

Ontología de monumentos de la villa de Madrid

Autor: Víctor Iglesias Castán

Tutora: Claudia Reina Fajardo

Profesor: Joan Arnedo Moreno

Grado de Ingeniería Informática

Computación

Enero del 2025

Créditos/Copyright



Esta obra está sujeta a una licencia de [Atribución / Reconocimiento – NoComercial - SinDerivados 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	Ontología de monumentos de la villa de Madrid
Nombre del autor:	Víctor Iglesias Castán
Nombre del colaborador/a docente:	Claudia Reina Fajardo
Nombre del PRA:	Joan Arnedo Moreno
Fecha de entrega (mm/aaaa):	01/2025
Titulación o programa:	Grado de Ingeniería Informática
Área del Trabajo Final:	Web semántica
Idioma del trabajo:	Español
Palabras clave:	Ontología Monumentos
Resumen del Trabajo (máximo 250 palabras): Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo	
Abstract (in English, 250 words or less):	

Abstract

Texto con la síntesis del proyecto, esto es, un texto en el cual se explica de manera concisa la definición del proyecto/problema abordado, sus objetivos/métodos de resolución, y los resultados y conclusiones (no puede ser una lista, sino un texto continuo redactado de manera estructurada). Si es necesario poner una referencia en este texto, ésta será anotada a pie de la misma página. En este apartado se puede usar un lenguaje más literario y coloquial que para el resto del documento.

El Abstract se escribirá por duplicado. Una de las versiones tiene que ser obligatoriamente en inglés. La otra versión tiene que estar escrita en catalán o español. En caso de no escribir el resto del documento en inglés, será necesario escribir la segunda versión del Abstract en el idioma utilizado para el resto de la memoria. La palabra Abstract se cambiará por “Resum” o “Resumen” en la versión catalana y española, respectivamente.

Extensión recomendada: 250 palabras máximo.

Como escribir un buen Abstract (en inglés):

<http://www.ece.cmu.edu/~koopman/essays/abstract.html>

Palabras clave

Ontología, Web semántica, Monumentos, Datos abiertos, Linked data, Trabajo de Final de Grado

Sumario

1	Introducción.....	7
1.1	Prefacio.....	7
1.2	Descripción.....	8
1.3	Objetivos generales.....	9
1.3.1	Objetivos principales.....	9
1.3.2	Objetivos secundarios.....	10
1.4	Metodología y proceso de trabajo.....	10
1.4.1	Requisitos.....	11
1.4.2	Dominio, alcance, usuarios y usos.....	12
1.5	Planificación.....	13
1.5.1	Tareas.....	13
1.5.2	Hitos.....	14
1.5.3	Diagrama de Gantt.....	14
1.6	Presupuesto.....	15
1.6.1	Equipamiento técnico.....	16
1.7	Estructura del resto del documento.....	17
2	Análisis de mercado.....	17
2.1	Web semántica.....	17
2.2	Ontología.....	18
2.3	XML.....	19
2.4	RDF y RDF Schema.....	20
2.5	OWL.....	20
2.6	SPARQL.....	21
2.7	Datos enlazados.....	21
2.8	Datos abiertos y conocimiento abierto.....	22
3	Propuesta.....	22
4	Diseño.....	23
	Bibliografía.....	26

Figuras y tablas

Índice de figuras

Figura 1: Diagrama de Gantt..... 14

Figura 2: Pila tecnológica de la web semántica.....18

Figura 3: Tripleta como grafo dirigido.....20

Índice de tablas

Tabla 1: Tareas por hito.....13

Tabla 2: Cronología de los hitos..... 14

Table 3: Horas por tarea.....16

Tabla 4: Campos del fichero CSV..... 24

1 Introducción

1.1 Prefacio

El presente documento es la memoria de mi trabajo de final de grado de ingeniería informática. Éste trabajo pretende explorar las posibilidades que ofrece actualmente la web semántica, sus herramientas y los datos abiertos. Para ello, se propone crear una ontología que pueda tener su interés y utilidad dentro del mundo de las ontologías, así como ser una pieza clave dentro de un entorno de aplicación cliente / servidor que empodere a las personas.

A partir de los datos abiertos acerca de los monumentos localizados en la villa de Madrid, se definirán las clases, jerarquías, propiedades y restricciones de una ontología. Este modelo no se diseñará de forma que quede circunscrito a los datos de Madrid, sino que será generalizado para que pueda contener monumentos de todas las ciudades del mundo. A continuación, la ontología se poblará con los datos de la capital.

Finalmente, se pondrán a prueba las posibilidades que ofrece la web semántica mediante la integración de la información contenida en la Wikidata acerca de monumentos. La Wikidata contiene información estructurada que mediante consultas semánticas puede relacionarse con los contenidos de la ontología, de tal manera que se enriquece la información disponible.

1.2 Descripción

En los desplazamientos por las ciudades, ya sea en la localidad de residencia del transeúnte o en una visita debida a diferentes motivos como turismo o trabajo, es frecuente encontrarse con monumentos. El espacio público no es un museo, por lo que toda la información acerca del monumento está contenida en una placa conmemorativa o inscripción, por lo que el observador carece de información más detallada, desde su contexto histórico a una breve biografía de su autor, o incluso a quién o qué se está homenajeando. De hecho, no es infrecuente que tan solo se tenga un conocimiento superficial incluso de monumentos por los que se pasa casi a diario.

Actualmente, cada vez más ayuntamientos se suman a la filosofía de datos abiertos, y comparten de forma abierta toda clase de datos acerca de su localidad, como monumentos, jardines y demás puntos de interés. Estos datos permitirían poblar una ontología de monumentos. Gracias a la omnipresencia del móvil y otros dispositivos portátiles, un transeúnte podría acceder a la información contenida en la ontología acerca del monumento que está viendo en ese momento mediante una aplicación móvil, ya sea a partir de las coordenadas de su posición actual o buscándolo por su nombre.

Además, la información contenida en dicha ontología podría ampliarse más allá de los datos suministrados por los entes públicos. Por un lado, la Wikipedia es el recurso colaborativo más grande construido por la humanidad y en ella se encuentra información detallada acerca de muchos monumentos de ciudades alrededor del mundo. Por otro lado, tenemos la Wikidata, un proyecto que sigue el paradigma de los datos enlazados. En su creación se realizó una gran importación de datos desde la Wikipedia y a partir de ahí siguió incorporándose más información. La estructura de la información que contiene está pensada para que pueda ser consumida por aplicaciones y servicios de la web semántica, así como sistemas de inteligencia artificial.

Finalmente, el hecho de tener toda esta información contenida en una ontología, la hace accesible a otras ontologías, programas y sistemas informáticos, por ejemplo relacionados con el urbanismo, el turismo, etc. Esto no es casual, pues las ontologías son la principal herramienta de la web semántica para conseguir su objetivo: que las aplicaciones puedan interpretar, compartir y reutilizar la información.

El presente trabajo no se centra en la realización de la aplicación móvil (lado cliente) ni de la aplicación del lado servidor, pero sí incluye la ontología a la que deberá consultar la aplicación servidor para satisfacer las peticiones del cliente, el poblado de esta con datos provenientes del Portal de datos abiertos del Ayuntamiento de Madrid, así como demostraciones de las consultas en SPARQL a la ontología, así de cómo se puede ampliar la información que contiene acerca de un monumento con la contenida en la Wikidata.

1.3 Objetivos generales

1.3.1 Objetivos principales

Los objetivos principales son dos. En primer lugar, crear una ontología de monumentos. Esta se diseñará para poder contener información de monumentos de cualquier ciudad del mundo, aunque por las constricciones de tiempo, para este proyecto se poblará sólo con los datos abiertos que proporciona Madrid. En segundo lugar, enriquecer la información contenida en la ontología con la información de la Wikidata mediante consultas semánticas en SPARQL.

El objetivo desde la perspectiva del usuario será poder obtener información acerca del monumento que está viendo: historia, descripción de los elementos, biografía de su autor, etc. Este objetivo se conseguirá mediante una aplicación para dispositivos móviles cuyo desarrollo no está en el alcance del presente trabajo. Esta aplicación del lado cliente consultará a una aplicación del lado servidor que a su vez obtendrá la información de la ontología.

En cuanto a los objetivos personales, por un lado y a pesar de llevar muchos años trabajando como programador, la mayoría dedicados al desarrollo web, me llama la atención la web semántica. En consecuencia, con este trabajo puedo mejorar mis conocimientos acerca de la web semántica y las tecnologías relacionadas: las ontologías, el estándar de notación RDF con sus tripletas y su representación en forma de grafo, el estándar OWL, el lenguaje SPARQL para hacer consultas sobre información basada en RDF y el concepto de datos abiertos.

Por otro lado, me llamó mucho la atención el proyecto de la Wikipedia desde que lo conocí hace muchos años, pues me pareció una forma excelente de hacer llegar conocimientos a cualquier ser humano con acceso a Internet, pudiendo así vencer las limitaciones económicas y físicas que padezca la persona en su entorno. Lamentablemente, en la última década los usuarios cada vez pasamos más tiempo en plataformas cuyos contenidos mayoritarios están pensados sólo para entretener, por ejemplo TikTok, Instagram o Facebook, de manera que parte de Internet se ha convertido en una nueva “caja tonta”. Frente a esta realidad, pueden explorarse los proyectos DBpedia y Wikidata como herramientas para enriquecer otras aplicaciones, colaborando así a que Internet mantenga el potencial que se le vio desde sus inicios para mejorar la cultura y la formación de las personas.

1.3.2 Objetivos secundarios

Esta ontología, a su vez, puede ser integrada en ontologías de dominio más general, por ejemplo de turismo, urbanismo o geografía. Esta posibilidad no es para nada algo secundario sino que la interoperabilidad es una de las aplicaciones de la web semántica. La única razón por la que se ha listado como objetivo secundario es porque no será abordado en este trabajo, si bien será posible pues la ontología será públicamente accesible y con una licencia que lo permite.

Finalmente, hay dos objetivos más que podrían lograrse si este proyecto avanza más allá del presente trabajo y llega a los usuarios. En primer lugar, el hecho de que las personas tengan un mayor conocimiento del arte y la cultura de su entorno favorecerá que haya mas atención en el cuidado y protección del patrimonio de nuestras ciudades, más allá de las localizaciones que atraen el turismo de masas. En segundo lugar, este sistema es más eficiente desde el punto de vista energético y de consumo de materiales que si los usuarios hacen las consultas en Google, no se diga si las hacen en una IA entrenada con cantidades ingentes de datos sin clasificar provenientes de Internet.

1.4 Metodología y proceso de trabajo

A la hora de crear una ontología existen diversas metodologías. Algunas de ellas resultado de la experiencia obtenida en el desarrollo de un proyecto concreto y que pueden adaptarse mejor a proyectos de características similares; otras con un enfoque más generalista. Algunas metodologías son: Ontology Development 101, Methontology, On-To-Knowledge, Text2Onto, SENSUS-Based y Grüninger y Fox. Algo frecuente en estas metodologías es que son un proceso iterativo, característica que no por casualidad es compartida con el desarrollo de software.

Probablemente, la que mejor encaja con este proyecto es Methontology [1], propuesta por el Laboratorio de Inteligencia Artificial de la Universidad Politécnica de Madrid. Esta metodología propone un ciclo de vida basado en iteraciones, donde cada iteración consiste en el mismo conjunto de actividades o pasos. Las actividades están inspiradas en la norma IEEE 1074 para el proceso de desarrollo de software. El fruto de cada iteración es un prototipo de la ontología.

Las razones por las que encaja mejor son dos:

1. El desarrollo de esta ontología se ve favorecido por un proceso iterativo. Por ejemplo, al crear las instancias mediante la carga de datos del fichero CSV, podrían descubrirse deficiencias en la ontología que deberán subsanarse.
2. Esta ontología es ampliable mediante la incorporación de nuevas fuentes de datos abiertos. Nuevas propiedades de los mismos podrían requerir cambios en la ontología, por lo que es correcto verla como un prototipo que evoluciona a medida que se enriquece con nuevas

fuentes de datos. En este sentido, la ontología resultante de este proyecto hay que verla como un prototipo.

Las actividades o pasos son:

- **Especificación:** Define el alcance de la ontología. En esta fase se debe decidir y delimitar:
 - Los objetivos de su creación.
 - El dominio de actuación.
 - Por quién y para qué será utilizada.
- **Conceptualización:** Se debe organizar y estructurar el conocimiento adquirido mediante la creación de un glosario de términos que pertenezcan al dominio, definirlos y estructurarlos. De las diferentes herramientas para lograr este objetivo, para este trabajo se empleará el diseño UML.
- **Formalización:** El modelo conceptual de la actividad anterior se pasa a un modelo formal, más riguroso.
- **Integración:** Debe contemplarse si existen y pueden integrarse otras ontologías para reducir costes y evitar duplicidades.
- **Implementación:** Se convierte el modelo formal en un modelo computable. En este trabajo se empleará Protégé para lograrlo. Además, en el desarrollo mediante Protégé que se realizará, se solaparán las actividades de formalización e implementación.
- **Evaluación:** Debe verificarse que la ontología cumpla con los objetivos planteados en la especificación.
- **Documentación:** Para que la ontología se pueda compartir y reutilizar, se debe documentar. Esta memoria será la documentación de la ontología.
- **Mantenimiento:** Consiste en establecer cuáles líneas de mantenimiento y modificación se llevarán a cabo. Esta actividad está fuera del alcance del presente trabajo, corresponderá a quienes reutilicen o amplíen la ontología en el futuro.

1.4.1 Requisitos

El producto final consta de la ontología como parte fundamental. En segundo lugar, consta de las consultas en SPARQL que sirven como prueba de concepto del potencial de las ontologías y el linked data.

Los requisitos de la ontología son los siguientes:

- Contener información acerca de los monumentos de Madrid provenientes de datos abiertos: nombre, descripción, año, dirección, barrio, distrito y enlace a la ficha del monumento en el sitio web “Patrimonio cultural y paisaje de Madrid”.
- Ser ampliable a monumentos de otras ciudades.
- Los monumentos estarán clasificados por tipos.

Los requisitos que deben cumplir las consultas son los siguientes:

- Localizar el monumento cercano a una geolocalización y proporcionar la información disponible en la ontología de ese monumento.
- Encontrar monumentos por nombre, barrio, distrito, año, autor y ciudad.
- Ampliar la información disponible mediante consultas cruzadas con el punto de acceso SPARQL de la Wikidata, de manera que se pueda incluir su historia, descripción de sus elementos, biografía del autor o autores, etc.

1.4.2 Dominio, alcance, usuarios y usos

El dominio del proyecto son los monumentos localizados en Madrid. El alcance del proyecto es la creación de la ontología de monumentos de Madrid a partir de los datos abiertos disponibles. Este alcance implica una primera iteración del producto, aunque tal vez requiera retornar a una actividad anterior cuando se descubra que son necesarios cambios. El alcance no incluye las nuevas iteraciones que requieran la incorporación de más datos de otras ciudades: se limita estrictamente a Madrid.

El perfil de los usuarios finales del producto desarrollado será técnico, ya sea para ampliar la ontología con más instancias o integrarla en otra ontología, o ya sea para crear las aplicaciones lado cliente y servidor, en ambos casos se requieren personas de suficiente perfil técnico.

Los posibles usos del producto final son tres:

- Ampliar la ontología con monumentos de más ciudades.
- Integrarla en ontologías de propósito más general.
- Crear una aplicación que permita a usuarios finales obtener información desde su móvil del monumento que están contemplando.

1.5 Planificación

La planificación del proyecto se estructura en hitos, los cuales derivan de las entregas de las PEC. Por cada hito se han definido sus tareas. La visualización en forma gráfica del espacio temporal de las tareas se realiza mediante el diagrama de Gantt.

1.5.1 Tareas

A continuación, se muestran las tareas requeridas de cada PEC o hito:

PEC 1
Lectura de los materiales de las asignaturas TFG y Web semántica
Informarse acerca de la web semántica, los datos abiertos, las ontologías y tecnologías relacionadas
Elaboración de una propuesta inicial
PEC 2
Análisis de la información acerca de monumentos disponible en ficheros open data
Análisis de la información acerca de monumentos contenida en Wikidata y en DBpedia
Elaboración del estado del arte
Crear proyecto en Github
PEC 3
Describir mediante UML el modelo que contiene toda la información necesaria
Creación de la ontología con Protégé a partir del diseño UML
Programación del script de importación de datos
Creación de consultas que demuestren las posibilidades de la ontología y su integración con Wikidata
PEC 4
Pulir la versión definitiva de la ontología
Versión definitiva en Github: fichero de la ontología y fichero README.md
Pulir la versión definitiva de la memoria
Creación del vídeo de defensa del TFG
Creación del vídeo tráiler del proyecto
PEC 5
Preparación de la defensa del trabajo
Defensa síncrona del trabajo

Tabla 1: Tareas por hito

1.5.2 Hitos

Los hitos coinciden con las entregas parciales de las PEC:

Tarea	Fecha inicio	Fecha fin
PEC 1	25/09/2024	20/10/2024
PEC 2	21/10/2024	17/11/2024
PEC 3	18/11/2024	08/12/2024
PEC 4	09/12/2024	12/01/2025
PEC 5	13/01/2025	?

Tabla 2: Cronología de los hitos

1.5.3 Diagrama de Gantt

A partir de las tareas e hitos definidos previamente, se ha elaborado el siguiente diagrama de Gantt:

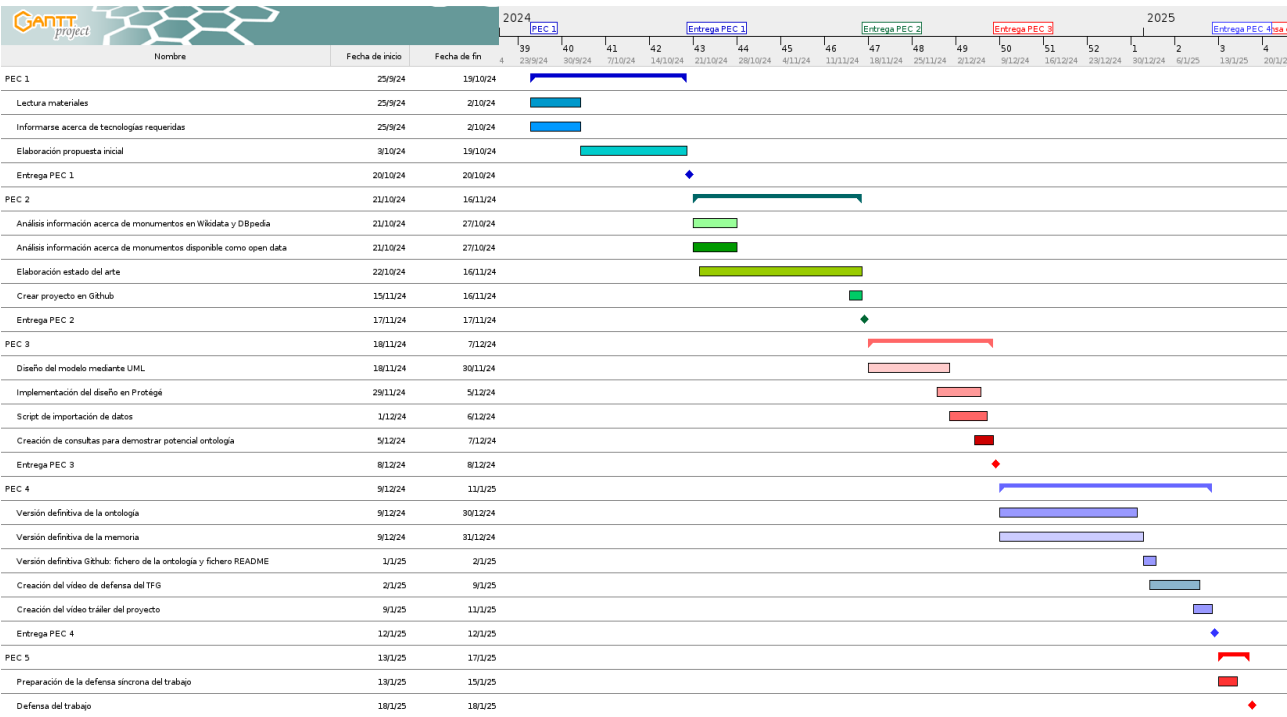


Figura 1: Diagrama de Gantt

Como todavía se desconoce la fecha exacta de la defensa del trabajo, se ha planteado la fecha del 18 de enero de 2025.

La planificación (tareas e hitos) se ha hecho acorde con los enunciados de cada PEC: estos especifican qué partes del proyecto se deben realizar. Ahora bien, existe cierta coincidencia entre la planificación dada y la metodología Methontology:

- La PEC 1 coincide con la actividad de especificación, así como las 2 primeras tareas de la PEC 2.
- La tarea del estado del arte coincide con la actividad de integración, pues tal vez se encuentre una ontología que pueda reutilizarse.
- En la PEC 3 se llevan a cabo las actividades de formalización e implementación. Las consultas que se crearán también servirán para realizar la actividad de evaluación de la ontología.
- En la PEC 4 aparecen las tareas de creación de las versiones definitivas de la ontología y la memoria, las cuales pueden verse como otra iteración después de la evaluación, pues según se haya visto en las consultas, podría ser necesario regresar a la actividad de conceptualización y avanzar hasta la de evaluación.

Un posible riesgo está en cumplir con los plazos de la PEC 3, pues a efectos prácticos concentra prácticamente todo el desarrollo del producto y tan sólo se dispone de 20 días.

1.6 Presupuesto

La limitación en este proyecto es de tiempo, éste viene dado por el total de créditos ECTS de la asignatura: 12. En la UOC, un crédito ECTS equivale a 25 horas de trabajo del estudiante [2], por lo que se dispone de 300 horas para este proyecto. Además, las fechas de entrega de los hitos también vienen estipuladas.

El periodo desde el inicio del semestre hasta el hito de la PEC 4 es de 106 días. De ahí a la PEC 5 se han puesto 5 días, si bien se desconoce la fecha exacta de la defensa del trabajo, es un tiempo razonable para la defensa del trabajo y la preparación de esta. Por lo tanto, puede considerarse que se disponen de 111 días.

El equipo humano del proyecto está compuesto por una única persona, por lo que las tareas del proyecto se desarrollarán en función de mi disponibilidad. Actualmente trabajo a jornada completa 40 horas semanales, además de cursar otras 2 asignaturas, pero dispongo de días de vacaciones para avanzar en el proyecto. Aproximadamente son 2,7 horas por día, unas 19 a la semana.

La siguiente tabla desglosa las tareas del proyecto y sus horas. A partir de aquí se podría realizar un presupuesto asignando a cada tipo de tarea un precio por hora.

Hito	Actividad	Horas
PEC 1		
	Lectura de los materiales	10
	Informarse acerca de tecnologías requeridas	11
	Elaboración propuesta inicial	46
PEC 2		
	Análisis información acerca de monumentos en Wikidata y DBpedia	9
	Análisis información acerca de monumentos disponible como open data	10
	Elaboración estado del arte	56
PEC 3		
	Diseño del modelo mediante UML	30
	Implementación del diseño en Protégé	16
	Script de importación de datos	6
	Creación de consultas para demostrar potencial ontología	10
PEC 4		
	Versión definitiva de la ontología	28
	Versión definitiva de la memoria	33
	Versión definitiva Github: fichero de la ontología y fichero README	2
	Creación del vídeo de defensa del TFG	21
	Creación del vídeo tráiler del proyecto	8
PEC 5		
	Preparación de la defensa síncrona del trabajo	5

Table 3: Horas por tarea

1.6.1 Equipamiento técnico

El equipamiento técnico se divide en software y hardware. En lo que respecta al último, basta con una estación de trabajo sin especiales prestaciones. Concretamente, se dispone de un equipo con CPU Intel Core i5-4460 a 3,2 GHz, 16 GB de RAM y 1TB de disco duro.

Respecto al software, el sistema operativo es Linux, si bien el software requerido es agnóstico del sistema operativo. El diseño UML de la ontología se realizará con la herramienta en línea “draw.io” y se implementará con el programa Protégé, posiblemente se deba añadir una extensión para poder trabajar con GeoSPARQL. No obstante, Protégé no tiene la capacidad de acceder a servicios remotos para realizar consultas distribuidas que integren la información contenida en la ontología con la de DBpedia o Wikidata, por lo que se empleará Apache Jena Fuseki como servidor SPARQL. GeoSPARQL también puede ser integrado en Fuseki, se analizará esta opción. Si bien Virtuoso podría funcionar mejor cuando hay un gran volumen de datos, Fuseki será suficiente para la prueba de concepto del presente trabajo.

1.7 Estructura del resto del documento

Breve descripción de los otros capítulos de la Memoria.

Explicación de los contenidos de cada capítulo y su relación con el trabajo en global.

2 Análisis de mercado

En este apartado se explica el concepto de web semántica, así como las tecnologías y estándares relacionados en los que se basa el presente trabajo.

2.1 Web semántica

La web actual es la 2.0. En sus páginas se presenta la información de forma comprensible para los humanos. La web semántica no es una nueva web, una especie de versión 3.0, sino una extensión de la actual que la complementa, cuyo objetivo final es la interrelación de los recursos disponibles en la web, es decir, debe convertir la web en una red de interconexión de datos que puedan ser comprendidos por aplicaciones.

Actualmente, la información de la web no está estructurada, Además, contiene gran cantidad de texto en lenguaje natural, siempre ambiguo. En consecuencia, para que la web semántica sea una realidad, debe lograrse describir el significado, es decir, la semántica.

Por otro lado, además de haberse logrado dotar de significado a la web, también se debe disponer de aplicaciones y tecnologías que empleen la información semántica para ofrecer nuevos servicios. Este software es lo que el W3C denomina servicios web y agentes.

Si se logran ambos objetivos, sería realidad el siguiente ejemplo de una aplicación: el buscador inteligente. Mediante el mismo, un usuario formula que desea irse de vacaciones a una localidad el próximo fin de semana. El buscador sabrá cuáles son las fechas del próximo fin de semana y sugerirá unos vuelos. Una vez el usuario selecciona los que considera mejores, el buscador se encarga automáticamente de las reservas. A continuación, el buscador sugiere diferentes alojamientos y se encarga también de reservar los seleccionados.

En la siguiente figura se muestra la pila de la web semántica, tal y como la presentó Tim Berners-Lee en la presentación “Semantic Web – XML2000” [3]:

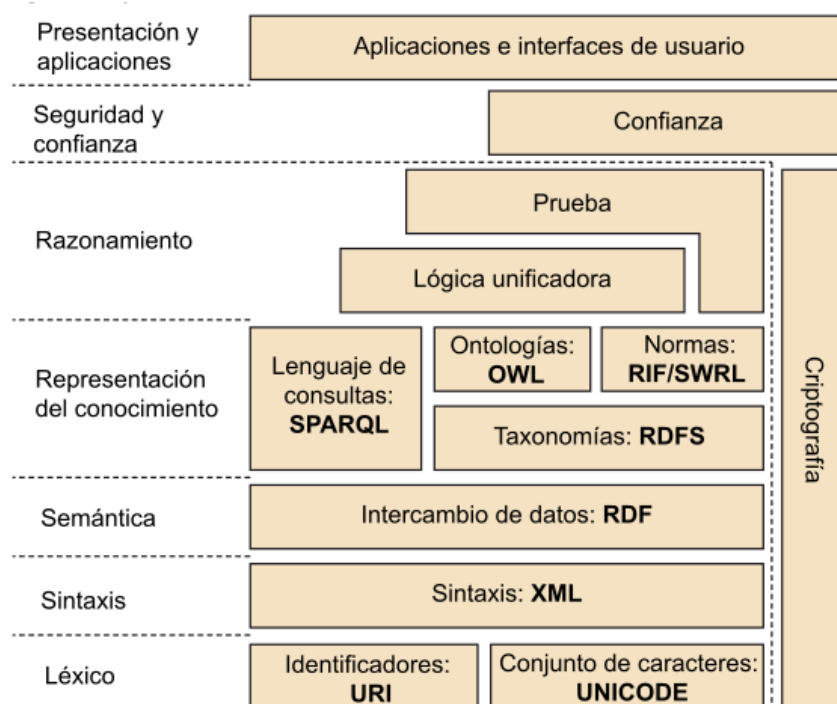


Figura 2: Pila tecnológica de la web semántica

En los siguientes subapartados se explican estas tecnologías y estándares.

2.2 Ontología

En filosofía, la ontología es la parte de la metafísica que estudia el ser. El ser abarca toda la realidad y todas las entidades que lo componen. La ontología examina qué tienen en común todas las cosas para así poder articular la estructura básica del ser. También investiga cómo pueden agruparse las cosas en tipos básicos, como las categorías de particulares y universales. Las particulares son entidades únicas, no repetibles, por ejemplo la persona Sócrates, mientras que las universales son generalizaciones, entidades repetibles, como el color azul. Otro contraste que realiza es entre objetos concretos que existen en el espacio y el tiempo, como un geranio, y objetos abstractos que existen fuera del espacio y del tiempo, como el número 3. Estos sistemas de categorías pretenden ofrecer un inventario exhaustivo de la realidad empleando categorías como sustancia, propiedad, relación, estado de cosas y acontecimiento.

Esta introducción no es baladí, pues la ontología en informática guarda cierta similitud con su homónima en la metafísica. Aquí, una ontología consiste en una definición formal de tipos, propiedades y relaciones entre entidades que existen para un dominio en particular [4]. Las tres principales propiedades de una ontología son:

- Utiliza una representación explícita, por lo que suele estar escrita en un lenguaje formal y en un soporte digital.
- Su conceptualización es compartida: representa la información que un conjunto de personas tienen respecto a un dominio de discurso, por lo que no puede representar el punto de vista de un único individuo.
- Representa el dominio de discurso relevante para un problema concreto.

La herramienta principal para conseguir el objetivo de la web semántica es la ontología, pues permite la definición formal de los conceptos del dominio, su jerarquía y las relaciones que los unen entre sí.

2.3 XML

Este acrónimo proviene de las siglas en inglés de *eXtensible Markup Language* y consiste en un lenguaje de marcas para almacenar, transmitir y reconstruir datos de cualquier índole, lo cual lo hace especialmente idóneo para el intercambio de información estructurada entre diferentes plataformas. Un lenguaje de marcas es aquel en el que se especifica la estructura y el formato del documento, y potencialmente las relaciones entre sus partes.

Un documento XML está formado por dos secciones principales:

- El prólogo: Aquí es donde se define que se trata de un documento XML, su versión, así como entradas opcionales con información acerca de su procesamiento. En el siguiente ejemplo se muestra un prólogo con las partes opcionales entre corchetes:

```
<?xml version="1.0" [encoding="UTF-8"] [standalone="yes"] ?>
```

- La estructura de datos o cuerpo del documento: contiene los datos y su jerarquía, es decir, contiene la información del documento. En el siguiente ejemplo se muestra una parte del cuerpo de un documento en el que dentro de la etiqueta “Personal” hay varias marcas “Empleado”, cada una con la misma estructura: NSS, Nombre, etc. Asimismo, la marca “Personal” podría estar formando parte de una estructura mayor:

```
<Personal>
  ...
  <Empleado>
    <NSS>111-11-1111</NSS>
    <Nombre>Ana</Nombre>
    <Apellidos>López Gómez</LastName>
    <Salario>65000.00</Salario>
    <Departamento>Contabilidad</Departamento>
  </Empleado>
```

```

    <Empleado>
      ...
    </Empleado>
  ...
</Personal>

```

Nótese que cada marca señala un segmento de información con un sentido claro y específico.

2.4 RDF y RDF Schema

Este acrónimo proviene de las siglas en inglés *Resource Description Framework* y es una notación estándar para la representación e intercambio de información sobre recursos, particularmente recursos web. Ésta notación representa un grafo dirigido compuesto de tripletas sujeto – predicado – objeto¹, en las que el sujeto y el objeto son nodos y el predicado es el arco que los une. Cada una de estas tres partes se identifica de forma inequívoca mediante un identificador de recursos uniforme, URI por sus siglas en inglés, para garantizar que no hay ambigüedades en la estructura semántica.

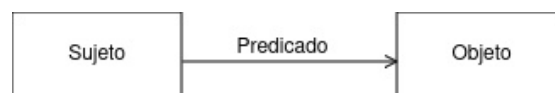


Figura 3: Tripleta como grafo dirigido

Por otro lado, *RDF Schema* (abreviado como RDFS) es una extensión de RDF para poder describir vocabularios de términos y relaciones sencillas entre ellos, por ejemplo jerárquicas. Para ello, proporciona los conceptos de clases y de propiedades, así como el de jerarquía entre ellas mediante las anotaciones `rdfs:Class`, `rdfs:subClassOf`, `rdf:property` y `rdfs:subPropertyOf`. Asimismo, se indica que un recurso es instancia de una clase mediante `rdf:type`. Nótese que tanto las instancias como las propiedades ya están incluidas en el estándar RDF.

2.5 OWL

Para poder representar definiciones y restricciones más complejas que las posibles con RDFS, existe una extensión del mismo: el estándar *Web Ontology Language* (OWL por sus siglas en inglés). Este estándar enriquece la definición de clases mediante las operaciones de unión, intersección y complementación de clases. Las propiedades son ampliadas con restricciones de cardinalidad y valores específicos, además de la existencia de las propiedades inversas, simétricas y transitivas. Asimismo, aporta el concepto de equivalencia tanto a las clases como a las propiedades.

¹ Curiosamente, en los lenguajes orientados a objetos, la llamada al método de un objeto tiende a seguir la estructura `sujeto.verbo(objeto)`

Finalmente, permite aplicar a la ontología herramientas de razonamiento inferencial, pues OWL está basado en la lógica descriptiva.

Gracias a estas características con las que amplía RDFS, la extensión OWL permite la definición de ontologías estructuradas basadas en la web, a diferencia de los lenguajes anteriores, que fueron pensados para comunidades específicas de usuarios, por ejemplo en el campo de las ciencias.

En los tres estándares descritos (RDF, RDFS y OWL), es necesario automatizar su interpretación, para lo cual deben almacenarse en un fichero. Los tres estándares definen el modelo de datos pero no su *serialización*, para la cual existen diferentes sintaxis como Notation3, Turtle y XML, siendo esta última la más usada y la que se empleará en este trabajo. No obstante, en vez de guardarse en un documento, también podría hacerse en un fichero con una *serialización* en formato binario.

2.6 SPARQL

SPARQL es un lenguaje para la consulta (*query*) de la información almacenada en el formato RDF, lo cual también incluye sus extensiones RDFS y OWL. Además de ser un lenguaje de consultas, también es el protocolo para acceder a datos RDF, tal y como indican sus siglas en inglés: *SPARQL Protocol and RDF Query Language*. Este lenguaje guarda ciertas similitudes con SQL, el lenguaje para hacer consultas en bases de datos relacionales, si bien en este caso los datos almacenados son tripletas.

Una de sus extensiones es GeoSPARQL, que permite representar y hacer consultas geoespaciales en el formato RDF, así como su integración en el contexto de los datos enlazados (*linked data*).

2.7 Datos enlazados

Los datos enlazados frecuentemente son referidos como datos vinculados o también por su nombre en inglés: *linked data*. Este término fue acuñado por Tim Berners-Lee en una nota de diseño acerca del proyecto de la web semántica. Su idea es que para conseguir que la web semántica sea una realidad, debe disponerse de un gran volumen de datos interconectados entre ellos y en un formato estándar, de manera que puedan ser tratados por ordenadores.

En la nota mencionada previamente, Tim Berners-Lee señaló estos cuatro principios:

1. Los recursos públicos de la web deben ser identificados mediante URIs.
2. Las URIs deben ser accesibles mediante el protocolo HTTP.
3. Cuando se accede a una URI, esta devuelve la información por medio de estándares como RDF o SPARQL.

4. Los datos deben incluir enlaces a otras URIs, es decir, a otros recursos externos, así se potencia el descubrimiento de datos relacionados en diferentes fuentes.

DBpedia y Wikidata son dos destacados ejemplos de proyectos que siguen el paradigma de los datos enlazados.

2.8 Datos abiertos y conocimiento abierto

Los datos abiertos (*open data* en inglés) son una filosofía y una práctica que persigue que ciertos datos estén disponibles para todo el mundo, sin restricciones como los derechos de autor o las patentes, mediante el empleo de licencias abiertas. Esta filosofía es parecida a la de otros movimientos “abiertos”, como por ejemplo el código abierto (*open source*), la educación abierta o el conocimiento abierto. Éste último amplía el concepto de datos abiertos: El conocimiento abierto consiste en una serie de principios y metodologías en la producción y difusión del conocimiento para garantizar el acceso, uso, modificación y redistribución de la información sin restricciones legales, sociales o tecnológicas.

3 Propuesta

Una vez explicadas las tecnologías y estándares en los que se basa el siguiente trabajo, se listan los elementos que se esperan obtener:

- Una ontología con los monumentos de la villa de Madrid.
- Un *script* de importación para cargar en la ontología los datos del fichero CSV del portal de datos abiertos del Ayuntamiento de Madrid.
- Un fichero con ejemplos de consultas SPARQL exclusivamente sobre la ontología y otras que también integren la información de la Wikidata para ampliar la información de la ontología, a modo de ejemplo de las posibilidades de la web semántica.

Sobre este último punto debe decirse que se hizo un pequeño estudio acerca de si era mejor hacer la integración de Wikidata o DBpedia. Dados los campos del fichero CSV (ver [punto 4](#)) y los campos disponibles en la DBpedia, se comprobó que sólo los campos de latitud y longitud podrían emplearse para relacionar ambas fuentes de datos, requiriendo además modificaciones, pues las coordenadas son mucho más precisas en el CSV. Mediante una consulta SPARQL se comprobó que de todos los monumentos ubicados en Madrid, la DBpedia sólo tiene las coordenadas de tres de ellos. Por otro lado, mediante consulta se comprobó que en la Wikidata muchos de los monumentos sí las tienen, aunque no todos.

Además, la Wikidata tiene estas otras ventajas:

- Se actualiza con mayor frecuencia y cuenta con una comunidad activa.
- De cada dato relacionado se muestra su identificador, con lo cual es fácil acceder a los datos del mismo. Así, por ejemplo, se puede mostrar la información del autor, así como más información de la cual carece la ontología.
- La documentación para hacer consultas es algo mejor.

Su principal desventaja es que hay más campos diferentes entre entidades (monumentos) que en la DBpedia, lo cual puede dificultar la integración de su información.

4 Diseño

Se continua con la fase de especificación de la metodología Methontology. Estos son los campos en el fichero CSV:

Campo	Comentarios	Tipo
PK	Se puede relacionar con el campo “catalogCode” de Wikidata	Entero
NOMBRE		Texto
SUBTIPO	Posibles valores: Escultura conceptual o abstracta Grupo Escultórico Elemento conmemorativo, Lápida Estatua Puente, construcción civil Puente, construcción civil Fuente, Estanque, Lámina de agua Elemento de ornamentación Edificación singular Puerta, Arco triunfal	Texto
DESCRIPCIÓN		Texto
FECHA	No es una fecha sino una descripción tipo “Cercano a 1808”	Texto
AUTORES	Primero los apellidos, después el nombre separado por una coma. Si hay más de uno se separan por un guión	Texto / array
PDF	URL del PDF que describe el monumento en la web de Patrimonio Cultural y Paisaje Urbano de Madrid	Texto
CONTENT-URL	URL a la web que describe el monumento en la misma web	Texto
NOMBRE-VIA		Texto
CLASE-VIAL	Posibles valores: Avenida Calle ...	Texto
TIPO-NUM	Posibles valores: V F	Texto
NUM		Entero
PLANTA		Vacío
PUERTA		Vacío
ESCALERAS		Vacío

ORIENTACIÓN		Vacío
LOCALIDAD	Siempre es “Madrid”	Texto
PROVINCIA	Siempre es “Madrid”	Texto
CODIGO-POSTAL		Entero
CODIGO-BARRIO		Entero
BARRIO		Texto
COD-DISTRITO		Entero
DISTRITO		Texto
COORDENADA-X		Entero
COORDENADA-Y		Entero
LATITUD		Decimal
LONGITUD		Decimal
TELÉFONO		Vacío
FAX		Vacío
EMAIL		Vacío
TIPO		Vacío

Tabla 4: Campos del fichero CSV

En el campo owl:sameAs de la DBpedia hay un enlace al elemento en la Wikidata para ese elemento, pero lamentablemente no desde Wikidata a DBpedia.

La combinación de la información de la ontología con la Wikidata se hará mediante la intersección de los campos “PK” y “catalogCode”. Incluso sería posible mediante “describedAtURL”, pues se trata de la URL a la página web del monumento en el sitio de Patrimonio y Paisaje de Madrid. No obstante, es más óptimo hacerlo mediante el identificador o clave primaria “PK”. Si en una futura ampliación de esta ontología se añadieran monumentos de otras ciudades que no pudieran relacionarse mediante el “catalogCode”, se deberán emplear las coordenadas.

A continuación, se listan campos en Wikidata que son interesantes para incorporar:

Campo	Comentarios
Autor / Arquitecto / Creador	Puede contener más de uno
Imagen	
Estilo arquitectónico	
Inicio (<i>Inception</i>)	Fecha de construcción
Evento significativo	Puede contener varios sucesos como elementos de la Wikidata (incendio, reconstrucción...) y el año del mismo

Depicts	Personaje histórico homenajeado
---------	---------------------------------

En general, la estructura de datos no es muy consistente en los diferentes monumentos, pues no todos tienen estos campos. Además, para unos se emplea el campo “Autor” y para otros “Creador”, sin que parezca que haya un criterio para usar uno u otro.

Bibliografía

- [1] «METHONTOLOGY: From Ontological Art Towards Ontological Engineering». [En línea]. Disponible en: https://oa.upm.es/5484/1/METHONTOLOGY_.pdf
- [2] «FAQ - Estructura de las enseñanzas - Calidad - (UOC)». Accedido: 15 de octubre de 2024. [En línea]. Disponible en: <https://www.uoc.edu/portal/es/qualitat/qualitat-titulacions/estructura-ensenyaments/faqs/faq-generals.html>
- [3] Robert Clarisó Viladrosa, Jordi Conesa i Caralt, y Robert Clarisó Viladrosa, «Ontologías y web semántica». Universitat Oberta de Catalunya.
- [4] «Ontología (informática)», *Wikipedia, la enciclopedia libre*. 14 de septiembre de 2024. Accedido: 31 de octubre de 2024. [En línea]. Disponible en: [https://es.wikipedia.org/w/index.php?title=Ontolog%C3%ADa_\(inform%C3%A1tica\)&oldid=162454337](https://es.wikipedia.org/w/index.php?title=Ontolog%C3%ADa_(inform%C3%A1tica)&oldid=162454337)
- [5] R. M. Kim Hamilton, *Learning UML 2.0*. [En línea]. Disponible en: <https://learning.oreilly.com/library/view/learning-uml-2-0/0596009828>
- [6] «Ontology Development 101: A Guide to Creating Your First Ontology». [En línea]. Disponible en: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- [7] *Methontology*. [En línea Video]. Disponible en: <https://tv.unir.net/videos/13955/47/106/2587/0/Methontology>
- [8] «MOD2-Ontologias-y-web-semantica».
- [9] Robert Clarisó Viladrosa, Jordi Conesa i Caralt, y Jordi Duran Cals, «Introducción a la representación del conocimiento». Universitat Oberta de Catalunya.
- [10] Robert Clarisó Viladrosa, «Introducción al trabajo final». Universitat Oberta de Catalunya.
- [11] José Ramón Rodríguez, «El trabajo final como proyecto». Universitat Oberta de Catalunya.
- [12] José Ramón Rodríguez, «La gestión del proyecto a lo largo del trabajo final». Universitat Oberta de Catalunya.
- [13] «Monumentos de la ciudad de Madrid - Portal de datos abiertos del Ayuntamiento de Madrid». Accedido: 20 de octubre de 2024. [En línea]. Disponible en: <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=eb8e993ae322b610VgnVCM1000001d4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnnextfmt=default>