

MÁSTER EN INTELIGENCIA ARTIFICIAL

INGENIERÍA ONTOLÓGICA

Diseño de una ontología para instalaciones deportivas

Autores

LUIS COUTO SELLER
IRENE MARBÁN ÁLVAREZ
AÍDA MUÑOZ MONJAS

November 6, 2022

Contenidos

1	Introducción	2
2	Metodología NeOn	2
3	Especificación de la ontología	2
3.1	Propósito	2
3.2	Ámbito	2
3.3	Lenguaje de la Implementación	2
3.4	Potenciales usuarios	2
3.5	Potenciales casos de uso	3
3.6	Requisitos de la ontología	3
3.6.1	Requisitos no funcionales	3
3.6.2	Requisitos funcionales: preguntas de competencia	3
3.7	Glosario de términos	4
4	Planificación temporal de la ontología	5
5	Búsqueda de ontologías de alto nivel	6
6	Recursos no ontológicos	7
7	Patrones en la ontología	8
8	Modelo conceptual	8
9	Clases Multilingües	8
10	Implementación de la ontología con OWL	8
11	Evaluación de la ontología con OOPS!	12
11.1	Mejoras implementadas tras la sugerencia de OOPS!	12
11.2	Resultados finales de la evaluación	13
12	Documentación de la ontología	13
13	Conclusiones	13
	Bibliografía	14

1 Introducción

El objetivo de este trabajo es diseñar e implementar una ontología que represente de manera correcta instalaciones deportivas y sus características y acciones relacionadas. Las ontologías y otras fuentes de conocimiento utilizadas durante el desarrollo de este trabajo serán citadas, y se puede acceder a ellas a través de los hipervínculos de la bibliografía.

2 Metodología NeOn

3 Especificación de la ontología

3.1 Propósito

El objetivo de esta Ontología se enmarca en el proyecto Ciudades abiertas, que tiene como objetivo representar datos de ciudades en diversas áreas. El objetivo de esta ontología será modelar el conocimiento de Instalaciones Deportivas para que pueda ser utilizado como ontología dentro del proyecto.

3.2 Ámbito

La definición de Instalaciones Deportivas puede llegar a ser ambigua, dado que existen múltiples instalaciones donde se realizan deportes con un fin muy distinto: un estadio de fútbol y un parque de barras podrían ser ambos considerados instalaciones deportivas.

Nuestra ontología se enmarca en la definición propuesta por el informe estadístico de 2022 publicado por el Ministerio de Educación, Cultura y Deporte, donde se proporcionan indicadores estadísticos para estimar las dimensiones y características de las infraestructuras en España. En el documento, se define instalación deportiva como: 'Instalaciones destinadas al deporte que incluyen uno o varios espacios deportivos donde puede desarrollarse la actividad físico-deportiva.'.

3.3 Lenguaje de la Implementación

La ontología será implementada en OWL utilizando Protegé.

3.4 Potenciales usuarios

Los usuarios identificados como posibles interesados en el uso de la ontología han sido los siguientes.

1. Ayuntamiento interesado en el proyecto ciudades abiertas que quiera aumentar la transparencia de los datos de sus instalaciones deportivas.
2. Persona responsable de una instalación deportiva que quiera almacenar estadísticas de la misma y/o crear una página web de la instalación.
3. Deportista que quiera conocer qué instalación deportiva le es más conveniente.

4. Ministerio u órgano del gobierno que quiera analizar estadísticas sobre el uso de las instalaciones deportivas en su país.

3.5 Potenciales casos de uso

Los posibles casos de uso identificados son los siguientes.

1. Conocer los subespacios forman una Instalación Deportiva concreta.
2. Buscar servicios deportivos como clases o alquileres de espacios ofrecidos por la organización que gestiona la Instalación Deportiva.
3. Enumerar los deportes que pueden realizarse en cada Instalación Deportiva, y en que subespacio se realizan.
4. Obtener estadísticas del uso de servicios ofrecidos en la Instalación Deportiva por cliente.
5. Obtener información sobre los trabajadores de la organización que lleva la instalación deportiva.

3.6 Requisitos de la ontología

3.6.1 Requisitos no funcionales

Los requisitos no funcionales de la ontología se refieren a las características y aspectos generales no directamente relacionados con el contenido de la ontología. En nuestro caso son los siguientes:

- NFR1: La ontología deberá soportar otros idiomas a parte de Español.
- NFR2: La documentación asociada a la ontología debe incluir fuentes fiables.
- NFR3: El formato utilizado para representar el conocimiento será OWL.
- NFR4: El sistema de nombrado en la ontología sera comprensible y coherente.

3.6.2 Requisitos funcionales: preguntas de competencia

Para las preguntas de competencia, hemos optado por un acercamiento *Middle out*. Comenzamos realizando preguntas *Bottom-Up*, sin mirar bases de datos ni contenido relacionado, simplemente buscando preguntas sencillas que podrían estar relacionadas con una instalación deportiva. Cuando llegamos a un limite donde nuestras preguntas carecían de complejidad, entonces tuvimos un acercamiento *Top-Down*, donde ya comenzamos a mirar información acerca de Instalaciones Deportivas, y pudimos desarrollar preguntas más complejas. El resultado puede verse en la tabla:

CQ1	¿Cómo se llama la instalación deportiva?
CQ2	¿Cuál es la geolocalización de la instalación deportiva?
CQ3	¿Cuál es la dirección de la instalación deportiva?
CQ4	¿Cuál es el horario de apertura de la instalación deportiva?
CQ5	¿Cómo se llama la organización que gestiona la instalación deportiva?
CQ6	¿Quién es el director de la organización que gestiona la instalación deportiva?
CQ7	¿Cuántos entrenadores se dedican al baloncesto?
CQ8	¿Qué tipo de trabajadores tiene la organización que gestiona la instalación deportiva?
CQ9	¿Cuál es el número de teléfono para contactar con la organización de la instalación deportiva?
CQ10	¿Qué roles puede tener un entrenador?
CQ11	¿La instalación deportiva tiene baños o vestuarios?
CQ12	¿Qué espacios deportivos tiene al aire libre?
CQ13	¿Qué espacios deportivos cubiertos tiene?
CQ14	¿La instalación deportiva tiene piscina?
CQ15	¿Cuántas pistas de fútbol tiene la instalación deportiva?
CQ16	¿Qué deportes se pueden realizar en la instalación deportiva?
CQ17	¿La instalación deportiva tiene cafetería?
CQ18	¿Qué servicios de día (individuales) se pueden comprar en la instalación deportiva?
CQ19	¿Qué monitores imparten las clases de zumba que ocurren en la sala multiusos?
CQ20	¿Cuánto cuesta alquilar una pista de tenis durante 1 hora?
CQ21	¿Qué días hay clases de tenis?
CQ22	¿A qué hora son las clases de natación los lunes?
CQ23	¿Qué bonos de servicios son ofrecidos por la organización de la instalación deportiva?
CQ24	¿Cuánto cuesta un bono mensual de sala de musculación?
CQ25	¿Cuánto dura la clase de baloncesto impartida por un monitor concreto?
CQ26	¿Qué reglas de vestimenta tiene la piscina de la instalación deportiva?
CQ27	¿A quién aplica la regla de vestimenta de la piscina?
CQ28	¿Cuántos clientes han contratado el bono de clases de zumba el mes de enero?

Figure 1: Preguntas de competencia

3.7 Glosario de términos

Tras realizar las preguntas de competencia, hemos contado los sustantivos relevantes que se repetían, tanto en las preguntas para identificar clases y relaciones, como en las respuestas.

Cabe destacar que las preguntas se han contestado de manera muy genérica, sin concretar para una Instalación específica, de forma que los términos recogidos están un nivel por encima de ser instancias. Los resultados de algunos de los términos más relevantes se recogen en la siguiente figura:

Pre glossary of terms			
Terms from questions + Frequency			
Instalación deportiva	16	Dirección/geolocalización	2
Deportes	8	Trabajadores	4
Organización	5	Regla	1
Clases	4	Pistas	1
Bono	4	Baño/Vestuario	2
Horario/hora/día	6	Piscina	1
Servicios	2	Aire libre / Cubierto	1
Cientes	1	Alquiler	1
Cuesta (precio)	1	Cafetería	1
Terms from answers + Frequency			
Nombre Instalación	1	Cantidad de dinero	1
Coordenadas	1	Entrenador/monitor/técnico ...	1
Teléfono	1	Zumba/Baloncesto/Tenis	1
Localidad/Código Postal...		Franja horaria	1
Pista de Baloncesto, Espacio Multiusos, Sala musculación	1	Bono de clases	1

Figure 2: Glosario de términos

4 Planificación temporal de la ontología

El ciclo de vida utilizado en el diseño y desarrollo de esta ontología es el ciclo de vida incremental, pudiéndose considerarse como la primera iteración de un modelo ágil.

Se decidió utilizar un ciclo de vida incremental debido a las características del proyecto, ya que la ontología no fue diseñada a partir de un set de datos proporcionados por el cliente, si no que se realizó el modelado a partir de las posibles necesidades de sus usuarios.

La planificación temporal se puede dividir en varias secciones. Durante la planificación de la ontología se buscó el ámbito sobre el que realizar el trabajo, un tema sobre el que no existiesen ontologías ya diseñadas, y se definieron los requisitos planteando las preguntas de competencia. Se diseñó la ontología reutilizando los recursos útiles encontrados plasmandolos en el modelo conceptual. A continuación se creó la implementación del modelo y su evaluación. La redacción de este documento se puede considerar un ejercicio transversal.

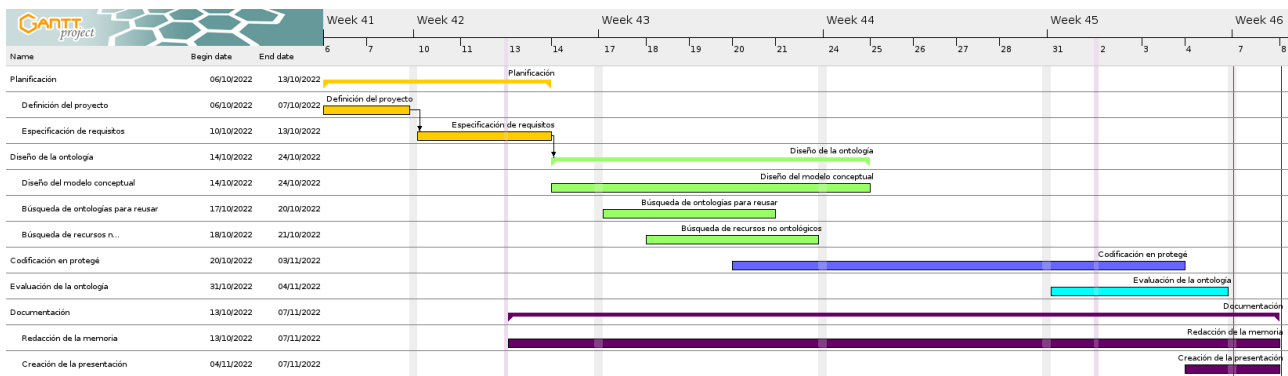


Figure 3: Diagrama de Gantt de la planificación temporal seguida.

5 Búsqueda de ontologías de alto nivel

Una vez identificados el glosario de términos mediante la especificación, y creado un modelo donde se relacionan las diferentes entidades, el siguiente paso es la formalización de la ontología. Para ello, hemos realizado una búsqueda de ontologías de alto nivel, detectando aquellas que cubren de manera total o parcial secciones de nuestro modelo para poder reutilizarlas.

Las primeras ontologías que hemos identificado para nuestra ontología, fueron mencionadas a lo largo del curso:

- **Org.** La ontología org ha sido utilizada para representar la entidad de la organización que lleva la instalación deportiva. No hemos utilizado la ontología completa, solo los módulos que nos hacían falta para representar la organización, su localización y sus miembros. Esta ontología es ampliamente utilizada: en LOV, podemos ver que 37 ontologías reutilizan contenido de org.
- **Foaf.** Para poder representar a todas las personas que pertenecen a nuestro modelo (director, entrenadores, monitores, deportistas...) hemos utilizado foaf. A efectos prácticos, nos ha servido con importar la ontología org, ya que esta contenía las entidades foaf que nos hacían falta
- **Odrl.** La ontología ODRL permite modelar reglas y políticas. Se ha utilizado para representar las reglas de vestimenta en los espacios deportivos. Para ello, se ha utilizado únicamente parte de la ontología, en concreto las entidades que son necesarias para generar el patrón Norm, que describiremos más adelante.
- **Geo.** La ontología Geo se ha utilizado para poder expresar la geolocalización de la instalación mediante la entidad GeoPoint. Esta ontología no tiene propiedades, así que hemos tenido que crear una propiedad que relacione la instalación con el GeoPoint.
- **Vcard.** Consiste en una ontología utilizada para representar la información que contiene una tarjeta de visita. En nuestro caso, hemos utilizado únicamente las entidades referentes a la dirección (Address) y nombre (OrganizationName).

Una vez identificadas estas ontologías de alto nivel, utilizamos el buscador de LOV (Linked Open Vocabularies) para buscar otras posibles ontologías que pudiéramos reutilizar. En concreto encontramos las siguientes:

- **Time.** La ontología time permite expresar tanto fechas como intervalos y duración de eventos. En nuestro caso, para expresar la fecha y la duración de los servicios hemos utilizado dos entidades diferentes.
- **Sport.** Esta ontología estaba disponible en el buscador, pero las URIs no llevaban a ninguna ontología, por lo que no pudimos reutilizarla.

A la hora de importar todas estas ontologías en protégé, utilizamos LOV como método de búsqueda de las URIs, lo que facilitó la tarea.

6 Recursos no ontológicos

Una de las fuentes de información utilizadas para generar la ontología propuesta en este documento es el siguiente informe [1]. A partir de este recurso no ontológico, mediante el uso de un T-Box se pudo realizar una modelización de la información descrita que representa la estructura taxonómica del documento en esta ontología de instalaciones deportivas.

El documento [1] del Ministerio de Cultura y Deporte del Gobierno de España, contiene los principales resultados del informe de explotación estadística del censo de instalaciones deportivas de 2005, así como las definiciones de las diferentes clases de instalaciones deportivas consideradas durante el censo.

Al plasmar el conocimiento presente en este documento en el modelo de la ontología mediante un T-Box, se decidió mantener las clases intermedias presentes en el documento como parte de la ontología pese a que serán probablemente rara vez utilizadas para mantener la estructura del documento mencionado y facilitar la reutilización de este.

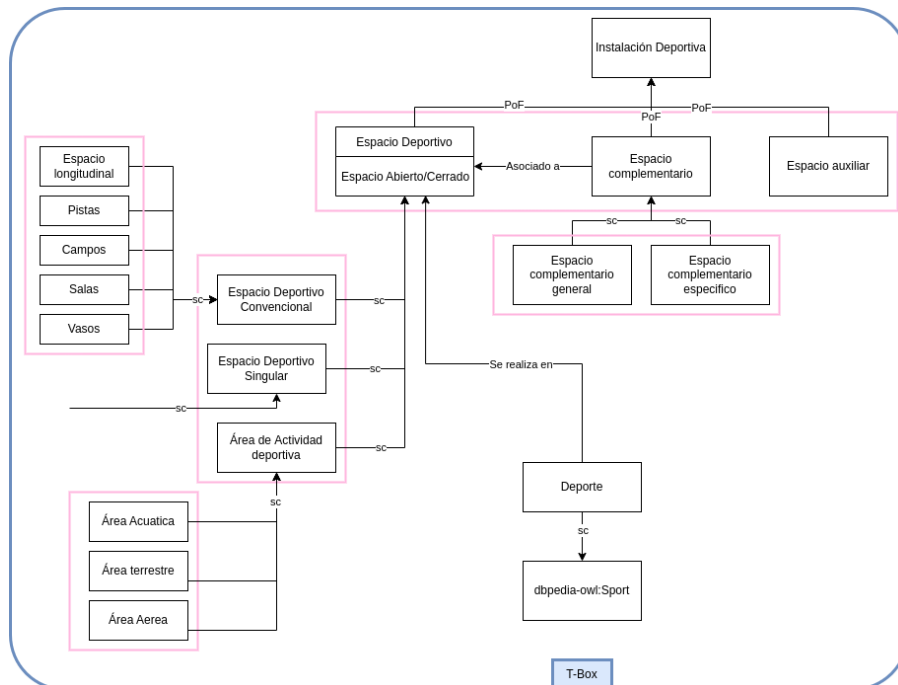


Figure 4: Sección correspondiente a los recursos no ontológicos en la ontología.

Es importante señalar que la estructura taxonómica del documento descrito exige deambiguar entre las relaciones "sub-clase de" (*sc* en el diagrama) y "parte de" (*PoF* en el

diagrama).

7 Patrones en la ontología

Para facilitar el diseño de los servicios ofrecidos por la organización en una instalación deportiva, se reutilizó el patrón de una relación N-aria según lo visto en clase. Este patrón se utiliza para representar una relación N-aria en el que todos los elementos tienen la misma importancia. Para representar esta relación N-aria se crea una clase, en nuestra ontología la clase *ServicioOfrecido*, a la que asocia todos los atributos de la relación.

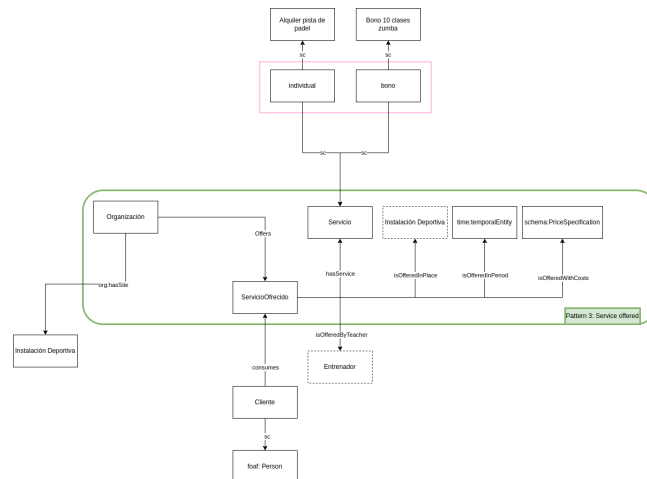


Figure 5: Sección correspondiente a los patrones utilizados en el diseño de la ontología.

8 Modelo conceptual

9 Clases Multilingües

10 Implementación de la ontología con OWL

Una vez identificados el glosario de términos mediante la especificación, y hemos conceptualizado un modelo donde se relacionan las diferentes entidades, el siguiente paso, la formalización de la ontología. Para ello, hemos realizado una búsqueda de ontologías de alto nivel, de forma que si alguna de ellas cubre toda o parte de alguna de nuestras entidades, podamos **reutilizarla**.

Las primeras ontologías que hemos identificado para nuestra ontología, fueron mencionadas a lo largo del curso:

- **Org:** La ontología org ha sido utilizada para representar la entidad de la organización que lleva la instalación deportiva. No hemos utilizado la ontología completa, solo los módulos que nos hacían falta para representar la organización, su localización y sus miembros. Esta ontología es ampliamente utilizada: en LOV, podemos ver que 37 ontologías reutilizan contenido de org.

- **Foaf:** Para poder representar a todas las personas que pertenecen a nuestro modelo (director, entrenadores, monitores, deportistas...) hemos utilizado foaf. A efectos prácticos, nos ha servido con importar la ontología org, ya que esta contenía las entidades foaf que nos hacían falta
- **Odrl:** La ontología ODRL permite modelar reglas y políticas. Se ha utilizado para representar las reglas de vestimenta en los espacios deportivos. Para ello, se ha utilizado únicamente parte de la ontología, en concreto las entidades que son necesarias para generar el patrón Norm, que describiremos más adelante.
- **Geo:** La ontología Geo se ha utilizado para poder expresar la geolocalización de la instalación mediante la entidad GeoPoint. Esta ontología no tiene propiedades, así que hemos tenido que crear una propiedad que relacione la instalación con el GeoPoint.
- **Vcard:** Consiste en una ontología utilizada para representar la información que contiene una tarjeta de visita. En nuestro caso, hemos utilizado únicamente las entidades referentes a la dirección (Address) y nombre (OrganizationName).

Una vez identificadas estas ontologías de alto nivel, utilizamos el buscador de LOV (Linked Open Vocabularies) para buscar otras posibles ontologías que pudieramos reutilizar. En concreto encontramos las siguientes:

- **Time:** La ontología time permite expresar tanto fechas como intervalos y duración de eventos. En nuestro caso, para expresar la fecha y la duración de los servicios hemos utilizado dos entidades diferentes.
- **Sport:** Esta ontología estaba disponible en el buscador, pero las URIs no llevaban a ninguna ontología, por lo que no pudimos reutilizarla.

A la hora de importar todas estas ontologías en protégé, utilizamos LOV como método de búsqueda de las URIs, lo que facilitó la tarea. Como se ha especificado en los requisitos, la ontología estará implementada en OWL. Para eso, hemos utilizado protégé, un editor de ontologías open source.

El primer paso que hemos hecho ha sido incluir todas las clases de la ontología que hemos creado mediante el T-Box. Después, hemos reutilizado de las ontologías nombradas en el apartado 5 todas las clases que nos hacían falta. Para ello, hemos importado las ontologías con su URI y movido los axiomas necesarios a nuestra ontología.

En general, la metodología para importar las ontologías ha sido idéntica para todas: primero hemos importando la ontología al completo con su uri, moviendo los axiomas que necesitábamos, y después hemos eliminando el resto. La única excepción ha sido la ontología Time.

Hemos importado la ontología Time de forma completa, ya que utilizábamos la totalidad de sus entidades. Al importar la ontología, nos hemos dado cuenta de que muchas de las propiedades estaban incompletas: les faltaba el rango o el dominio. Hemos optado por mantenerla como estaba, ya que muchos de estos problemas se debían a términos deprecados, y no hemos encontrado una versión más reciente de la ontología.

Una vez importadas las clases necesarias, hemos definido aquellas que nos hacían falta para completar nuestro modelo, que no se encontraban ni en el T-Box ni en las ontologías de

alto nivel, como puede ser por ejemplo la clase Monitor, Entrenador, Pista de Baloncesto... A continuación se muestra una imagen de la jerarquía de las clases.

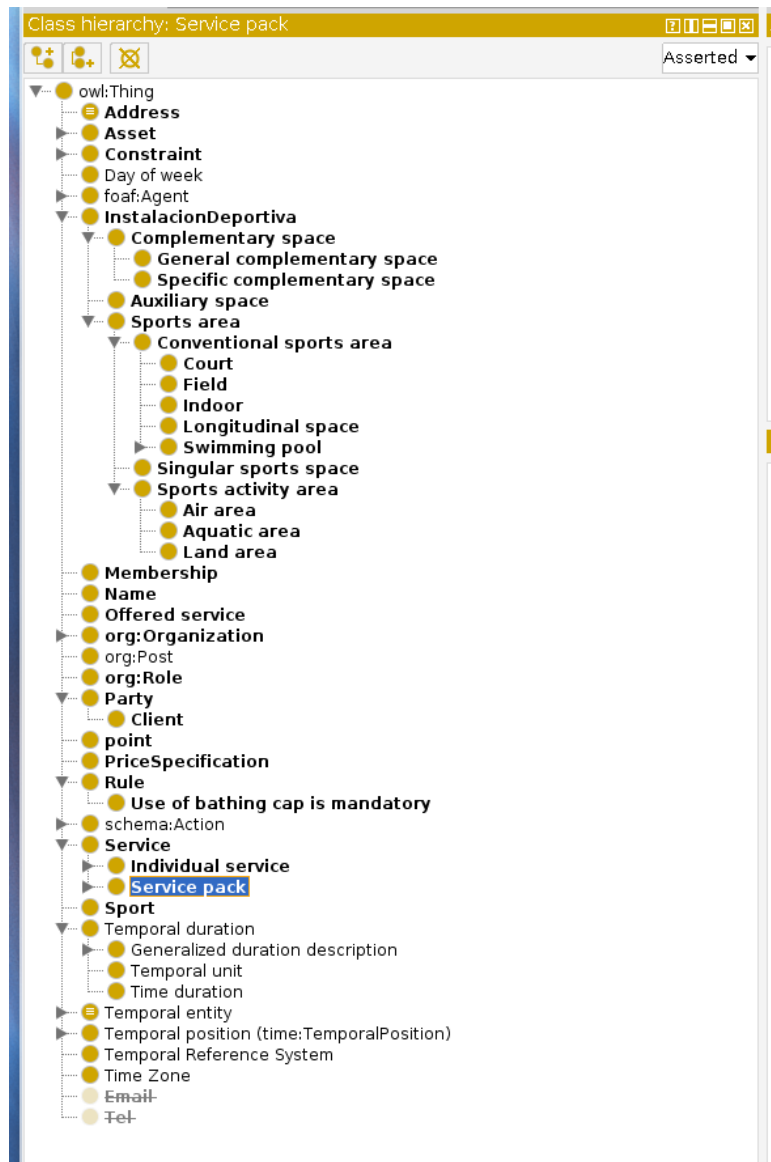


Figure 6: Classes

Este mismo procedimiento lo hemos realizado para los Object Properties. En este caso, a demás de importar los necesarios de las diferentes ontologías y crear los que necesitabamos, hemos definido las propiedades de cada propiedad.

Hemos tenido cierta dificultad a la hora de crear la propiedad Part Of, ya que tiene varios dominios para un único rango (tanto los espacios deportivos, como los complementarios y auxiliares son part of de una instalación deportiva). Para evitar interferencias, hemos incluido los diferentes dominios con la condición OR.

En la imagen a continuación se pueden ver los diferentes Object Properties de la ontología.



Figure 7: Object Properties

De forma similar hemos añadido los Data Properties. Estos en lugar de apuntar a una entidad, apuntan a un literal, por lo que hemos tenido que especificar los tipos de datos para literales como la divisa, el precio... Un caso particular ha sido la propiedad `hasCeiling` (tiene techo), para definir los Espacios Deportivos. En este caso el literal era un booleano. A continuación se muestran las diferentes Data Properties.



Figure 8: Data Properties

Finalmente, hemos completado las labels y comentarios de aquellas entidades que no han sido directamente importadas, asegurandonos de que estaban disponibles tanto en inglés como español.

Una vez finalizada la implementación de la ontología, la hemos exportado como RDF para poder evaluarla con OOPS!

11 Evaluación de la ontología con OOPS!

La ontología diseñada se ha evaluado utilizando la herramienta Ontology Pitfall Scanner! (OOPS!)[2], que identifica errores en la ontología dividiéndolos en críticos, importantes y menores, de acuerdo a una batería errores comunes.

11.1 Mejoras implementadas tras la sugerencia de OOPS!

Tras la primera iteración de la evaluación con OOPS!, la herramienta destacó los siguientes problemas críticos:

- **P19:** Propiedades con múltiples rangos o dominios.

Una de las relaciones definidas tenía más de un dominio, habiendo escrito en Protegé accidentalmente *and* en lugar de *or* en el dominio. Este error fue solucionado.

- **P29:** Relaciones transitivas mal definidas.

Las relaciones *partOf*, *hasService*, *managed_by* y *constraint* estaban definidas como transitivas, teniendo rangos y dominios distintos. Por definición una propiedad transitiva debe tener el mismo rango y dominio. Este error fue solucionado.

Gracias a los errores destacados por OOPS! se pudieron solucionar todos los errores críticos y varios errores clasificados como importantes. Se decidió no solucionar los errores no críticos de las ontologías importadas debido a las restricciones temporales sobre la entrega del trabajo.

11.2 Resultados finales de la evaluación

Tras aplicar las correcciones sobre las sugerencias de OOPS! la ontología tiene **xx** errores importantes y **xx** errores menores. Los resultados completos de esta evaluación se pueden encontrar en el Anexo.

12 Documentación de la ontología

[3]

13 Conclusiones

Bibliografía

References

- [1] Ministerio de Cultura y Deporte, “Anuario de Estadísticas Deportivas 2022. Instalaciones y espacios deportivos.” <https://www.culturaydeporte.gob.es/dam/jcr:90d8389b-60eb-41b4-96c1-6f4734d356e5/nota-resumen-instalaciones-y-espacios-deportivos.pdf>, 2022, (Accessed on 18/10/2022).
- [2] M. Poveda-Villalón, A. Gómez-Pérez, and M. C. Suárez-Figueroa, “OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation,” *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 10, no. 2, pp. 7–34, 2014.
- [3] Garijo, Daniel, “WIDOCO: a wizard for documenting ontologies,” in *International Semantic Web Conference*. Springer, Cham, 2017, pp. 94–102. [Online]. Available: <http://dgarijo.com/papers/widoco-iswc2017.pdf>

Anexos

Figure 9: Resultados finales de la evaluación con OOPS!