user interface of OCEANS

The left hand side of the tab shows the hierarchy of ontology ecosystem knowledge base. There also displays the number of notes attached to an entity, indicated with a comment icon next to a name in the tree. In the same window, the user may also write his or her own notes or reply to existing posts. The middle panel is used for three things. It can show the details of an entity which allows users to edit. It can present the knowledge graph related to the selected entity displayed in a static visualisation provided default by WebProtégé, or present an alternative dynamic visualisation added by OCEANS. It can also highlight the changes of the entity. The right hand side is for tracking comments and reading project feed.

Figure 3shows an example of the default visualised knowledge graph of the tool 'Apache Jena'.

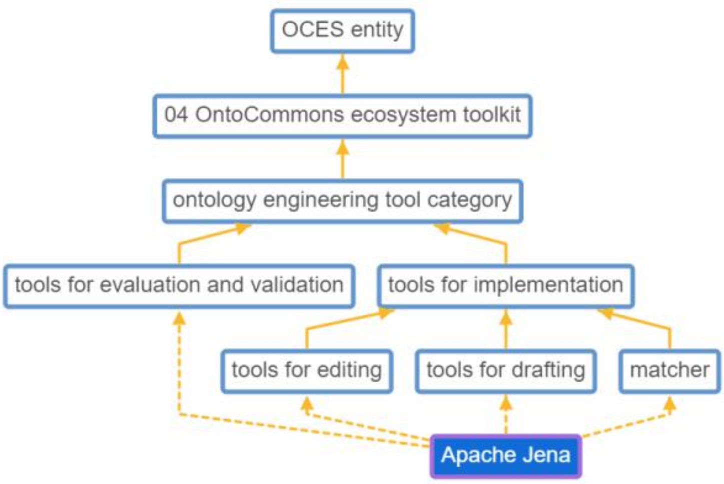

Figure 3 - Default visualisation of knowledge graphs

Figure 4 presents the new visualisation of the knowledge graph for the same tool.

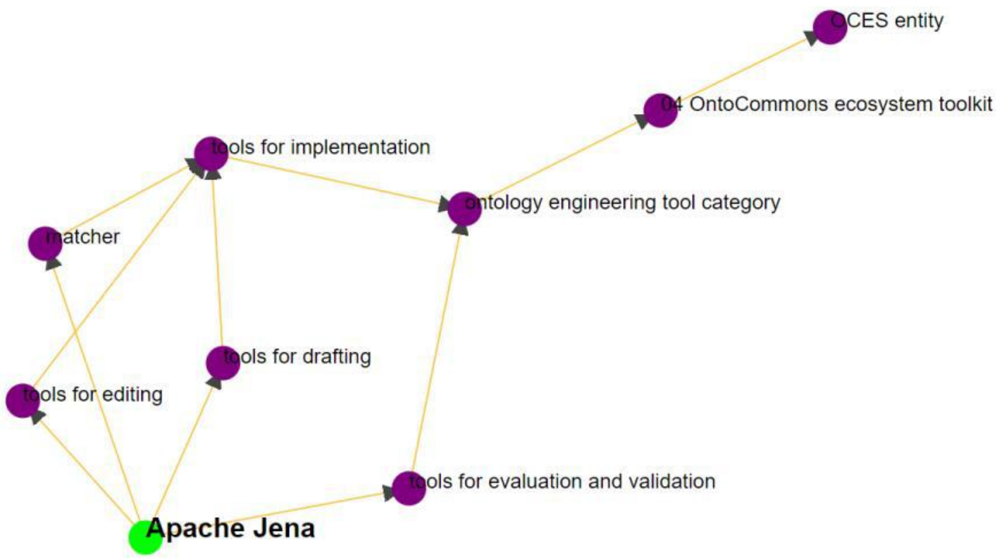

Figure 4 - New visualisation of knowledge graphs

5. Ontology Ecosystem Knowledge Base and Knowledge Graphs, First Version

The OWL file of the ontology ecosystem knowledge base can be view or downloaded from three sources.

OCEANS: http://srvgal91.deri.ie:8080/

GitHub: https://github.com/ontology-ecosystem/knowledge-graph

Industry Portal: not applicable for the first version, the second version will be published on Industry Portal.

The first version contains 6 classes of industrial domain ontologies, namely

- domain ontologies of computer science, systems and electrical engineering (9 instances),
- domain ontologies of material science and engineering (29 instances),
- domain ontologies of mechanical and industrial engineering (53 instances),
- domain ontologies of physics and chemistry (11 instances),
- domain ontologies of thermal and process engineering (6 instances), and
- domain ontologies, not elsewhere classified (22 instances).

There are 40 classes, 4 object properties, 3 annotation properties, 40 subclasses of relations, and 124 RDF 3-tuples relations used for describing the methodological framework of OntoCommons ecosystem.

The ontology engineering tool category contains 25 classes, with 7 main categories, namely

- tools for evaluation and validation,
- tools for implementation,
- tools for maintenance,
- tools for publication,
- tools for requirement specification,
- tools for use, and
- tools for specific purposes, not elsewhere classified.

A handful of instances has been added to the ontology engineering tool category in the first version for a demonstration purpose. For example, in Figures 3 and 4, it shows Apache Jena is a direct instance of tool for evaluation and validation, tools for editing, tools for drafting, and matcher. Annotation properties such as dc:description, hasDocumentation, hasPublication, hasRepository and so on are used to capture more detailed information about each tool.

Figure 5shows an excerpt of the top-level hierarchy. Various entities can be populated into 'OCES entity' collaboratively by each WP and across WPs. 'OCES metadata' is created to highlight the important meta-level knowledge, such as the methodological framework along with the relevant tools. This class will be continuous updated and enriched during the course of the project.

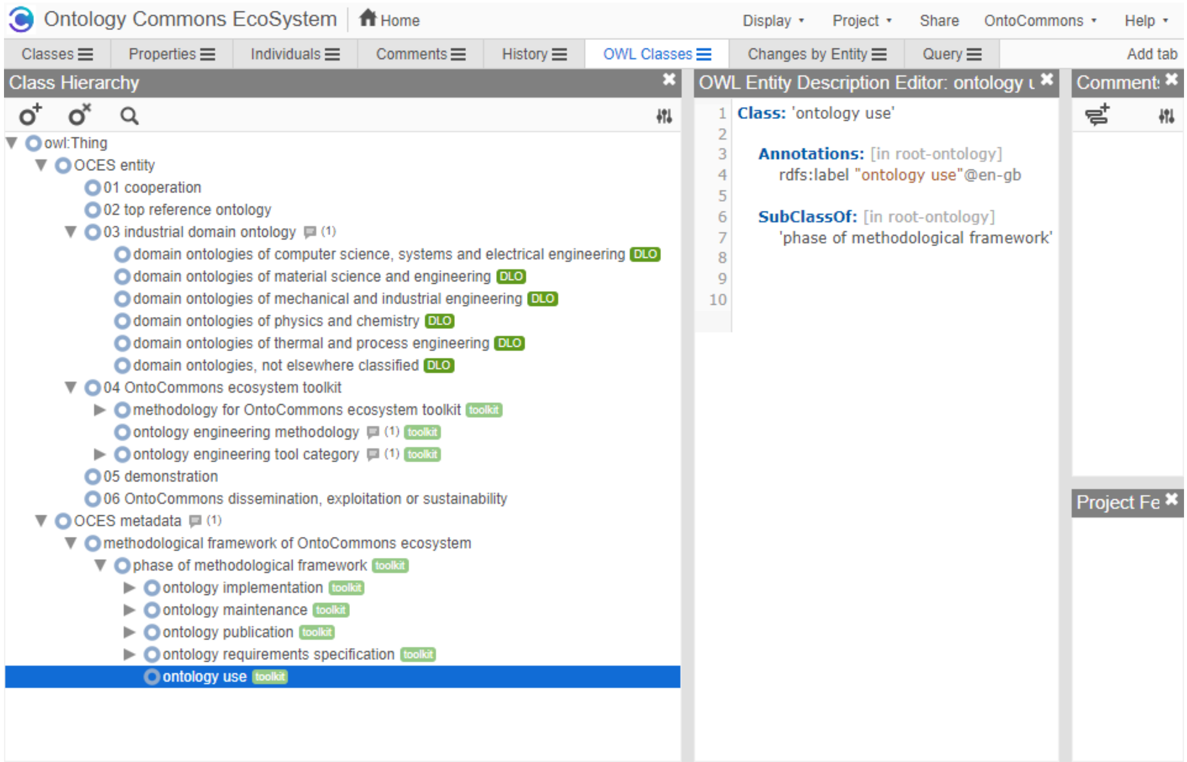

Figure 5 - Excerpt of the top-level hierarchy

Queries can be built arbitrarily with commonly used search criteria and the basic expressivity of the SPARQL query language. Figure 6 exemplifies a query for tools that are both for drafting and editing.

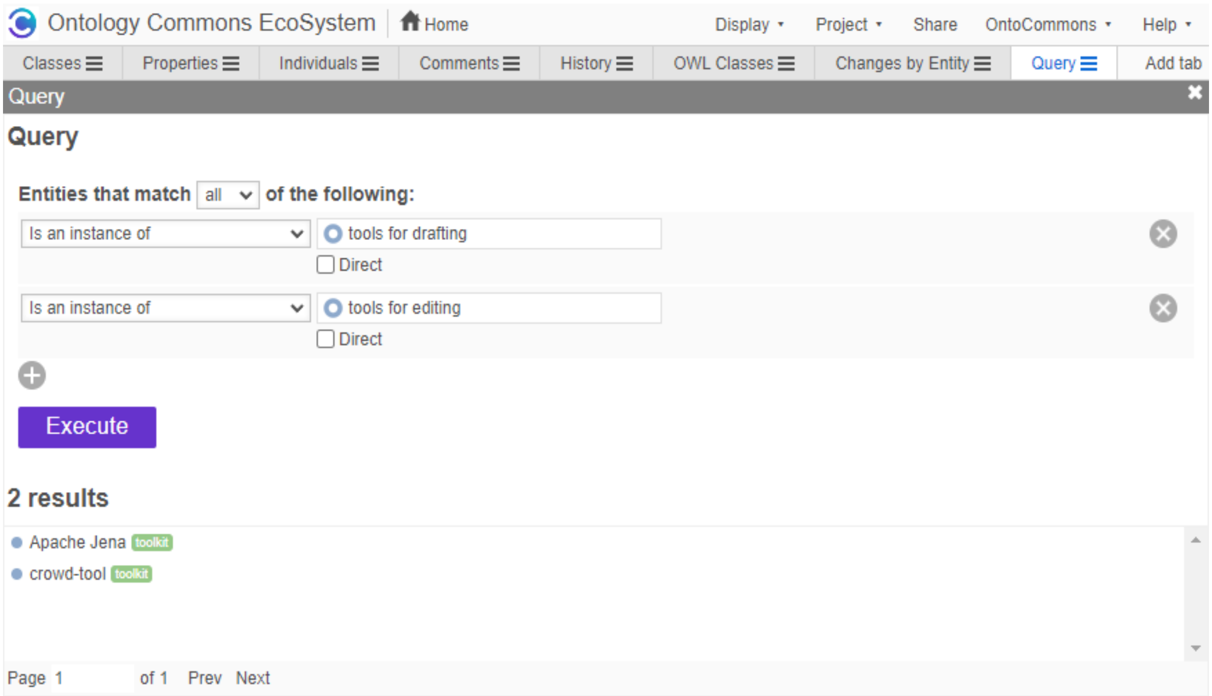

Figure 6 - Querying in OCEANS

6. Future Directions for Ontology Ecosystem Knowledge Graphs, Second Version

The following directions are for the future development of the second version of the knowledge graph:

- Describe the LOT methodology, OntoCommons version in the form of knowledge graph.
- Further add details of existing ontology engineering tools collected from landscape analysis into the knowledge base.
- Encode the top reference ontology and industrial domain level ontologies of the ontology ecosystem based on their specifications and classifications.
- Enrich the specifications of the industrial domain level ontologies by integrating the knowledge graphs with the OntoCommons ontology catalogue [María and Arkopaul, 2022], e.g., URI link, licence status, development language, syntax, subdomains, and natural language.
- Link the demonstration cases with the reference implementation in the knowledge graph.
- Publish the knowledge graph so that the public can view and understand the meta-level description of the ontology ecosystem.

7. Conclusion

T4.6 ensures that all the information generated within Work Package 4 (WP4) and across Work Packages (WPs) is encoded into a machine-readable knowledge base that describes the deliverables especially the ontology commons ecosystem at the meta-data level. This knowledge base can be viewed by knowledge graphs and indexed by formal knowledge engineering languages.

8. References

[Fernández-Izquierdo et al., 2021] Alba Fernández-Izquierdo, María Poveda-Villalón, Emna Amdouni, Hedi Karray. "D4.2 Methodological framework for ontology management". OntoCommons project deliverable. 2021.

[Skjæveland et al., 2021] Martin G. Skjæveland, Laura Slaughter, Christian Kindermann, Jinzhi Lu, Emna Amdouni, Francesca L. Bleken, Jesper Friis, Kevin Haller, Casper W. Andersen, Daniel Garijo, María Poveda-Villalón, Serge Chavez, Lan Yang, Laura Waslmayr, Stefani Tsaneva. "D4.3 - Landscape Analysis of Ontology Engineering Tools". OntoCommons project deliverable. 2021.

[Tania et al., 2013] Tania Tudorache, Csongor Nyulas, Natalya F. Noy, Mark A. Musen. "WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web". Semantic web 4, no. 1 (2013): 89-99.

[María and Arkopaul, 2022] María Poveda-Villalón, Arkopaul Sarkar. "D3.3 - Populated Domain Ontology Registry". OntoCommons project deliverable. 2022.