

Implementación de Ontología aplicada a la clasificación de los palos flamencos.

Autor: Raúl López Vicente
Tutor: Felipe Geva Urbano
Profesor: Felipe Geva Urbano

Grado de Ingeniería Informática
Web Semántica / Computación



Esta obra está sujeta a una licencia de Reconocimiento- NoComercial-SinObraDerivada [3.0 España de Creative Commons.](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Implementación de Ontología aplicada a la clasificación de los palos flamencos</i>
Nombre del autor:	<i>Raúl López Vicente</i>
Nombre del colaborador/a docente :	<i>Felipe Geva Urbano</i>
Nombre del PRA:	<i>Felipe Geva Urbano</i>
Fecha de entrega (mm/aaaa):	<i>06/2024</i>
Titulación o programa:	<i>Grado Ingeniería Informática</i>
Área del Trabajo Final:	<i>Web Semántica</i>
Idioma del trabajo:	<i>Español</i>
Palabras clave	<i>Web Semántica, Ontologías, Palos Flamencos</i>
Resumen del Trabajo (máximo 250 palabras):	
<p><i>El Flamenco significa una forma de expresión artística que se desarrolla en diferentes ámbitos como lo son el cante, el toque o el baile.</i></p> <p><i>Cualquier representación del mismo, en cualquiera de sus variedades, se enmarca en unos esquemas predefinidos de forma natural por la tradición oral y popular. Dicha transmisión se remonta siglos atrás y el origen de algunas de sus variantes es del todo incierto.</i></p> <p><i>Esto ha provocado que se hayan creado diferentes formas de interpretar este arte tan arraigado. Estas vienen determinadas por aspectos como la rítmica, la métrica, el origen o la armonía, entre otras muchas. La particularidad de cada una de estas variaciones viene conformada por la combinación de las características que se acaban de enumerar creando el concepto de palo flamenco.</i></p> <p><i>La clasificación de los palos tiene diferentes raíces que derivan en palos menores, o en variantes de los mismos atendiendo a las características que se les atribuyen.</i></p> <p><i>Con este trabajo se pretende realizar una ontología que permita el acceso a la descripción de los mismos y el reconocimiento de una obra en función de las propiedades del mismo.</i></p> <p><i>El conjunto de datos que se tomará como punto de partida es el desarrollado por un estudio previo de la materia llamado Flabase. Se trata de un proyecto llevado a cabo por Sergio Oramas para la universidad Pompeu i Fabra en el año 2015 enmarcado en el ya desaparecido Cofla (Análisis Computational de la Música Flamenca, Proyectos de Excelencia de la Junta de Andalucía, P09-TIC-4840, Junta de Andalucía (Consejería de Innovación, Ciencia y Empresas).</i></p>	

<https://www.upf.edu/web/mtg/flabase>

Se pretende que toda esta información se pueda interrogar a través de una aplicación, Wiki o portal Web.

Abstract:

Flamenco represents a form of artistic expression that unfolds in various realms such as singing, guitar playing, or dancing. Any representation of it, in any of its varieties, is framed within predefined patterns naturally established by oral and folk tradition. Such transmission dates back centuries, and the origins of some of its variants are entirely uncertain.

This has led to the creation of different ways of interpreting this deeply rooted art. These interpretations are determined by aspects such as rhythm, meter, origin, or harmony, among many others. The peculiarity of each of these variations is shaped by the combination of the aforementioned characteristics, thus forming the concept of flamenco style or "palo flamenco."

The classification of flamenco styles has different roots leading to minor styles or variations thereof based on the attributed characteristics.

This project aims to create an ontology allowing access to the description of these styles and the recognition of a piece based on its properties.

The dataset serving as a starting point is one developed by a previous study in the field called Flabase. It is a project conducted by Sergio Oramas for Pompeu i Fabra University in 2015, part of the now-defunct Cofla (Computational Analysis of Flamenco Music, Excellence Projects of the Andalusian Government, P09-TIC-4840, Andalusian Government (Ministry of Innovation, Science, and Enterprises)).

<https://www.upf.edu/web/mtg/flabase>

It is intended that all this information be accessible through an application, Wiki or Website.

Agradecimientos

A todas aquellas personas que me han acompañado en estos años de estudio y que me han mostrado todo su afecto y apoyo.

Abstract

El Flamenco significa una forma de expresión artística que se desarrolla en diferentes ámbitos como lo son el cante, el toque o el baile.

Cualquier representación del mismo, en cualquiera de sus variedades, se enmarca en unos esquemas predefinidos de forma natural por la tradición oral y popular. Dicha transmisión se remonta siglos atrás y el origen de algunas de sus variantes es del todo incierto.

Esto ha provocado que se hayan creado diferentes formas de interpretar este arte tan arraigado. Estas vienen determinadas por aspectos como la rítmica, la métrica, el origen o la armonía, entre otras muchas. La particularidad de cada una de estas variaciones viene conformada por la combinación de las características que se acaban de enumerar creando el concepto de palo flamenco.

La clasificación de los palos tiene diferentes raíces que derivan en palos menores, o en variantes de los mismos atendiendo a las características que se les atribuyan.

Con este trabajo se pretende realizar una ontología que permita el acceso a la descripción de los mismos y el reconocimiento de una obra en función de las propiedades del mismo.

El conjunto de datos que se tomará como punto de partida es el desarrollado por un estudio previo de la materia llamado Flabase. Se trata de un proyecto llevado a cabo por Sergio Oramas para la universidad Pompeu i Fabra en el año 2015 enmarcado en el ya desaparecido Cofla (Análisis Computacional de la Música Flamenca, Proyectos de Excelencia de la Junta de Andalucía, P09-TIC-4840, Junta de Andalucía (Consejería de Innovación, Ciencia y Empresas).

<https://www.upf.edu/web/mtg/flabase>

Se pretende que toda esta información se pueda consultar a través de una aplicación.

La estructura del proyecto se repartirá de la siguiente manera:

- Una aproximación a los objetivos que se desean alcanzar
- El estudio de los conceptos teóricos con respecto a la web semántica y las ontologías
- Poner en práctica todo lo anteriormente desarrollado.

English:

Flamenco represents a form of artistic expression that unfolds in various realms such as singing, guitar playing, or dancing. Any representation of it, in any of its varieties, is framed within predefined patterns naturally established by oral and folk tradition. Such transmission dates back centuries, and the origins of some of its variants are entirely uncertain.

This has led to the creation of different ways of interpreting this deeply rooted art. These interpretations are determined by aspects such as rhythm, meter, origin, or harmony, among many others. The peculiarity of each of these variations is shaped by the combination of the aforementioned characteristics, thus forming the concept of flamenco style or "palo flamenco."

The classification of flamenco styles has different roots leading to minor styles or variations thereof based on the attributed characteristics.

This project aims to create an ontology allowing access to the description of these styles and the recognition of a piece based on its properties.

The dataset serving as a starting point is one developed by a previous study in the field called Flabase. It is a project conducted by Sergio Oramas for Pompeu i Fabra University in 2015, part of the now-defunct Cofla (Computational Analysis of Flamenco Music, Excellence Projects of the Andalusian Government, P09-TIC-4840, Andalusian Government (Ministry of Innovation, Science, and Enterprises)).

<https://www.upf.edu/web/mtg/flabase>

It is intended that all this information be accessible through an application. The project structure will be divided as follows:

- An outline of the objectives to be achieved
- Study of theoretical concepts regarding the semantic web and ontologies
- Implementation of all previously developed concepts

Palabras clave

Web Semántica, Ontologías, Palos Flamencos, Flamenco, Andalucía, España, Arte, Folclore, Tradición, Memoria, Trabajo de Final de Grado/Máster.

Índice

1. Introducción.....	13
1.1. Contexto/Justificación	13
1.1.1. Motivación personal	13
1.2. Descripción/Definición	13
1.3. Objetivos generales	14
1.4. Metodología y proceso de trabajo.....	15
1.5. Planificación.....	15
2. Análisis de mercado	17
2.1. Web Semántica	17
2.2. Ontologías	18
2.3. Tecnologías y estándares	20
2.3.1. Conjunto de caracteres: UNICODE.....	21
2.3.2. Lenguaje de marcado extendido: XML	21
2.3.3. Esquema de XML: XML Schema.....	22
2.3.4. Identificadores: URI (IRI, URL, URN).....	22
2.3.5. Intercambio de datos: RDF	23
2.3.6. Taxonomías: RDF-Schema	24
2.3.7. Consultas: SPARQL.....	24
2.3.8. SQWRL/SWRL	25
2.3.9. Ontologías: OWL 2.....	26
2.4. Herramientas para la creación y tratamiento de ontologías.....	27
2.5. Proyectos relacionados	27
3. Propuesta.....	29
4. Diseño.....	30
4.1. Metodología.....	30
4.2. Dominio y alcance de la ontología	30
4.3. Reutilización de ontologías existentes.....	31
4.4. Análisis de los datos.....	31

4.4.1.	Origen de datos	31
4.4.2.	Estructura original de los datos	32
4.4.3.	Conversión y tratado de los datos obtenidos	36
4.4.4.	Unión de los archivos JSON.....	36
4.4.5.	Conversión final de los datos.....	43
4.5.	Diseño de la ontología	44
4.5.1.	Software utilizado	44
4.5.2.	Identificación de los conceptos	44
4.5.3.	Propiedades de objeto de las clases.....	47
4.5.4.	Restricciones.....	48
4.5.5.	Propiedades de datos de las clases.....	49
4.5.6.	Instancias.....	50
4.5.7.	Diseño final del sistema.....	56
5.	Implementación.....	57
5.1.	Instrucciones de instalación.....	57
5.2.	Consultas y reglas de inferencia.....	59
5.2.1.	Consultas y reglas de inferencia con SWRL y SQWRL.....	59
5.2.2.	Consultas SPARQL.....	66
6.	Conclusiones y líneas de futuro	68
6.1.	Líneas de futuro.....	68
	Bibliografía	70
	Anexo A Glosario	71

Figuras y tablas

Índice de figuras

Figura 1. Diagrama de Gantt de la PEC1.....	16
Figura 2. Diagrama de Gantt de la PEC2.....	16
Figura 3. Diagrama de Gantt de la PEC3.....	16
Figura 4. Diagrama de Gantt de la PEC4.....	16
Figura 5. Diagrama de Gantt de la PEC5.....	16
Figura 6. Esquema Web semántica	18
Figura 7. Esquema niveles ontología	20
Figura 8. Esquema Web semántica y Ontologías	21
Figura 9. Diagrama sintaxis URI	22
Figura 10. Esquema URI.....	23
Figura 11. Triplete RDF.....	23
Figura 12. Ejemplo estructura consulta SPARQL	25
Figura 13. Esquema OWL 2.....	26
Figura 14. Niveles De OWL.	27
Figura 15. Estructura JSON Artists	33
Figura 16. Estructura JSON Albums	34
Figura 17 Estructura JSON Palos	35
Figura 18. Estructura JSON Places	35
Figura 19. Código Python Extraer Artists	37
Figura 20. CSV datos extraídos Artists	37
Figura 21. CSV datos extraídos Provincias.....	38
Figura 22. Código Python Extraer Palos	39
Figura 23. CSV datos extraídos Palos	39
Figura 24. Código Python Extraer Albums	40
Figura 25. Nueva estructura JSON Albums	41
Figura 26. CSV datos extraídos Albums	42
Figura 27. Excel final Canciones.....	43
Figura 28. Artistas_excel.xlsx.....	43
Figura 29. Canciones_excel.xlsx.....	43
Figura 30.Palos_excel.xlsx	44
Figura 31. Provincias_excel.xlsx	44
Figura 32. Primera versión arquitectura de clases de ontología	45
Figura 33. 2ª versión Arquitectura de clases de la ontología.	46
Figura 34. Ejemplo Clase Enum TipoArtista	47
Figura 35. Esquema Ontograf del modelo de datos.....	47
Figura 36. Propiedades de Objeto	48
Figura 37. Lista propiedades de datos.....	49
Figura 38. Cellfie inserción Albums.....	51

Figura 39. Cellfie Inserción Palos.....	52
Figura 40. Cellfie Inserción Artistas.....	53
Figura 41. Instancias/individuales de los albums	54
Figura 42. Instancias/individuales de los albums	54
Figura 43. Instancias/individuales de los canciones.....	55
Figura 44. Instancias/individuales de los palos	55
Figura 45. Instancias/individuales de los provincias.....	56
Figura 46. Diseño final del sistema	56
Figura 47. Descarga/Instalación de Protégé.	57
Figura 48. Instalación de plugins Protégé	58
Figura 49. Pestañas Protégé	59
Figura 50. Cuestiones SQWRL.....	60
Figura 51. SQWRL Artistas por palo	60
Figura 52. SQWRL Resultado Artistas por palo	61
Figura 53. SQWRL Artistas por provincia	61
Figura 54. SQWRL Resultado Artistas por provincia	61
Figura 55. SQWRL Canciones por provincia	62
Figura 56. SQWRL Resultado Canciones por provincia	62
Figura 57. SQWRL Mujeres flamencas	63
Figura 58. SQWRL Resultado Mujeres flamencas.....	63
Figura 59. SQWRL Palos por compás	63
Figura 60. SQWRL Resultado Palos por compás	64
Figura 61. Regla SWRL EsDeLaFamilia	64
Figura 62. Propiedad esDeLaFamilia.....	65
Figura 63. Regla SWRL esDelRitmo.....	65
Figura 64. Propiedad esDelRitmo	66
Figura 65. SPARQL Palos por compases.	66
Figura 66. SPARQL Resultado Palos por compases.	67
Figura 67. SPARQL Canciones por artistas.....	67
Figura 68. SPARQL Resultado Canciones por artistas.....	67

Índice de tablas

Tabla 1. I Versión modelado de clases	45
Tabla 2. II Version Modelado de clases	46
Tabla 3. Tabla Propiedades de Objeto.....	48
Tabla 4. Tabla de restricciones de propiedades.....	49
Tabla 5. Propiedades de datos	49
Tabla 6. Propiedades Datos Inferidas	49

1. Introducción

1.1. Contexto/Justificación

Acercarse al Flamenco ya sea como aficionado, oyente o como estudioso del mismo requiere la capacidad de poder distinguir perfectamente qué está sucediendo en la canción. Saber interpretar sus métricas, ritmos o contextualizarla en un momento y lugar adecuados es una tarea que requiere de un conocimiento previo y profundo. Adquirirlo implica cierto grado de complejidad al que se ha de enfrentar quien quiera adentrarse en el estudio de esta música.

Por ello, se ha considerado necesario poder trasladar este conocimiento al ámbito de la web semántica y de este modo poder contextualizar todos estos términos en el mundo digital.

Para llevarlo a cabo se ha optado por la creación de una ontología que facilite el aprendizaje y las posibles consultas que puedan surgir en un momento dado.

Con este trabajo se pretende posibilitar el acceso a la información de esta materia a través de una aplicación enfocada a la docencia o al mero acercamiento a la base musical. Con ello se pretende reducir la complejidad que conlleva.

1.1.1. Motivación personal

El gusto por esta música ha hecho que me haya interesado por la misma sin dejar de sentir una frustración enorme por la dificultad que conlleva su aprendizaje.

Por ello, las principales causas que han provocado el desarrollo de este proyecto podrían resumirse:

- Facilitar la tarea de aprendizaje del Flamenco.
- Contribuir a su difusión a través del mundo digital y de la Web semántica

1.2. Descripción/Definición

Para las personas ajenas al mundo del flamenco es muy probable que el mismo es pueda resultar un tanto monótono o con una cierta uniformidad en todas sus expresiones. Cuando el oyente va mostrando interés y afinando un poco el oído es cuando empieza a distinguir una serie de matices y de patrones que determinan diferentes formas de expresión que son las que hacen de esta manifestación artística algo tan único y especial.

Son estos matices los que suelen resultar más complicados de comprender y de localizar en según qué canciones o bailes. Por ello, y de forma natural y popular, se han ido constituyendo una serie de estilos o palos que clasifican estas virtudes de una forma un tanto anárquica.

Dentro de ese supuesto desorden, los palos flamencos se terminan definiendo por una serie de características muy claras que en la mayor parte de las ocasiones se hacen patentes. El ritmo, el origen, el estado de ánimo, la finalidad de la composición, etc.. son algunas de ellas.

El objetivo de crear una web semántica que permita concretar en el mundo digital todas estas cualidades es un tanto ambicioso debido a su origen popular y natural. Pero tras un estudio de una serie de patrones se considera más que posible poder llevarlo a cabo.

El propósito de la web semántica es proporcionar respuestas precisas a las consultas formuladas. Por ejemplo, si buscamos información referente a los palos flamencos se mostrarán muchas webs en las que se recoja información sobre el tema tal y como haría un buscador convencional.

Hoy en día, la web semántica está ganando importancia al establecer un estándar para el intercambio de información entre sistemas completamente heterogéneos. Esto se basa en el uso de ontologías que permiten representar un modelo conceptual para la representación, almacenamiento y, especialmente, el intercambio de datos. Su potencial radica en la capacidad de incorporar datos de diversas fuentes para mejorar la calidad y precisión de las respuestas.

Si se navega por internet, hay repositorios que proporcionan información sobre temáticas referentes al mundo de la música. Tan sólo se ha dado con un repositorio (FLABASE) con el que poder trabajar para extraer las estructuras de datos necesarias. Estas se utilizarán para crear una taxonomía que formará el núcleo de la ontología final. Se elegirá el programa Protégé para finalmente generar un archivo con la ontología en el lenguaje OWL.

Una vez llegados a este punto, se procederá a cargar los datos de estos repositorios, pero para realizar este trabajo tendremos que utilizar alguna herramienta, API o script, que permita poblar la ontología. Finalmente, se interrogará la ontología mediante los lenguajes SQWRL, SPARQL y crear reglas de inferencia con SWRL para obtener respuestas concretas.

Con este sistema se podrá utilizar esta ontología tipo para poder implementarla en algún sistema, aplicación, portal web o wiki.

1.3. Objetivos generales

Este trabajo pretende establecer una solución que facilite y proporcione el suficiente conocimiento del concepto de los palos flamencos.

Se pretende crear una ontología que pueda servir de punto de partida para futuras ampliaciones y que a su vez sirva para poder aplicarla para diferentes usos didácticos.

Para ello, se desarrollará un modelo ontológico personalizado que permita poder realizar consultas e inferir diferentes razonamientos. El objetivo perseguido en este proyecto es diseñar, crear, implementar y poner en marcha una web semántica con la finalidad de poder facilitar el acceso al conocimiento de la materia ya sea a través de su implementación en un portal, en una wiki, o aplicación.

Dicha herramienta es necesaria para el indexado de los contenidos agilizando de ese modo la búsqueda y que permita el acceso a todos estos conceptos.

Ello supondrá un ahorro del tiempo empleado en la búsqueda de información referente a la definición de los estilos flamencos. La conexión de los conceptos entre si colaborará con este cometido.

Apenas hay datos digitalizados con los que poder trabajar del tema en la actualidad. Por ello, este proyecto podría resultar de gran ayuda si se enfoca como herramienta destinada a la docencia.

También servirá como base de conocimiento para futuras plataformas a través de la información ordenada en metadatos y en ontologías. De ese modo, a través de agentes externos, también se podrá interactuar con la información.

1.4. Metodología y proceso de trabajo

Primeramente se llevará a cabo la descripción del estado del arte de la web semántica. En ella se tratará toda la información que versa sobre su filosofía, necesidad y objetivos. También se desarrollarán los posibles objetivos a futuro al igual que los pilares y herramientas en los que se sustenta.

Previo al análisis del negocio, se realizará una búsqueda de posibles referencias que arrojen conocimiento sobre el tema. Ello incluye la búsqueda de ontologías dedicadas al flamenco previamente creadas, repositorios de datos abiertos, etc.. Una vez localizados se procederá a su estudio y, en el caso de que proceda, a la descarga y reutilización de estas fuentes.

A continuación se procederá a determinar qué datos nos resultaran útiles de estos archivos. Con ello se generará un modelo conceptual que estructure la idea del proyecto.

Una vez que el modelado esté creado se determinará con que herramientas se desarrollará la estructura de la taxonomía. Con ellas se desarrollará y creará la ontología.

Se realizará un poblado de la ontología para poder ser utilizada con los datos extraídos de los repositorios de open data.

Por último, se desarrollarán las reglas de inferencia y las consultas con las que poder trabajar y comprobar que todo funciona perfectamente.

1.5. Planificación

El reparto de tareas para la elaboración de la ontología se lleva a cabo a través de diferentes fechas en forma de 5 PECS. Los hitos de cada una de ellas se centran en la búsqueda de una idea para el proyecto, el modelado de los datos, la elaboración de la ontología, la redacción de la memoria y la defensa final.

A continuación se muestra la planificación mediante un diagrama de Gantt haciendo uso de la herramienta online tomsplanner.com.

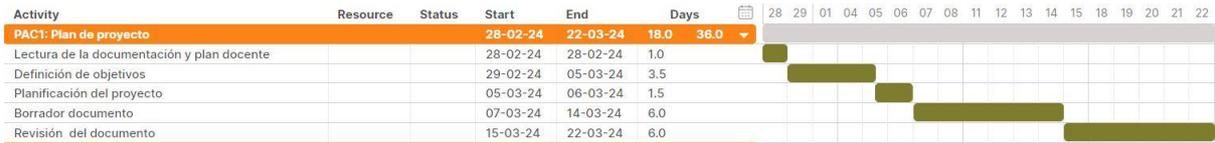


Figura 1. Diagrama de Gantt de la PEC1



Figura 2. Diagrama de Gantt de la PEC2

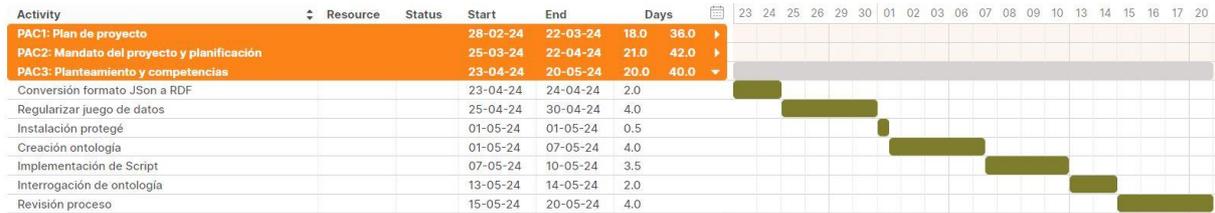


Figura 3. Diagrama de Gantt de la PEC3

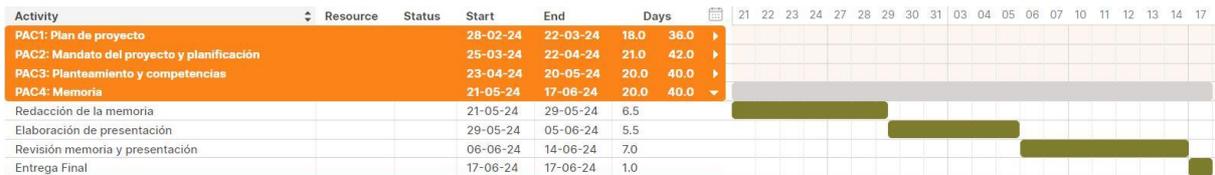


Figura 4. Diagrama de Gantt de la PEC4

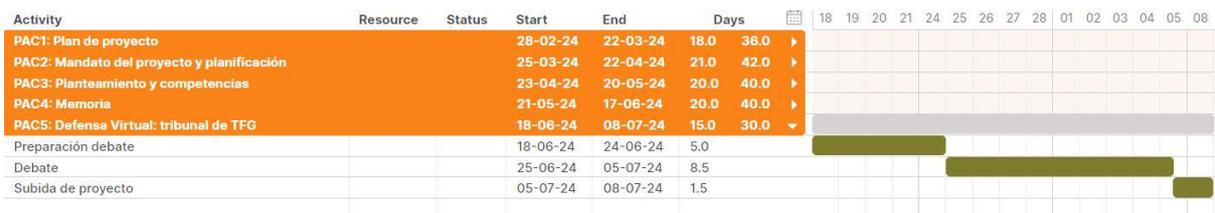


Figura 5. Diagrama de Gantt de la PEC5

2. Análisis de mercado

En este apartado se realizará exposición detallada de los conceptos que forman parte del presente proyecto. Inicialmente se hace referencia a la web semántica y la ontología como elementos principales, seguidamente se presentan las tecnologías y los estándares. A continuación se presenta una descripción de las herramientas utilizadas y finalmente se detallan los proyectos previos relacionados con el tema.

2.1. Web Semántica

La World Wide Web (WWW) ha ido sufriendo una serie de cambios de paradigma reflejados a través de sus diferentes versiones.

La Web 3.0. se presenta como la Web Semántica, siendo acuñado el término a principios del 2000 por el propio creador de la WWW y presidente del consorcio W3C, Tim Berners-Lee, quien planteará la necesidad de una evolución de la red.

“Mi sueño es una Web en la que las máquinas sean capaces de analizar todos los datos/contenido, enlaces y transacciones entre la gente y los ordenadores-. La ‘Web Semántica’ que lo haría posible es para llegar todavía, pero cuando llegue, la rutina de nuestras compras, burocracia y vida diaria será gestionada por máquinas comunicándose con máquinas. Los ‘agentes inteligentes’ que han sido anunciados durante décadas, se convertirán por fin en una realidad.”

(Berners-Lee et al, 2001)

Hasta el momento la información que se mostraba en la misma venía determinada por el uso de los documentos HTML que la trataban en forma de texto plano o elementos multimedia.

"El desafío de la Web Semántica, por tanto, es proporcionar un lenguaje que exprese tanto datos y reglas para razonar sobre los datos y que permite exportar a la Web reglas de cualquier sistema de representación de conocimiento existente." (b1)

Esta forma de presentar los contenidos resultaban de gran utilidad para los usuarios. Sin embargo, no lo sucede lo mismo con los sistemas que suelen trabajar con la información ya estructurada y gestionada a través de Bases de Datos.

Con la aparición de la Web semántica se dota de significación y se consigue contextualizar los datos que pueblan la red, resultando útil ya no sólo para el ser humano que interactúa sino también para las propias máquinas (aplicaciones informáticas). (b2)

Esto se consigue creando relaciones entre los conceptos a través de etiquetas semánticas. Para ello se basa en los estándares de las propiedades características del lenguaje natural.

Para solucionarlo se crean las ontologías que suponen una categorización de las palabras. La misma define la relación entre las mismas dependiendo del dominio al que pertenezcan. De hecho, también reciben la denominación "modelos de dominio".

El concepto de Ontología tiene su origen en la Filosofía (trata del ser en general y de sus propiedades trascendentales), adoptado por la Informática.

Por ello, en estos términos, se podría definir como la representación de la conceptualización explícita compartida de un dominio en particular (**b3**).

Para poder entender esta definición habrá que tener en cuenta estas tres características fundamentales:

- **Representación explícita:** Se encuentra en lenguaje formal y en soporte digital que permita que pueda ser interpretado por cualquier sistema informático.
- **Conceptualización compartida:** representación de la información que comparten un conjunto de individuos respecto a un dominio en particular.
- **Dominio Particular:** Representa el ámbito requerido para un problema preciso.

Según la teoría al respecto de T. R. Gruber (1993) el conocimiento en las ontologías se formaliza principalmente utilizando cinco tipos de componentes: relaciones, clases, funciones, axiomas e instancias.

- **Relaciones:** Es la representación de las conexiones entre los diferentes conceptos que se manejan en el dominio.
- **Clases:** La idea que tenemos de lo que se quiere representar. Se sustentan en taxonomías y pueden ser de objetos, tareas, métodos, etc..
- **Funciones:** Relaciones en las que se calcula un elemento dependiendo de otros elementos del dominio.
- **Axiomas:** Expresiones evidentes. Siempre son ciertas. Su uso es aplicable a la definición de restricciones como del propio significado del elemento.
- **Instancias:** la representación de los objetos de una clase o concepto

En cuanto a la clasificación de las ontologías Se pueden encontrar diferentes tipos. Dos criterios que se podrían aplicar a la categorización podrían ser según el tipo de conocimiento o por la motivación.

En el segundo caso, propuesto por Mizoguchi en 1995, se pueden apreciar las siguientes:

- **Ontologías generales o de alto nivel:** Utilizada para conceptos genéricos como el tiempo, el espacio, etc..
- **Ontologías del dominio:** Utilizan el lenguaje propio perteneciente a un dominio concreto.
- **Ontologías de tarea:** Utilizan un vocabulario referente a tareas o actividades más genéricas.
- **Ontologías de aplicación:** De un ámbito concreto. Diseñadas para una aplicación concreta.

En el caso a tratar se plantea la ontología como Ontología de Dominio aunque interactúe reutilizando ontologías generales y de tarea.

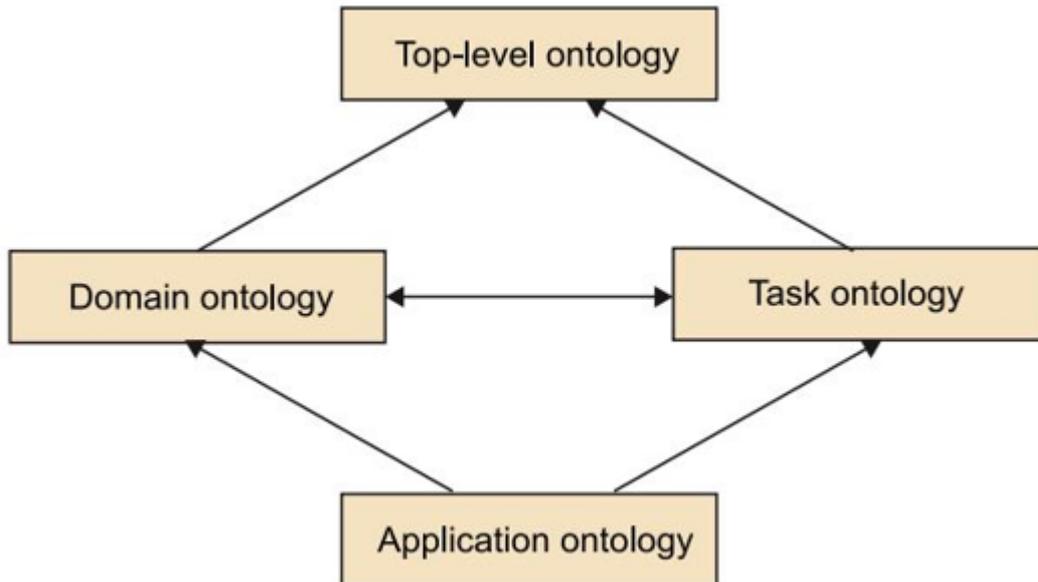


Figura 7. Esquema niveles ontología

Las ontologías, por tanto, podrían considerarse como un conocimiento compartido y común de algún dominio al que pueden acceder diferentes individuos desde cualquier situación geográfica.

A día de hoy se investigan en diferentes ámbitos de la IA como Procesamiento del Lenguaje Natural, Ingeniería del Conocimiento, Representación del Conocimiento, Integración Inteligente de Información, Recuperación de la Información, Sistemas Cooperativos de Información o el Comercio Electrónico.

2.3. Tecnologías y estándares

Se encuentra un conjunto de tecnologías estándar para la gestión de la información en la web semántica propuestas por la W3C.

La figura 3 representa la arquitectura en la que se estructura la WS.

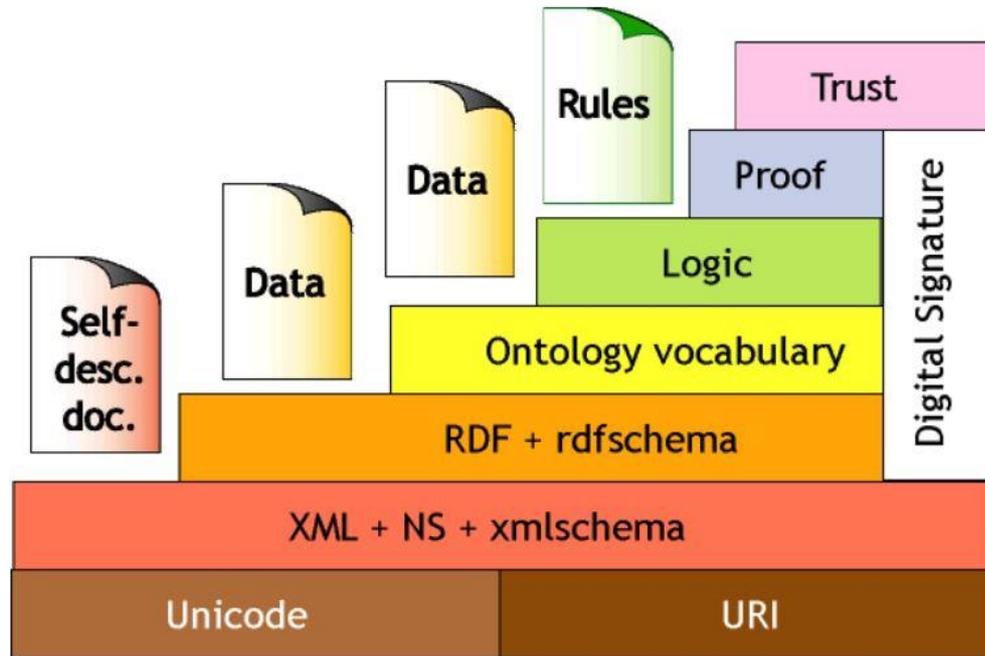


Figura 8. Esquema Web semántica y Ontologías

En las capas superiores se aprecian las aplicaciones semánticas, a continuación se muestran los sistemas razonadores o de inferencia que gestionan la información semántica a nivel lógico. En la siguiente se encuentran las descripciones aplicando las ontologías y las reglas y la inferior muestra la información a nivel sintáctico.

2.3.1. Conjunto de caracteres: UNICODE

Se trata de un estándar de codificación de caracteres. UNICODE posee tres formatos para la codificación de los mismos (UTF-8, UTF-16 y UTF-32) en función del número de bits utilizados. Lenguajes aplicados al mundo de la informática y el mundo web como el lenguaje de marcado XML hacen uso de este estándar. Siempre se especifica en su cabecera el formato de transformación (UTF).

Ejemplo:

```
<?xml version="1.0" encoding="UTF-8"?>
```

2.3.2. Lenguaje de marcado extendido: XML

XML es un lenguaje de marcado extensible, al igual que el HTML es un lenguaje de marcado de hipertexto en el que se basa la Web Semántica.

Está enfocado en la transferencia de datos y está basado en el contenido. Ello facilita la comunicación entre aplicaciones compartiendo el conocimiento de forma estructurada. Resulta legible tanto para el humano como para la máquina.

Permite a los usuarios definir sus propias etiquetas para poder darle forma al conocimiento. El etiquetado sigue un conjunto de reglas que conforman lo que se conoce como esquemas.

Ejemplo XML:

```
<?xml version="1.0">
  <contact>
    <name>Raúl López</name>
    <email>rlopezvic@uoc</email>
    <telephone>657623676</telephone>
  </contact>
```

2.3.3. Esquema de XML: XML Schema

Tal y como se comentaba en el punto anterior referente al XML el XML Schema se basa en una serie de reglas y restricciones a la hora de componer la estructura del contenido. Si las normas se cumplen satisfactoriamente, el documento se considerará como válido.

2.3.4. Identificadores: URI (IRI, URL, URN)

Identificador de Recursos Unificado. Se trata de un identificador único para cada uno de los recursos que pueblan la red (emails, página Web, libro, etc.), ya sea abstracto o físico.

Utilizan el UNICODE como estándar de codificación de sus caracteres.

Además del URI podemos encontrar otros identificadores:

- **IRI:** Identificador de Recursos Internacionalizado. Similar a los URI pero permite utilizar caracteres de otras lenguas distintas a la inglesa
- **URL:** Localizador Uniforme de Recursos. Como su nombre indica, hace referencia a la localización del recurso. Pertenece al conjunto de las URI.
- **URN:** Nombre Uniforme de Recurso. Permite identificar recursos pero sin necesidad de estén disponibles.

Un ejemplo del esquema de URI podría ser el siguiente:

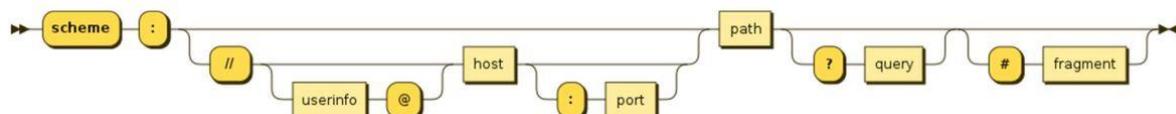


Figura 9. Diagrama sintaxis URI

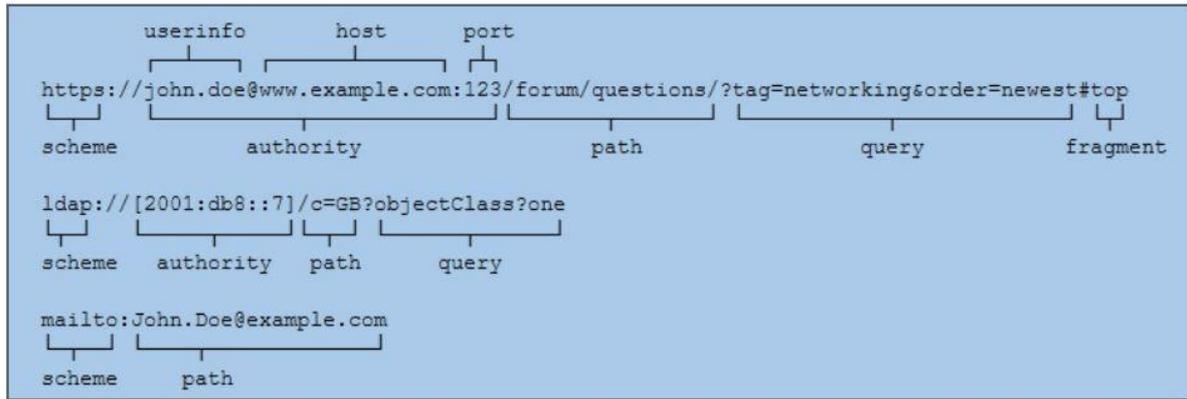


Figura 10. Esquema URI

2.3.5. Intercambio de datos: RDF

Marco de Descripción de Recursos es un estándar utilizado para la descripción de metadatos. Se trata de una creación del W3C y sirve para detallar los recursos de la red.

Define los modelos de datos en forma de grafo dirigido expresando la información en forma de tripletes. Estas estructuras se basan en la expresión del tipo sujeto-predicado-objeto.

- **Sujeto:** Origen. Recurso.
- **Predicado:** Arista. Propiedad o relación del recurso.
- **Objeto:** Nodo destino. El valor.

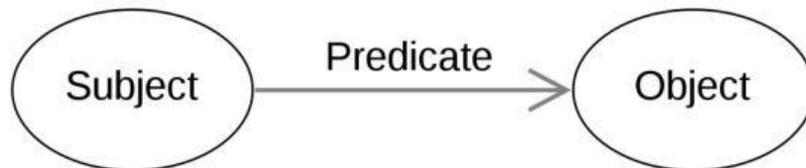


Figura 11. Triplete RDF

La representación de los nodos en el grafo se hacen a través de URIs que sirvan como identificación de una entidad concreta, un nodo en blanco o un literal. También se llevará a cabo la representación de las aristas por medio de URIs.

Esto es así para el establecimiento de la relación entre nodos en forma de par atributo/valor.

A diferencia del XML que funciona a nivel sintáctico, el RDF lo hace a nivel semántico.

Para representar la serialización de los RDF se pueden encontrar diferentes estándares como son RDF/XML, N-Triples, Turtle, TriG, N-Quads o JSON-LD.

2.3.6. Taxonomías: RDF-Schema

El Esquema RDF es una extensión del RDF que amplía sus capacidades proporcionando los elementos necesarios para describir ontologías. Es decir clases, propiedades e instancias y restricciones.

RDF-S permite establecer jerarquías entre las clases y las subclases con tres clases básicas como son: *rdfs:Resource*, *rdfs:Property* y *rdfs:Class*.

Para las propiedades se determina la jerarquía especificando el tipo de clase mediante *rdf:type*, *rdfs:subClassOf* y *rdfs:subPropertyOf*, quedando para determinar las restricciones *rdfs:domain* y *rdfs:range*.

Classes	Properties
<i>rdfs:Resource</i>	<i>rdfs:range</i>
<i>rdfs:Class</i>	<i>rdfs:domain</i>
<i>rdfs:Literal</i>	<i>rdfs:type</i>
<i>rdfs:Datatype</i>	<i>rdfs:subClassOf</i>
<i>rdfs:langString</i>	<i>rdfs:subPropertyOf</i>
<i>rdf:HTML</i>	<i>rdfs:label</i>
<i>rdf:XMLLiteral</i>	
<i>rdf:Property</i>	

2.3.7. Consultas: SPARQL

Se trata del lenguaje empleado para realizar consultas a los tripletes de RDF y RDF-S. Para ello incluye un subconjunto de referencias URI y optimiza el espacio para adquirir más velocidad.

Puede realizar 4 tipos de consultas:

- **SELECT:** búsqueda de expresión
- **CONSTRUCT:** Crea tripletes RDF en función de datos recuperados previamente.
- **ASK:** Verdadero o Falso
- **DESCRIBE:** extrae conjuntos de caracteres de un gráfico.

Incluye funciones modificadoras de consulta como **DISTINCT**, **ORDER BY** o **LIMIT**.

Ejemplo de la estructura de una consulta SPARQL:

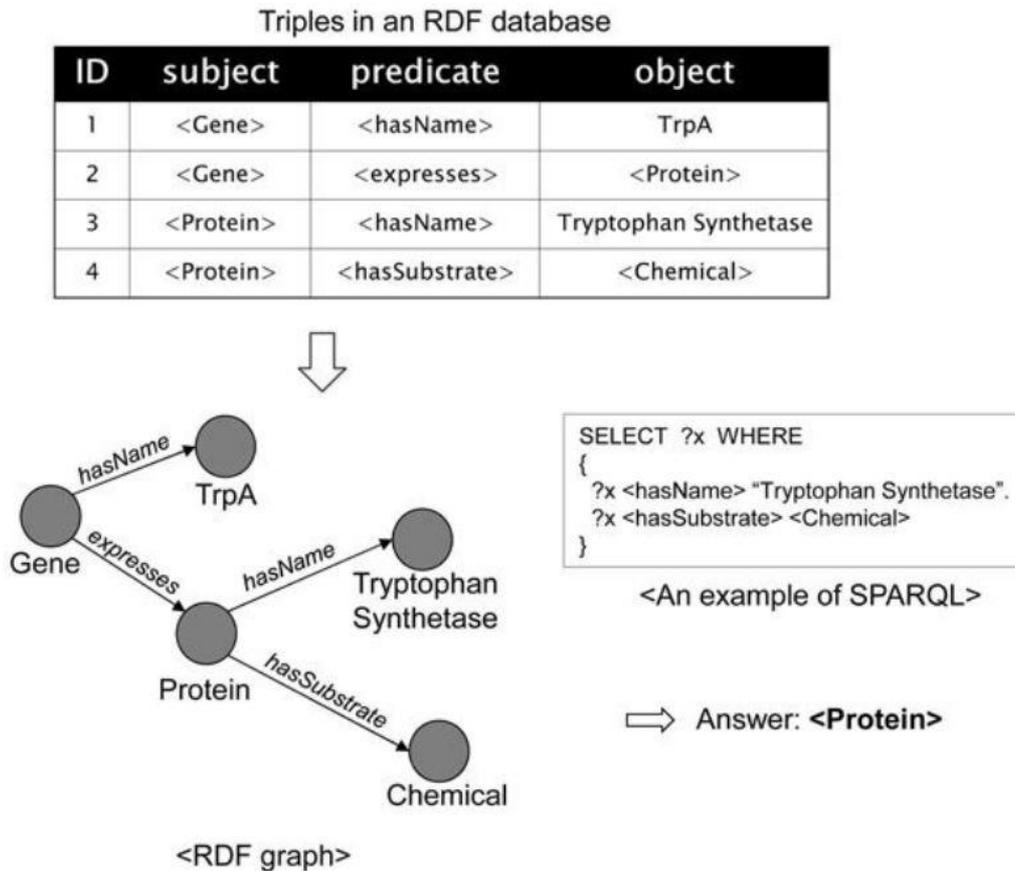


Figura 12. Ejemplo estructura consulta SPARQL

2.3.8. SQWRL/SWRL

Las ontologías requieren de una serie de reglas lógicas que permitan gestionar correctamente cualquier dominio. Para poder definir las correctamente se aplica el lenguaje SWRL (Semantic Web Rule Language) que posibilita la creación de reglas de lógica de primer orden, permitiendo realizar inferencias complejas y expresivas. Se componen de antecedentes (si...) y consecuentes (entonces...), y se aplican para derivar nuevos hechos a partir de los hechos existentes en una ontología.

Ejemplo SWRL:

Person(?p) ^ hasParent(?p, ?parent) -> Child(?p)

A su vez se pueden realizar consultas de forma similar a como se realizan a través de SQL o incluso SPARQL a través de SQWRL. Ésta es una extensión de SWRL que es utilizada para recuperar información y obtener resultados específicos basados en los datos almacenados en una ontología.

Ejemplo SQWRL:

Person(?p) ^ hasAge(?p, ?age) ^ swrlb:greaterThan(?age, 18) -> sqwrl:select(?p)

2.3.9. Ontologías: OWL 2

OWL 2 es la versión actual de OWL. Se trata de un lenguaje de representación del conocimiento que permite desarrollar ontologías: clases, propiedades, tipos, valores de datos que son utilizados en la WS.

Estas ontologías de facilitan un modelo de marcado basado en RDF y codificado en XML. A grandes rasgos, se podría afirmar que es una extensión de RDFS diseñada para representar recursos y sus relaciones proporcionando semántica adicional a los datos.

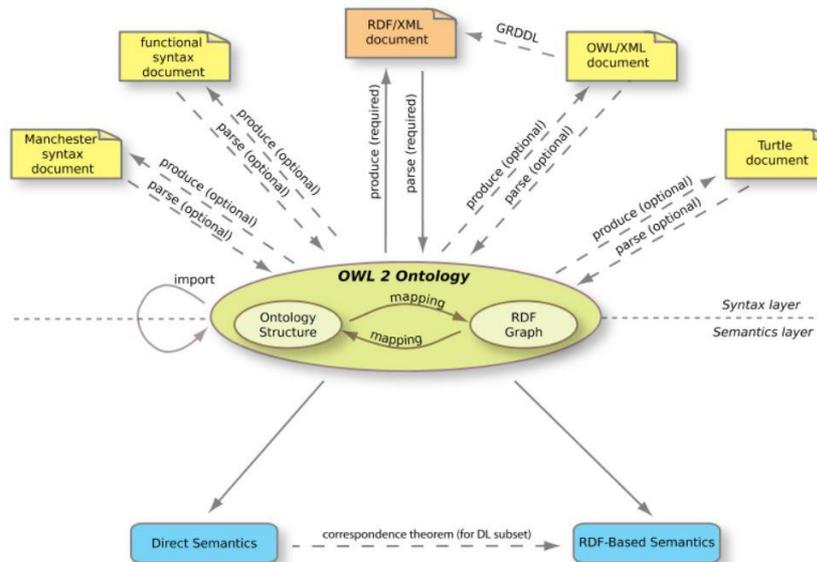


Figura 13. Esquema OWL 2

Hay tres niveles principales dentro de OWL 2:

- **OWL Lite:** Restricciones enfocadas al establecimiento de jerarquías.
- **OWI DL:** Máxima expresividad. Hay garantía de resultados en un tiempo finito.
- **OWL Full:** Expresividad total. No hay garantía de resultados en un tiempo finito.

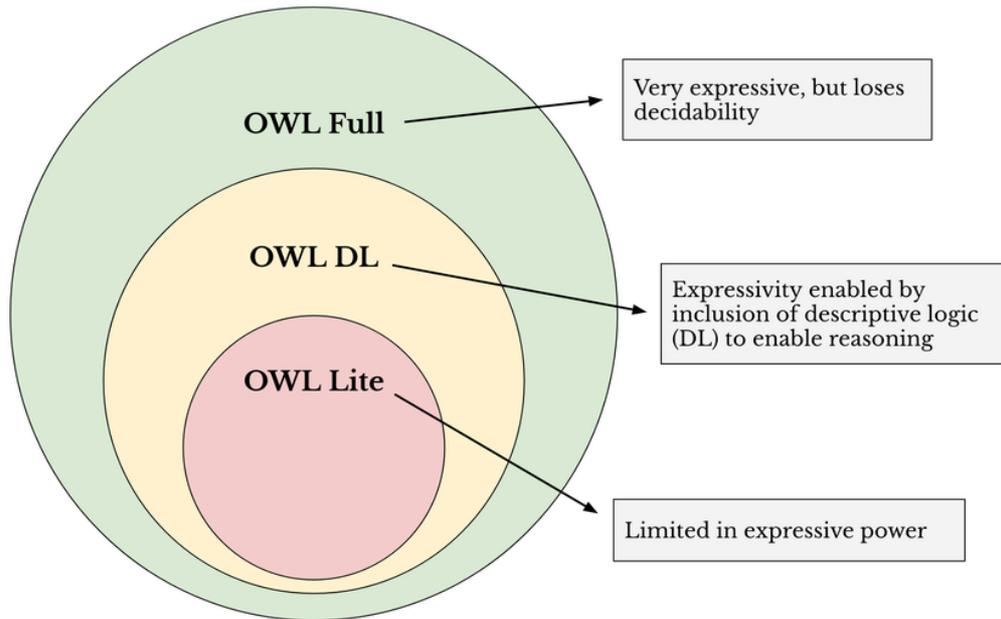


Figura 14. Niveles De OWL.

Algunas de las sintaxis que puede adoptar OWL son OWL2, OWL2/XML, RDF/XML o la Sintaxis Manchester.

2.4. Herramientas para la creación y tratamiento de ontologías

Del mismo modo que la forma de abordar las ontologías es muy variada, también lo son las herramientas con las que se pueden desarrollar.

Estas herramientas son utilizadas para definir la estructura con la que clasificar la información realizada por medio de anotaciones. Todas ellas se podrían clasificar como herramientas de desarrollo de ontologías, de fusión e integración de ontologías, de evaluación de ontologías, basadas en la anotación, de aprendizaje o de almacenamiento y consulta.

Algunos ejemplos de herramientas son LinkFactory, WebOnto, Ontolingua o Protegé. Esta última es la más recomendada y la más utilizada. Esta será la herramienta elegida por desarrollar este proyecto debido a la documentación existente y las recomendaciones de la misma.

2.5. Proyectos relacionados

En la web apenas se puede localizar información precisa respecto a la elaboración de ontologías que clasifiquen de una forma tan concreta los palos flamencos. No obstante, la información respecto a los mismos y a sus estructuras sí que es relativamente accesible en la red.

El principal proyecto que se ha hecho eco de una clasificación como la que se desea desarrollar sería [COFLA II](#): análisis Computacional de la música Flamenca inmerso en el marco de las Ayudas a la Investigación de la Agencia Andaluza para el Desarrollo del Flamenco.

Dicho trabajo culminó con el resultado de un Website que mostraba los resultados del mismo.

Lamentablemente no se encuentra operativa en la actualidad aunque sí que se sigue manteniendo accesible el informe final.

Convive junto a COFLA el proyecto [FLABASE](#) que se ha dedicado a reunir las diferentes bases de conocimiento general (Wikipedia, DBPedia..), enciclopedias musicales (Musicbrainz) y sitios web especializados en flamenco.

Dicho trabajo de investigación ha sido llevado a cabo por Sergio Oramas para la Universidad Pompeu Fabra.

Será esta la principal fuente de datos con la que trabajar debido a que consta de una serie de Datasets que contienen una parte esencial del conocimiento con el que se pretende trabajar en este proyecto.

FLABASE contiene información sobre 1.174 artistas, 76 palos (géneros flamencos), 2.913 álbumes, 14.078 temas y 771 localizaciones andaluzas.

Está disponible gratuitamente en formato JSON. La precisión de los resultados muestra un alto grado de calidad e integridad de la base de conocimientos.

En cuanto a la clasificación de los palos flamencos se accede a diferentes fuentes de conocimiento, como lo son las siguientes publicaciones en formato libro:

Pulpitarrita con pasaporte flamenco. Silvia Marin. Editorial FlamencoVive. **(b4)**

En el capítulo "La mar de estilos flamencos" aparecen los diferentes estilos (los de arriba, los alegres (festeros), los de abajo, los tristes).

En esta parte se reparten los cantes en cuatro familias : fandango, soleá, seguiriya y tangos.

Investigación y Flamenco J.M. Díaz-Báñez, F.J. Escobar-Borrego (Eds.) , Signatura Ediciones, ISBN 978-8496210-90-5, 2011. **(b5)**

Libro publicado por José Manuel Díaz-Báñez, director del proyecto COFLA 2 que aborda los diferentes estilos flamencos desde diferentes perspectivas.

Palos y el estilo Flamenco. Felipe Gertrudix. <https://www.bubok.es/libros/9562/palos-y-estilos-del-flamenco> **(b6)**

Publicación que repasa y clasifica los palos y los estilos flamencos.

Flamencopolis (Web): https://flamenco.plus/flamencopolis/index.php?id_palo=cantinas

Portal Web que aglutina información muy relevante y contrastada del arte flamenco en todas sus variantes. La sección explicativa de los palos resulta determinante para este trabajo. **(b10)**

Además de estas referencias bibliográficas también caben destacar los siguientes proyectos previos que han servido como inspiración para desarrollar este proyecto:

- *Modelo web semántico para contactar grupos musicales en la ciudad de Bogotá (b7)*
Se trata de un estudio y creación de una ontología con la que poder clasificar los géneros musicales tradicionales de la ciudad para poder realizar búsquedas de artistas que coincidan con esa clasificación. El autor del estudio es Juan Esteban Katz Delgado para la Universidad Distrital FJC de Colombia.
<http://portal.amelica.org/ameli/journal/536/5362548005/html/>
- *Performed Music Ontology (b8)* es un proyecto perteneciente a BIBFRAME en el que se establecen una serie de clasificaciones de diferentes aspectos y características de la música interpretada. BIBFRAME es un modelo para expresar y conectar datos bibliográficos creado por la Library of Congress. (<https://www.loc.gov/>).
<http://performedmusicontology.org/>
- *Music Representation Ontology.(b9)* Ontología creada para clasificar las polifonías y perteneciente al proyecto polifonía-project (<https://github.com/polifonia-project>).
<https://github.com/polifonia-project/music-representation-ontology>

3.Propuesta

Tras realizar el análisis de los elementos necesarios para formar parte de un proyecto enfocado a la web semántica, se describirán los elementos generados a continuación:

- Ontología que estructure elementos claves del flamenco para poder trabajar con ella posteriormente.
- Los códigos necesarios para poder añadir elementos a la ontología con los que poder inferir en un futuro.
- Un documento explicativo en el que quede claro cómo realizar consultas, inferencias, etc..
(Esta misma memoria)

4. Diseño

A continuación se llevará a cabo el correspondiente análisis de los datos disponibles. En los siguientes apartados se detallan las bases de una metodología que será utilizada para el diseño de nuestro modelo, al igual que cada una de las etapas del proceso.

4.1. Metodología

La implementación de la ontología cuyo prototipo se describió en los capítulos anteriores ha dado lugar a ligeros cambios de modelización debido a la naturaleza de los datos contenidos en los repositorios con los que se pretende trabajar en este proyecto. En el presente capítulo se explican estos cambios y se detalla el mecanismo de descarga e inclusión de estos datos en la ontología de los palos flamencos.

Para llevarlo a cabo se procederá a seguir los siguientes pasos:

- Concretar el dominio y alcance de la ontología.
- Localizar y reutilizar ontologías existentes.
- Detallar los términos imprescindibles para la ontología.
- Definir las clases y su jerarquía.
- Definir las propiedades de las clases (atributos).
- Crear instancias.

4.2. Dominio y alcance de la ontología

La primera fase a la hora de desarrollar una ontología es determinar cuál será el dominio y el alcance de la misma. En este punto se pretende definir los actores y los elementos que intervienen en la expresión artística del flamenco.

A la hora de establecer cuál es el alcance del mismo se plantean una serie de preguntas.

¿Para qué se utilizará la ontología?

La creación de la misma tiene como objetivo crear una base de conocimiento que pueda ser implementada e incrementada por futuros desarrollos y estudios relacionados con la materia. Es decir, Se podría considerar una primera aproximación a la clasificación semántica (web) del Flamenco.

¿Para qué tipo de preguntas la información en la ontología deberá proveer respuestas?

La tipología de las preguntas que se pueden responder con la misma son bastante diversas:

- ¿Qué tipo de rítmicas tienen los palos flamencos?
- ¿De qué tipo es la canción X: Alegre o Triste?
- ¿Cuál es la ciudad de nacimiento del intérprete de la canción X?

- ¿Podrías darme un listado de cantaoras?
- ¿Qué canciones tienen el compás X?

¿Quién la utilizará en el futuro?

El uso y el mantenimiento serían gestionados por profesionales de las TIC que opten por ampliar o reutilizar la ontología.

4.3. Reutilización de ontologías existentes

Tras una exhaustiva búsqueda en las redes para poder encontrar alguna ontología que contemple alguno de los aspectos y elementos del dominio elegido (Flamenco) se concluye que los resultados no han resultado satisfactorios.

Algunos de los portales en los que se ha realizado dicha prospección:

- <http://swoogle.umbc.edu/>
- <http://www.ksl.stanford.edu/software/ontolingua/>
- <http://www.daml.org/ontologies/>
- <https://www.hakia.com/>

Se han tenido en cuenta ontologías como *Modelo web semántico para contactar grupos musicales en la ciudad de Bogotá* (b7) y *Music Representation Ontology*. (b8) Ambas abordan la temática de la gestión de artistas o la de elementos musicales pero se terminan desestimando debido a que no terminaban de encajar en el modelo planteado inicialmente.

4.4. Análisis de los datos

Los orígenes de datos que constituirán la base de conocimiento a la hora de desarrollar el modelo tienen diferentes orígenes. Para que resulten útiles tanto a la hora de utilizarlos como referencia como para poblar nuestra ontología será necesario realizar una serie de modificaciones.

4.4.1. Origen de datos

Tal y como se ha planteado en el punto anterior los datos obtenidos de los repositorios a los que se ha acudido han resultado un tanto confusos a la hora de poder aplicar un orden con el que poder trabajar y elaborar una ontología adecuada.

Los orígenes de datos principales son los siguientes:

- FLABASE(Web): contiene información sobre 1.174 artistas, 76 palos (géneros flamencos), 2.913 álbumes, 14.078 temas y 771 localizaciones andaluzas.
<https://www.upf.edu/web/mtg/flabase>

- *Pulpitarrita con pasaporte flamenco.* (libro)Silvia Marín. Editorial FlamencoVive. **(b4)**
En el capítulo "La mar de estilos flamencos" aparecen los diferentes estilos (los de arriba, los alegres (festeros), los de abajo, los tristes).
- Flamencopolis.(Web) **(b10)**: Portal que explica en detalle todos y cada uno de los palos flamencos <https://flamenco.plus/flamencopolis/index.php?>

El principal recurso al que se acude es al proyecto FLABASE que muestra una serie de datasets en formato JSON aglutinando información de diferentes fuentes como son:

- Wikipedia,
- DBpedia
- Andalucia.org
- elartedevivirelflamenco.com
- MusicBrainz
- flun.cica.es/index.php/grabaciones/base-datos-grabaciones
- juntadeandalucia.es/institutodeestadisticaycartografia/sima

A la hora de determinar y pensar en el diseño de la ontología se comprueba que los datos obtenidos de dichos ficheros requiere de un limpiado y un tratado de los mismos para poder ser reutilizados y aplicados a la ontología planteada de una forma óptima.

4.4.2. Estructura original de los datos

Tras revisar los datasets disponibles en FLABASE se opta por descargar los siguientes:

- Artists
- Albums_Db
- Palos
- Places

Se descarta la utilización del DataSet Albums_mb . Los ficheros obtenidos se encuentran en formato JSON. Se muestran como archivos individuales que complican la carga posterior en la ontología a través del programa Protégé.

La información contenida en los archivos de los Datasets se muestra con los siguientes campos y estructuras:

Artists:

andaluciaorg_full_name, andaluciaorg_html, andaluciaorg_html_en, andaluciaorg_origin,
andaluciaorg_text, andaluciaorg_text_en, andaluciaorg_title, andaluciaorg_type, andaluciaorg_url,

andaluciaorg_url_en, dbpedia_filtered_type, dbpedia_types, dbpedia_uri, elartede_text, elartede_title, elartede_url, extracted_birth_year, extracted_death_year, extracted_origin, extracted_origin_id, extracted_province, flabase_artist_id, flabase_type, freebase_uri, mb_type, mbid, title, wikidata_uri, wikipedia_categories, wikipedia_html_en, wikipedia_html_es, wikipedia_text_en, wikipedia_text_es, wikipedia_url

```
{
  "elartede_url": "http://elartedevivirelflamenco.com/bailaoras137.html",
  "elartede_text": "ADELA CAMPALLO \n \nADELA SILVA CAMPALLO, bailaora de flamenco, conocida con su propio nombre Cerro del Águila en el seno de una familia de tradición flamenca. De hecho, es en el grupo que forman sus padres: José Galván y Manolo Marín. Pronto encuentra oportunidad de medir su valía, al quedar finalista de los Concurso: colaboración con la compañía de Antonio Canales, de la que es solista. \nLa tierra y el cielo tiembla \ncon solt emociona \ncon solo su cadera moldear, \n transmitiéndoselo al público \ncuando se pone a bailar. \nYa en 2004 Sala Joaquín Turina de Sevilla. El montaje volvió a representarse posteriormente en foros como el Festival de J A Adela Campallo formó parte de la compañía de Javier Latorre en la obra 'Rinconete y Cortadillo', representada c otros escenarios. También colaboró con la compañía de Farruquito en 'Alma vieja', que se estrenó en el Festival de 2005 clausuró la VII Muestra Andaluza de Baile Flamenco, en el emblemático escenario del Corral del Carbón de 2005/2006 en el Teatro Central de Sevilla, tomando parte en la obra 'Otra generación', dirigida por José Miguel didáctica 'Ritmo Flamenco Rythm'. De cara a 2006, la bailaora sevillana participa en Flamenco Festival USA y Fe: bien templadas, en el bronce de Juan José Amador padre, dan paso a las seguriyas con las que comienza la joven grandes maestros de la danza flamenca. Su entrega sin condiciones es reconocida por un público que ovaciona sin fiestas por bulerías. En lo que llevamos de Muestra, éste es el remate más serio y profesional de cuantos se ha sentido se prolongue la fiesta a su término. Tampoco se trata de poner el punto cómico a los postres y hacer ba: flamenco es en que, con tono desenfadado, se ofrezca una nueva guinda, cuasi preparada (pues todos saben se les blanco y negro, los dos tocaores, Mariano Campallo y Paco Iglesias, nos sorprenden con unas deliciosas granaína: foráneos con tanto sentimiento. Adela vuelve, tras estos minutos musicales, con su habitual fuerza, a bailarlo: elegancia sólo queda mermada por las bastedades que propone en sus alegrías, que si bien son comunes en la trad: con recursos más sutiles. Después de unas bulerías que hace el cuadro de atrás, Campallo se prepara para termin: que seduce a los asistentes al espectáculo. La finura de sus movimientos es sobresaliente y su baile en general
  "andaluciaorg_html": "\u003Cp\u003E\u003C/p\u003E\u003Cp\u003EAdela Silva Campallo nació en Sevilla en 1977, en sus inicios en la academia de Manolo Marín y luego pasó a manos de José Galván, donde terminó de forjarse la Hierbabuena, Antonio Canales, Javier Barón, su hermano Rafael Campallo, Adrián Galia y Manolete \u003C/p\u003E\u003Cp\u003E Andaluza de Danza para pasar después a la de Manuela Carrasco Actualmente trabaja con Andrés Marín y tiene a
  "andaluciaorg_url_en": "http://www.andalucia.org/en/flamenco/artists/adela-campallo/",
  "andaluciaorg_full_name": "Adela Silva Campallo",
  "andaluciaorg_url": "http://www.andalucia.org/es/flamenco/artistas/adela-campallo/",
  "andaluciaorg_html_en": "\u003Cp\u003EAdela Silva Campallo was born in Seville in 1977, in the flamenco neighb flamenco school and after that she finished her dancing training with José Galván However, she has improved Campallo, Adrián Galia and Manolete She has taken part in many shows until he joined the Compañía Andaluza de Andrés Marín and has a long experience performing in tablaos of Seville, Barcelona and Tokyo\u003C/p\u003E",
  "flabase_type": "Bailaor",
  "extracted_origin_id": "place760",
  "andaluciaorg_text": "Adela Silva Campallo nació en Sevilla en 1977, concretamente en el flamenquísimo barrio de José Galván, donde terminó de forjarse como bailaora Sin embargo, ha perfeccionado su arte posteriormente Campallo, Adrián Galia y ManoleteHa participado en numerosos espectáculos hasta que se introdujo en la Compañía Marín y tiene a sus espaldas una amplia etapa en tablaos de Sevilla, Barcelona y Tokio",
  "andaluciaorg_text_en": "Adela Silva Campallo was born in Seville in 1977, in the flamenco neighbourhood of C and after that she finished her dancing training with José Galván However, she has improved her art with art Galia and Manolete She has taken part in many shows until he joined the Compañía Andaluza de Danza, after wh: long experience performing in tablaos of Seville, Barcelona and Tokyo",
  "andaluciaorg_origin": "Sevilla",
  "flabase_artist_id": 1.
}
```

Figura 15. Estructura JSON Artists

Albums_db: Se destaca este Dataset porque cada uno de los registros referentes a los álbumes incluye una lista de canciones que maneja valores indispensables como lo son los palos en sí y los artistas. Posteriormente se procederá a extraer estos datos de estas listas en un nuevo conjunto de datos

Campos: artist, flabase_album_id, flabase_artist_id, formato, songs, title. [Songs Fields]: album, album_id, artist, duration, flabase_artist_id, flabase_guitarist_id, flabase_palo_id, flabase_song_id, guitarist, palo, style, title

```
{
  "flabase_album_id": 1,
  "artist": "LA RUBIA",
  "title": "TARANTAS Y EL ANGEL CAIDO",
  "formato": "PIZARRA",
  "songs": [
    {
      "album": "TARANTAS Y EL ANGEL CAIDO",
      "album_id": 1,
      "title": "CANTOS POPULARES",
      "flabase_palo_id": 15,
      "palo": "TARANTAS",
      "flabase_song_id": 1,
      "artist": "LA RUBIA"
    },
    {
      "album": "TARANTAS Y EL ANGEL CAIDO",
      "title": "EL ANGEL CAIDO",
      "album_id": 1,
      "palo": "CANCION",
      "flabase_song_id": 2,
      "artist": "LA RUBIA"
    }
  ]
}
```

Figura 16. Estructura JSON Albums

Palos:

Campos: andaluciaorg_title, andaluciaorg_url, dbpedia_uri, flabase_palo_id, flabase_type, wikipedia_html, wikipedia_text, wikipedia_title, wikipedia_url

por clasificar cada uno individualmente con dos características como lo son la familia de los cantes (Alegres/Tristes) y el compás del mismo (Binario, Ternario, etc..). Esta información se ha obtenido manualmente de la Web *Flamencópolis* (**b10**) y del libro *Pulpitarrita* (**b4**) actualizando finalmente 52 palos.

4.4.3. Conversión y tratado de los datos obtenidos

El primer problema que se plantea es el de poder unir todos los datos en ficheros únicos que aglutinen la información.

4.4.4. Unión de los archivos JSON

Tras un estudio pormenorizado de las estructuras en las que se encuentran los datos obtenidos de FLABASE se decide establecer como formato final para cada uno de los Datasets el CSV separado por comas.

Para conseguirlo se desarrollan una serie de Scripts en Python que permitan la extracción de los campos deseados en un fichero único (uno por cada Dataset).

Conversión Artists:

Los datos que son considerados esenciales para ser usados en la ontología serían :

- **artist_id**: Identificador del artista. flabase_artist_id
- **artist**: Nombre del artista. flabase_title
- **type**: Tipo de artista (Bailaor, Cantaor o Guitarist). flabase_type
- **provincia**: Provincia donde nació el artista. extracted_province

El script desarrollado quedaría como figura a continuación `combinar_ficheros_json.py`:

```

main.py x
1 import os
2 import json
3 import csv
4
5 def combinar_ficheros_json(carpeta, nombre_salida, campos_seleccionados):
6     datos_totales = []
7
8     # Enumerar archivos en la carpeta
9     archivos = [archivo for archivo in os.listdir(carpeta) if archivo.endswith('.json')]
10
11     # Leer y combinar los datos de cada archivo
12     for archivo in archivos:
13         with open(os.path.join(carpeta, archivo), 'r') as f:
14             datos = json.load(f)
15             registro = [datos.get(campo, None) for campo in campos_seleccionados]
16             datos_totales.append(registro)
17
18     # Escribir los datos combinados con una cabecera en un nuevo archivo CSV
19     with open(nombre_salida, 'w', newline='') as f:
20         writer = csv.writer(f)
21         writer.writerow(campos_seleccionados) # Escribir la cabecera
22         writer.writerows(datos_totales) # Escribir los datos
23
24     print(f"Se han combinado {len(archivos)} archivos JSON en '{nombre_salida}'")
25
26 # Especifica los campos que quieres seleccionar de cada archivo
27 campos_seleccionados = ['flabase_artist_id', 'flabase_title', 'flabase_type', 'extracted_province'] # Re
28
29 # Especifica la carpeta y el nombre del archivo de salida
30 carpeta = 'AB'
31 nombre_salida = 'artistas_combinad.csv'
    
```

Figura 19. Código Python Extraer Artists

Tras la ejecución del mismo se obtiene el archivo artistsprov_combined.csv :

	A	B	C	D
1	artist_id,artist,type,province			
2	1,Adela Campallo,Bailaor,Sevilla			
3	10,Alba Carmona,Cantaor,Sevilla			
4	100,Cardo Saban Torres Irigaray,Bailaor,Sevilla			
5	101,Carlos Haro,Guitarist,Málaga			
6	102,Carlos Ledesmann,Guitarist,			
7	103,Carlos Marín,Guitarist,			
8	104,Carlos Montoya,Guitarist,			
9	105,Carlos Piñana,Guitarist,			
10	106,Carmen Amaya,Bailaor,			
11	107,Carmen Cortés,Bailaor,			
12	108,Carmen De La Jara,Cantaor,Cádiz			
13	109,Carmen Flores,Cantaor,Cádiz			
14	11,Alberto San Miguel,Guitarist,Cádiz			
15	110,Carmen Ledesma,Bailaor,Córdoba			
16	111,Carmen Linares,Cantaor,Jaén			
17	112,Carmen María Ros,Guitarist,Almería			
18	113,Carrete,Bailaor,			
19	114,Cascabel De Mairena,Cantaor,Sevilla			
20	115,Cándido De Málaga,Cantaor,Málaga			
21	116,Celedonio Romero,Guitarist,			
22	117,Celia Morales,Guitarist,Málaga			
23	118,Celia Romero,Cantaor,Sevilla			
24	119,Chaconcito,Cantaor,Córdoba			
25	12,Alejandro Conde (cantante),Cantaor,Córdoba			
26	120,Chano Domínguez,Pianist,Córdoba			
27	121,Chano Lobato,Cantaor,Cádiz			
28	122,Chaparro De Málaga,Guitarist,Málaga			
29	123,Chato De La Isla,Cantaor,Cádiz			

Figura 20. CSV datos extraídos Artists

Conversión Places:

Se determina que el dato que puede tener cierta relevancia es el de la provincia. Es por ello que se decide extraer las 8 provincias que componen Andalucía: Almería, Cádiz, Córdoba, Granada, Huelva, Jaén, Málaga y Sevilla. Se opta por hacerlo a mano ya que son muy pocos los registros y no es necesario el desarrollo de un script para dicho fin.

El fichero obtenido será provincias_combined.csv

	A	B
1	province_id,province	
2	1,Almería	
3	2,Cádiz	
4	3,Córdoba	
5	4,Granada	
6	5,Huelva	
7	6,Jaén	
8	7,Málaga	
9	8,Sevilla	
10		
11		

Figura 21. CSV datos extraídos Provincias

Conversión Palos:

La conformación de un fichero único formado por los campos que se consideran esenciales constará de dos partes:

- La creación de un script (JuntaFicheros.py) que extraiga los datos deseados del archivo JSON en un fichero único en formato csv separado por comas.
- Actualización manual de los datos añadiendo los campos familia y compás. Para ello se accede a la información proporcionada por la Web Flamencopolis (**b10**) y el libro Pulpitarrita. (**b4**) Se rellena cada uno de los registros consultando dichas fuentes concluyendo con un fichero csv con 76 registros correspondientes a los palos.

Script:

```

25
26 # Especifica los campos que quieres seleccionar de cada archivo
27 campos_seleccionados = ['flabase_palo_id', 'andaluciaorg_title', 'flabase_type'] # Reem
28
29 # Especifica la carpeta y el nombre del archivo de salida
30 carpeta = 'palos'
31 nombre_salida = 'palos_combined.csv'
32
33 # Llama a la función para unir los archivos JSON
34 combinar_ficheros_json(carpeta, nombre_salida, campos_seleccionados)
    
```

Run: main x

E:\UOC_NEW\AprendizajeComputacional\PRAC\junta_ficheros\Scripts\python.exe C:/Users/rlope/PycharmProjects/junta_ficheros/main.py
 Se han combinado 74 archivos JSON en 'palos_combined.csv'.
 Process finished with exit code 0

Figura 22. Código Python Extraer Palos

El fichero resultante será palos_combined.csv

	A	B	C
1	palo_id,palo,family,beat		
2	1,ALEGRÍAS,alegre, alterno		
3	11,GILIANA,triste,libre		
4	13,MURCIANAS,triste,libre		
5	15,TARANTAS,triste,libre		
6	16,ROMERAS,alegre, ternario		
7	17,RONDEÑAS,triste,binario		
8	18,POLO,triste,alterno		
9	19,MEDIA GRANAINA,triste, libre		
10	2,MILONGA,alegre,alterno		
11	21,ZAMBRA,alegre,libre		
12	22,CAÑA,triste,alterno		
13	23,BANDOLÁ,alegre,ternario		
14	24,CABALES,alegre,binario		
15	25,SERRANAS,triste,alterno		
16	27,MINERAS,triste,libre		
17	3,SEVILLANAS,alegre,ternario		
18	30,FANDANGOS,triste,libre		
19	32,MARTINETES,triste,alterno		
20	33,ROMANCE,triste,libre		
21	34,CARCELERAS,triste,alterno		
22	35,FARRUCA,triste,binario		
23	36,TANGOS,alegre,binario		
24	38,CARTAGENERAS,triste,libre		
25	39,MALAGUEÑAS,triste,libre		
26	4,MIRABRAS,alegre,libre		
27	40.SAETAS.triste.libre		

Figura 23. CSV datos extraídos Palos

(NOTA: Este fichero se continuará filtrando posteriormente por los datos que hay en las canciones. Por tanto, los palos que no son utilizados se eliminan de la lista quedando 52 palos en total)

Conversión Albums_db:

La complejidad de esta extracción viene dada por la inclusión de una lista de canciones en cada uno de los registros de los álbumes.

Cada una de las canciones de dicha lista contiene la información necesaria para poder trabajar con la ontología. Por tanto, se crea el script convJSON2CsvAlbums.py para extraer todos los datos de la lista en un fichero JSON único:

```

1 import json
2 import os
3
4 # Directorio que contiene los archivos JSON
5 directorio = 'AB'
6
7 # se crea la lista para almacenar los datos
8 lista_datos = []
9
10 # se recorre los registros
11 for filename in os.listdir(directorio):
12     if filename.endswith('.json'):
13         # se especifica la ruta del archivo
14         ruta = os.path.join(directorio, filename)
15
16         # se lee archivo JSON y se convierte en diccionario Python
17         with open(ruta, 'r') as file:
18             datos = json.load(file)
19
20         # Se extraen los datos de la lista songs
21         if 'songs' in datos:
22             lista_datos.extend(datos['songs'])
23
24 # Se crea un nuevo diccionario con la lista de los datos
25 nuevo_json = lista_datos
with open('albums_json.json', 'w') as f:
    f.write(json.dumps(nuevo_json))

```

Run: main x
E:\UOC_NEW\AprendizajeComputacional\PRAC\convJSON2CsvAlbums\Scripts\python.exe C:/Users/rLope/PycharmProjects/convJS
Process finished with exit code 0

Figura 24. Código Python Extraer Albums

El Script formará un fichero JSON con la siguiente estructura de datos:

```
[
  {
    "album": "TARANTAS Y EL ANGEL CAIDO",
    "album_id": 1,
    "title": "CANTOS POPULARES",
    "flabase_palo_id": 15,
    "palo": "TARANTAS",
    "flabase_song_id": 1,
    "artist": "LA RUBIA"
  },
  {
    "album": "TARANTAS Y EL ANGEL CAIDO",
    "title": "EL ANGEL CAIDO",
    "album_id": 1,
    "palo": "CANCION",
    "flabase_song_id": 2,
    "artist": "LA RUBIA"
  },
  {
    "album": "SAETAS",
    "palo": "SAETAS",
    "album_id": 10,
    "title": "HABIA SONAO UN CLARIN RONCO",
    "flabase_artist_id": "221",
    "flabase_palo_id": 40,
    "guitarrist": "BANDA DE CORNETAS Y TAMBORES",
    "flabase_song_id": 17,
    "duration": "2:48",
    "artist": "EL GLORIA"
  },
  {
    "album": "SAETAS",
    "palo": "SAETAS",
    "album_id": 10,
    "title": "TODAS LAS MARES TIENEN PENAS",
    "flabase_artist_id": "221",
    "flabase_palo_id": 40,
    "guitarrist": "BANDA DE CORNETAS Y TAMBORES",
    "flabase_song_id": 18,
    "duration": "3:00",
    "artist": "EL GLORIA"
  },
  {
    "album": "SAETAS",
    "palo": "SAETAS",
    "album_id": 10,
    "title": "ERES GUAPA Y SEVILLANA",
    "flabase_artist_id": "221",
    "flabase_palo_id": 40,
    "guitarrist": "BANDA DE CORNETAS Y TAMBORES",
    "flabase_song_id": 19,
    "artist": "EL GLORIA"
  }
].
```

Figura 25. Nueva estructura JSON Albums

Para la conversión a csv se utiliza el programa Excel Se abre el JSON con el mismo y se formatea el resultado quedando del siguiente modo:

album	album_id	title	flabase_palo_id	palo	flabase_song_id	artist	flabas	guitarist	durati
TARANTAS Y EL ANGEL CAIDO	1	CANTOS POPULARES	15	TARANTAS		1	LA RUBIA		
TARANTAS Y EL ANGEL CAIDO		EL ANGEL CAIDO		CANCION		2	LA RUBIA		
SAETAS	10	HABIA SONAO UN CLARIN RONCO	40	SAETAS		17	EL GLORIA	221	BANDA DE CORNETAS Y TAMBORES
SAETAS	10	TODAS LAS MARES TIENEN PENAS	40	SAETAS		18	EL GLORIA	221	BANDA DE CORNETAS Y TAMBORES
SAETAS	10	ERES GLAPA Y SEVILLANA	40	SAETAS		19	EL GLORIA	221	BANDA DE CORNETAS Y TAMBORES
SAETAS	10	MIRARLE AL SEÑOR LA CARA	40	SAETAS		20	EL GLORIA	221	BANDA DE CORNETAS Y TAMBORES
TANGO NÚMERO 2	100		36	TANGOS		187	NIÑA DE JEREZ		
SEGUIDILLAS GITANAS Y FANDANGUILLOS	1000		30	FANDANGOS		1947	EL MOCHUELO	240	MANUEL LOPEZ
SEGUIDILLAS GITANAS Y FANDANGUILLOS	1000		5	SIGUIRIYAS		1948	EL MOCHUELO	240	MANUEL LOPEZ
BULERIAS Y SEGUIDILLAS	1001	QUE DOLOR DE MI MADRE	5	SIGUIRIYAS		1949	JOSE CEPERO	457	NIÑO RICARDO
BULERIAS Y SEGUIDILLAS	1001	TE HE QUERIO TANTO	42	BULERIAS		1950	JOSE CEPERO	457	NIÑO RICARDO
ALEGRÍAS Y TARANTAS DEL NIÑO DE MARCHENA	1002		15	TARANTAS		1951	NIÑO DE MARCHENA		RAMON MONTOYA
ALEGRÍAS Y TARANTAS DEL NIÑO DE MARCHENA	1002		1	ALEGRÍAS		1952	NIÑO DE MARCHENA		RAMON MONTOYA
TIENTOS Y JOTA ARAGONESA NUMERO 1	1003		63	TIENTOS		1953	EL MOCHUELO	240	
VAMOS A ACORDARNOS	1004	CON QUE DUQUELAS	5	SIGUIRIYAS		1954	SERNITA DE JEREZ		PACO HERNANDEZ
VAMOS A ACORDARNOS	1004	CUANDO VIENE EL DIA	40	SAETA		1955	SERNITA DE JEREZ		PACO HERNANDEZ
VAMOS A ACORDARNOS	1004	EL QUERER DE ESTA FLAMENCA		SOLEARES		1956	SERNITA DE JEREZ		PACO HERNANDEZ
VAMOS A ACORDARNOS	1004	POR LA CALLE VAN DICIENDO	66	RUMBA		1957	SERNITA DE JEREZ		PACO HERNANDEZ
VAMOS A ACORDARNOS	1004	QUE ESTOY ETICO DE PENAS	1	ALEGRÍAS		1958	SERNITA DE JEREZ		PACO HERNANDEZ
VAMOS A ACORDARNOS	1004	RUGE DE CELOS		SERRANA Y VERDIAL		1959	SERNITA DE JEREZ		PACO HERNANDEZ
VAMOS A ACORDARNOS	1004	SOY EL LUCERO DEL ALBA	30	FANDANGOS		1960	SERNITA DE JEREZ		PACO HERNANDEZ
FANDANGOS Y MALAGUEÑAS	1005	LA CANTAORA	30	FANDANGOS		1961	MANUEL VALLEJO	702	ANTONIO MORENO
SOLEARES NUMERO 1 Y POLO	1006			SOLEARES		1962	LA RUBIA		
LOS OJITOS NEGROS	1007					1963	TERESA ESPAÑA	1082	
SEGUIDILLAS 1 Y 2	1008	A CLAVITO Y CANELA	5	SIGUIRIYAS		1964	FERNANDO EL HERRERO	340	RAMON MONTOYA
SEGUIDILLAS 1 Y 2	1008	HASTA LAS PIEDRAS	5	SIGUIRIYAS		1965	FERNANDO EL HERRERO	340	RAMON MONTOYA
SEVILLANAS NUMERO 2 Y RONDEÑAS	1009		3	SEVILLANAS		1966	LA RUBIA		
FANDANGOS Y MILONGA	101	ANOCHÉ MIENTRAS DORMIA	2	MILONGA		188	JOSE CEPERO	457	NIÑO RICARDO
FANDANGOS Y MILONGA	101	CUANDO ME VEAS VENIR	30	FANDANGOS		189	JOSE CEPERO	457	NIÑO RICARDO
SEGUIDILLAS Y BULERIAS	1010	QUE DOLOR DE MI MARE	5	SIGUIRIYAS		1967	ISABELITA DE JEREZ	411	MANOLO EL DE BADAJOZ
SEGUIDILLAS Y BULERIAS	1010	YO DE NADIE HE MURMURAO		BULERIAS POR SOLEÁ		1968	ISABELITA DE JEREZ	411	MANOLO EL DE BADAJOZ

Figura 26. CSV datos extraídos Albums

Se procede a un limpiado de los datos del siguiente modo:

- Se eliminan los registros que carecen de palos
- Se eliminan los registros que no son interpretados por un artista reconocido.
- Se eliminan todas la columnas que carecen de relevancia
- Se cambian los nombres de las cabeceras.

Una vez limpiado y filtrado el contenido del csv los campos quedarán como se observa a continuación:

- album_id: Identificador del disco que incluye la canción.
- album: Nombre del disco.
- palo_id: Identificador del palo de la canción.
- palo: Nombre del palo de la canción.
- artist_id: Identificador del artista que interpreta la canción.
- artist: Identificador del artista.
- guitarist_id: Identificador del guitarrista que toca en la canción
- guitarist: Identificador del guitarrista

	C	D	E	F	G	H	I	J
1	album_id	album	palo_id	palo	artist_id	artist	guitarist_id	guitarist
2	1	TARANTAS Y EL ANGEL CAIDO	15	TARANTAS		LA RUBIA		
3	10	SAETAS	40	SAETAS	221	EL GLORIA		BANDA DE CORNETAS Y TAMBORES
4	10	SAETAS	40	SAETAS	221	EL GLORIA		BANDA DE CORNETAS Y TAMBORES
5	10	SAETAS	40	SAETAS	221	EL GLORIA		BANDA DE CORNETAS Y TAMBORES
6	10	SAETAS	40	SAETAS	221	EL GLORIA		BANDA DE CORNETAS Y TAMBORES
7	1001	BULERIAS Y SEGUIDILLAS	5	SIGUIRIYAS	457	JOSE CEPERO	857	NIÑO RICARDO
8	1001	BULERIAS Y SEGUIDILLAS	42	BULERÍAS	457	JOSE CEPERO	857	NIÑO RICARDO
15	1008	SEGUIDILLAS 1 Y 2	5	SIGUIRIYAS	340	FERNANDO EL HERRERO	1028	RAMON MONTOYA
16	1008	SEGUIDILLAS 1 Y 2	5	SIGUIRIYAS	340	FERNANDO EL HERRERO	1028	RAMON MONTOYA
17	101	FANDANGOS Y MILONGA	2	MILONGA	457	JOSE CEPERO	857	NIÑO RICARDO
18	101	FANDANGOS Y MILONGA	30	FANDANGOS	457	JOSE CEPERO	857	NIÑO RICARDO
19	1010	SEGUIDILLAS Y BULERIAS	5	SIGUIRIYAS	411	ISABELITA DE JEREZ		MANOLO EL DE BADAJOZ
20	1012	GRANAINAS Y CARTAGENERAS	39	MALAGUEÑAS	34	DON ANTONIO CHACON		PERICO DEL LUNAR
21	1015	TARANTA DE LA GABRIELA Y MALAGUEÑA DE CHACOI	15	TARANTAS	578	NIÑA DE LOS PEINES		LUIS MOLINA
22	1015	TARANTA DE LA GABRIELA Y MALAGUEÑA DE CHACOI	38	CARTAGENERAS	578	NIÑA DE LOS PEINES		LUIS MOLINA
23	1016	BULERÍAS AL GOLPE	42	BULERÍAS		ANTONIA SUAREZ LA CHIVA		ANTONIO PEREZ
24	1017	BULERIAS Y SEVILLANAS NUMERO 1	3	SEVILLANAS	578	NIÑA DE LOS PEINES		LUIS MOLINA
25	1017	BULERIAS Y SEVILLANAS NUMERO 1	42	BULERÍAS	578	NIÑA DE LOS PEINES		LUIS MOLINA
26	1018	MALAGUEÑA Y FANDANGUILLO	39	MALAGUEÑAS		NIÑO DE MARCHENA	1028	RAMON MONTOYA
27	1018	MALAGUEÑA Y FANDANGUILLO	30	FANDANGOS		NIÑO DE MARCHENA	1028	RAMON MONTOYA
28	1019	RONDEÑAS	17	RONDEÑAS	78	BENI DE CALDIZ		ANDRES HEREDIA
29	102	CIEÑ AÑOS DE CANTE GITANO	57	TONÁ	54	ANTONIO MAIRENA	778	MELCHOR DE MARCHENA
30	102	CIEÑ AÑOS DE CANTE GITANO	5	SIGUIRIYAS	54	ANTONIO MAIRENA	778	MELCHOR DE MARCHENA
31	102	CIEÑ AÑOS DE CANTE GITANO	61	CANTIÑAS	54	ANTONIO MAIRENA	778	MELCHOR DE MARCHENA
32	102	CIEÑ AÑOS DE CANTE GITANO	30	FANDANGOS	54	ANTONIO MAIRENA	778	MELCHOR DE MARCHENA
33	102	CIEÑ AÑOS DE CANTE GITANO	6	SIGUIRIYAS	54	ANTONIO MAIRENA	778	MELCHOR DE MARCHENA

Figura 27. Excel final Canciones

Finalmente se extraen los campos Album e Album_id y se crea un nuevo CSV con los mismos.

4.4.5. Conversión final de los datos

Debido al posterior intento de poblar la ontología a través de Protégé se comprueba que los ficheros finales en el formato CSV separados por comas no son aptos para dicha tarea.

Por tanto debido al nuevo método de carga de instancias se vuelven a transformar estos ficheros con la extensión .xlsx (archivo de Microsoft Excel) que es el que admite el plugin Cellfie de Protégé para la importación de los mismos.

Los ficheros finales para la importación quedarán del siguiente modo:

Artista	Genero	TipoArtista	Provincia
Adela Campallo	Mujer	Bailaor	Sevilla
Alba Carmona	Mujer	Cantaor	Sevilla
Cardo Saban Torre	Hombre	Bailaor	Sevilla
Carlos Haro	Hombre	Guitarist	Málaga

Figura 28. Artistas_excel.xlsx

Cancion	Album	Palo	Artist
CANTOS POPULARES	TARANTAS Y EL ANGEL CAIDO	TARANTAS	LA RUBIA
HABIA SONAO UN CLARIN RONCO	SAETAS	SAETAS	EL GLORIA
TODAS LAS MARES TIENEN PENAS	SAETAS	SAETAS	EL GLORIA
ERES GUAPA Y SEVILLANA	SAETAS	SAETAS	EL GLORIA
MIRARLE AL SEÑOR LA CARA	SAETAS	SAETAS	EL GLORIA
QUE DOLOR DE MI MADRE	BULERIAS Y SEGUIDILLAS	SIGUIRIYAS	JOSE CEPERO
TE HE QUERIO TANTO	BULERIAS Y SEGUIDILLAS	BULERÍAS	JOSE CEPERO

Figura 29. Canciones_excel.xlsx

palo	familia	compas
ALEGRÍAS	alegre	alterno
GILIANA	triste	libre
MURCIANAS	triste	libre
TARANTAS	triste	libre
ROMERAS	alegre	ternario
RONDEÑAS	triste	binario
POLO	triste	alterno

Figura 30. Palos_excel.xlsx

provincia_id	Provincia
1	Almería
2	Cádiz
3	Córdoba

Figura 31. Provincias_excel.xlsx

4.5. Diseño de la ontología

El objetivo a la hora de elaborar una ontología es el de poder responder una serie de preguntas al sistema y que pueda responder correctamente con el juego de datos proporcionado.

Tal y como se especificaba en el [punto 4.3](#), tras una búsqueda exhaustiva de no se haya ninguna que resulte óptima. Por tanto la ontología se establecerá de cero.

4.5.1. Software utilizado

Para la elaboración de la misma se utilizará Protégé en su versión de escritorio con la última versión publicada hasta la fecha: Protégé 5.6.4.

4.5.2. Identificación de los conceptos

Teniendo como punto de partida el modelado final extraído de nuestro juego de datos se obtiene una serie de entidades sobre las que pivotará el axioma.

En un primer momento se establece una primera versión sobre la que trabajar. La jerarquía quedaría estructurada en un conjunto de clases y subclases que tendrían el siguiente esquema:

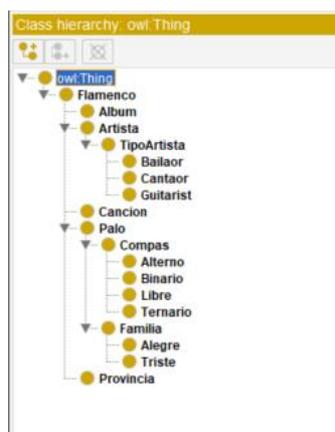


Figura 32. Primera versión arquitectura de clases de ontología

Album	Conjunto de canciones	Clase
Artista	Persona que interpreta	Clase
TipoArtista	Especialidad artística del artista	Subclase Artista
Bailaor	Expresión artística con el baile	Subclase de TipoArtista
Cantaor	Expresión artística con la voz	Subclase de TipoArtista
Guitarist	Expresión artística con la guitarra	Subclase de TipoArtista
Canción:	Entidad	Clase
Palo	Tipo o estilo de composición	Clase
Compas	Ritmo de una canción o característica del Palo	Subclase de Palo
Binario	Ritmo de canción binario	Subclase de compás
Ternario	Ritmo de canción ternario	Subclase de compás
Alterno	Ritmo de canción alterno	Subclase de compás
Libre	Ritmo de canción libre	Subclase de compás
Familia	Si los palos o las canciones son tristes o alegres.	Subclase de Palo
Alegre	La temática pertenece a los cantos alegres	Subclase de Familia
Triste	La temática pertenece a los cantos tristes	Subclase de Familia
Provincia	Provincias de andalucia.	Clase

Tabla 1. I Versión modelado de clases

Una vez creado un conjunto de propiedades de objeto, de datos y un conjunto de restricciones se intenta poblar e importar los datos a través de la elaboración de una serie de Scripts sin ningún éxito. Tras varios intentos fallidos se opta por utilizar el plugin Cellfie el cual requiere los datos en otro formato. Al realizar las consultas de inserción en sintaxis Manhattan el acceso a las subclases del Compas y de la Familia no terminan de funcionar correctamente por lo que se opta por una nueva versión en la que sólo habrá clases y las subclases anteriores pasarán a ser elementos de las clases de tipo enum.

La nueva concepción de las clases quería del siguiente modo en la versión II:

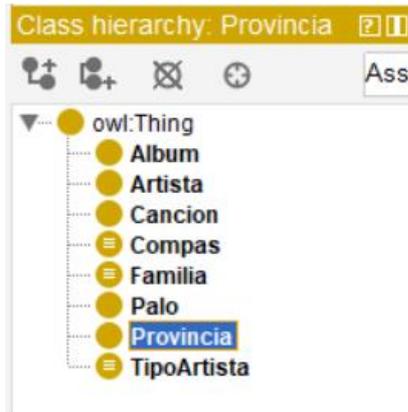


Figura 33. 2ª versión Arquitectura de clases de la ontología.

Album	Conjunto de canciones	Clase
Artista	Persona que interpreta	Clase
TipoArtista	Especialidad artística del artista	Clase enum con los valores: Bailaor, Cantaor y Guitarist
Canción:	Entidad	Clase
Palo	Tipo o estilo de composición	Clase
Compas	Ritmo de una canción o característica del Palo	Clase enum con los valores: Binario, Ternario, Libre y Alternativo
Familia	Si los palos o las canciones son tristes o alegres.	Clase enum con los valores: Alegre, Triste
Provincia	Provincias de andalucia.	Clase

Tabla 2. II Version Modelado de clases

Las clases Enum se crean añadiendo instancias de los elementos a enumerar y posteriormente se genera una equivalencia de los mismos:

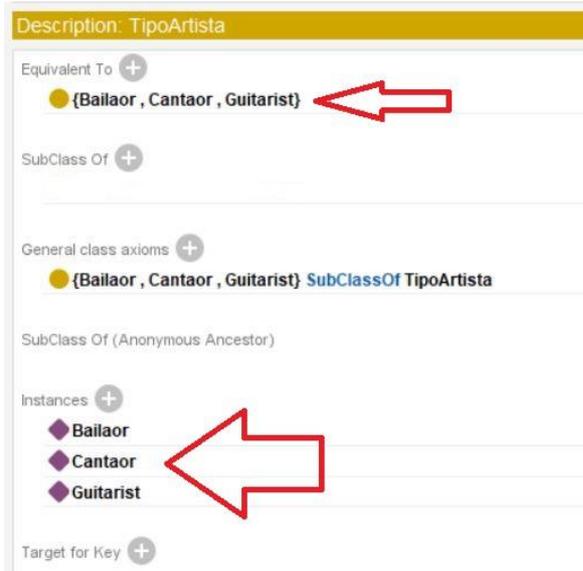


Figura 34. Ejemplo Clase Enum TipoArtista

El esquema Ontograf de la ontología quedaría del siguiente modo:

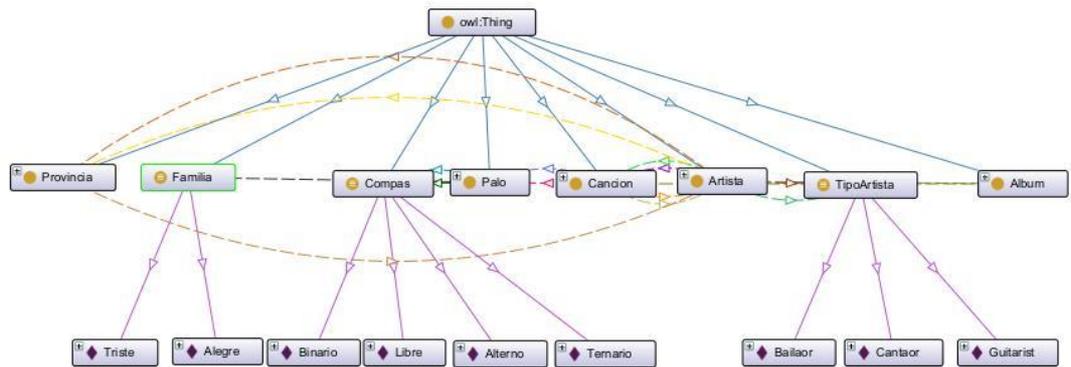


Figura 35. Esquema Ontograf del modelo de datos

4.5.3. Propiedades de objeto de las clases

Se determinan las siguientes clases de objeto:



Figura 36. Propiedades de Objeto

Nombre	Dominio	Rango	Inversa de
esDelTipo	Palo	Familia	
perteneceAlAlbum	Canción	Album	contieneLaCancion
contieneLaCancion	Album	Canción	perteneceAlAlbum
esInterpretadaPor	Canción	Artista	interpretaCancion
interpretaCancion	Artista	Canción	esInterpretadaPor
tieneCompas	Palo	Compás	
esDelPalo	Canción	Palo	
nacidoEn	Artista	Provincia	esLugarDeNacimientoDe
esLugarDeNacimientoDe	Provincia	Artista	nacidoEn
suExpresionArtisiticaEs	Artista	TipoArtista	

Tabla 3. Tabla Propiedades de Objeto

4.5.4. Restricciones

Se perfilan las propiedades con las siguientes restricciones:

Propiedad	Valor	Clase
perteneceAlAlbum	Some	album
esInterpretadaPor	Max 1	Artista
esDelPalo	Max 1	Palo
tieneCompas	Max 1	Compás
interpretaCancion	Min 1	Canción
nacidoEn	Max 1	Provincia

suExpresionArtisticaEs	Max 1	TipoArtista
------------------------	-------	-------------

Tabla 4. Tabla de restricciones de propiedades

4.5.5. Propiedades de datos de las clases

Se determinan las siguientes propiedades de datos:



Figura 37. Lista propiedades de datos.

Nombre	Dominio	Tipo
tieneGenero	Artista	String
tieneNombreAlbum	Album	String
tieneNombreArtista	Artista	String
tieneNombreCancion	Canción	String
tieneNombreCompas	Compás	String
tieneNombreExpresion	TipoArtista	String
tieneNombreFamilia	Familia	String
tieneNombrePalo	Palo	String
tieneNombreProvincia	Provincia	String

Tabla 5. Propiedades de datos

Propiedades de datos inferidas a través de reglas SWRL.

Nombre	Dominio	Tipo
esDeLaFamilia	Canción	String
esDelRitmo	Canción	String

Tabla 6. Propiedades Datos Inferidas

Estas propiedades inferidas se generan tal y como se explica en el punto [“Consultas y reglas de inferencia con SWRL y SQWRL”](#) .

4.5.6. Instancias

Se tratan de importar los datos obtenidos de los datasets especificados y tratados en el apartado Datos.

La primera opción es la utilización la librería Jena en Eclipse para el lenguaje Java.

Surge el problema de que ya no permite la inserción de datos en csv a ontología. Tal y como se especifica en los foros de la propia plataforma de Protégé.

<https://protege-user.stanford.narkive.com/OV96Tj4b/csv-files-in-jena-ontology-api> (b11)

Se intenta, sin éxito, desarrollar una serie de Scripts en Python que permitan la importación de los datasets como individuals en la ontología. Para ello se instalan las bibliotecas Panda y rdflib.

Otra de las posibilidades en las que se está trabajando es el uso de la aplicación de comandos ROBOT que trabaja con este tipo de funcionalidades para ontologías.

<https://robot.obolibrary.org/>

Finalmente, y tras una exhaustiva búsqueda de cómo poder llevarlo a cabo, se opta por la utilización de la aplicación Cellfie a través de la opción “Create Axioms from Excel workbook” que proporciona Protégé.

A través de Cellfie se crean una serie de scripts que permiten realizar la inserción de los datasets en formato Excel con extensión .xlsx.

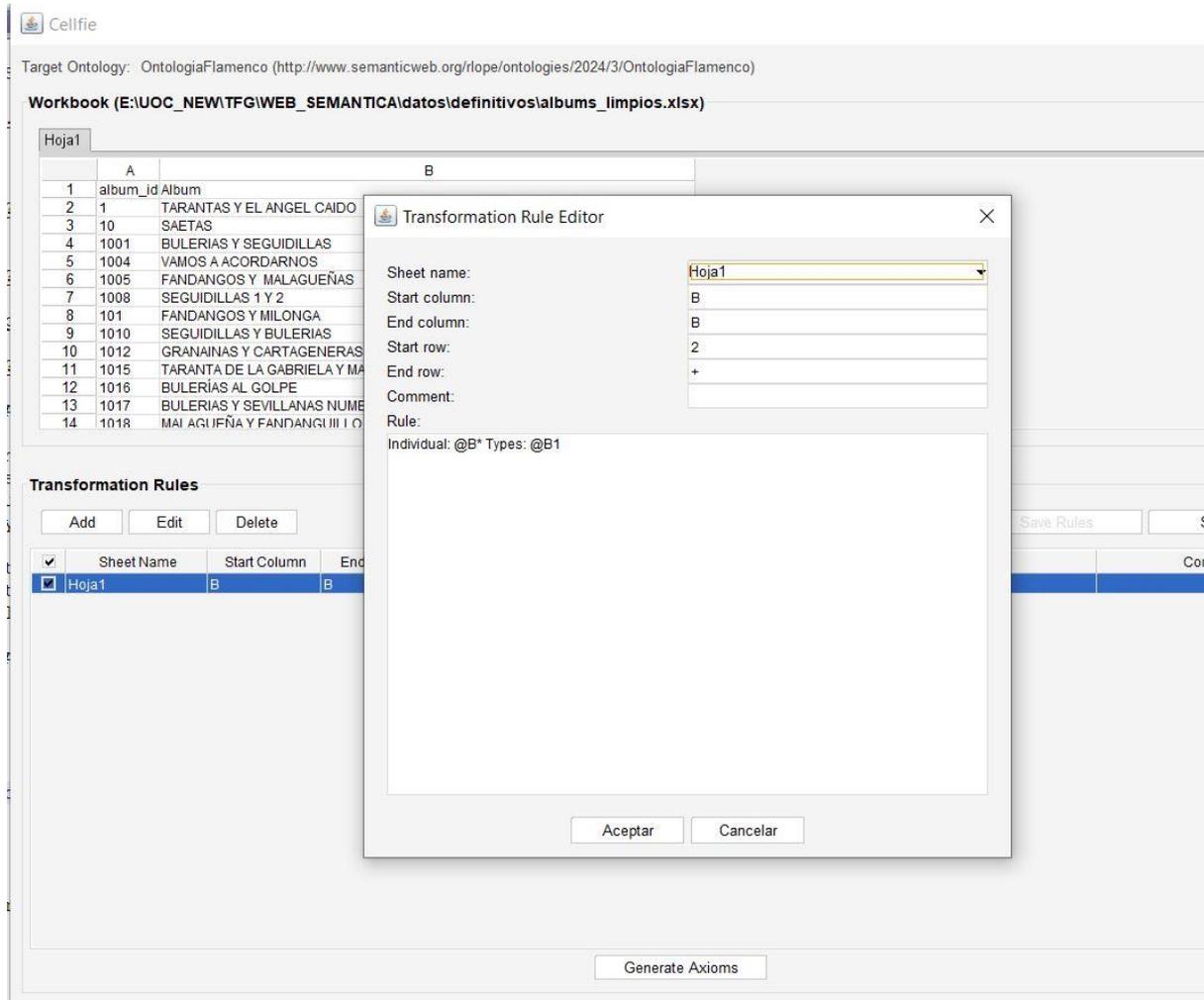


Figura 38. Cellfie inserción Albums

Cellfie

Target Ontology: OntologiaFlamenco (<http://www.semanticweb.org/rlope/ontologies/2024/3/OntologiaFlamenco>)

Workbook (E:\UOC_NEW\TFG\WEB_SEMANTICA(datosdefinitivos)\palos_excel.xlsx)

palos_combined Hoja1

	A	B	C	D
1	palo_id	Palo	Familia	Compás
2	1	ALEGRÍAS	alegre	alterno
3	11	GILIANA	triste	libre
4	13	MURCIANAS	triste	libre
5	15	TARANTAS	triste	libre
6	16	ROMERAS	alegre	ternario
7	17	RONDEÑAS	triste	binario
8	18	POLO	triste	alterno
9	19	MEDIA GRANAINA	triste	libre
10	2	MILONGA	alegre	alterno
11	21	ZAMBRA	alegre	libre
12	22	CAÑA	triste	alterno
13	23	BANDOLÁ	alegre	ternario
14	24	CABALES	alegre	binario
15	25	SEPIANAS	triste	alterno

Transformation Rules

Add Edit Delete

<input checked="" type="checkbox"/>	Sheet Name	Start Column	End Column	Start Row
<input checked="" type="checkbox"/>	palos_combined	B	B	2

Transformation Rule Editor

Sheet name: palos_combined

Start column: B

End column: B

Start row: 2

End row: 52

Comment:

Rule:

Individual: @B* Types: @B1 Facts: tieneCompas @D*, esDelTipo @C*

Aceptar Cancelar

Figura 39. Cellfie Inserción Palos

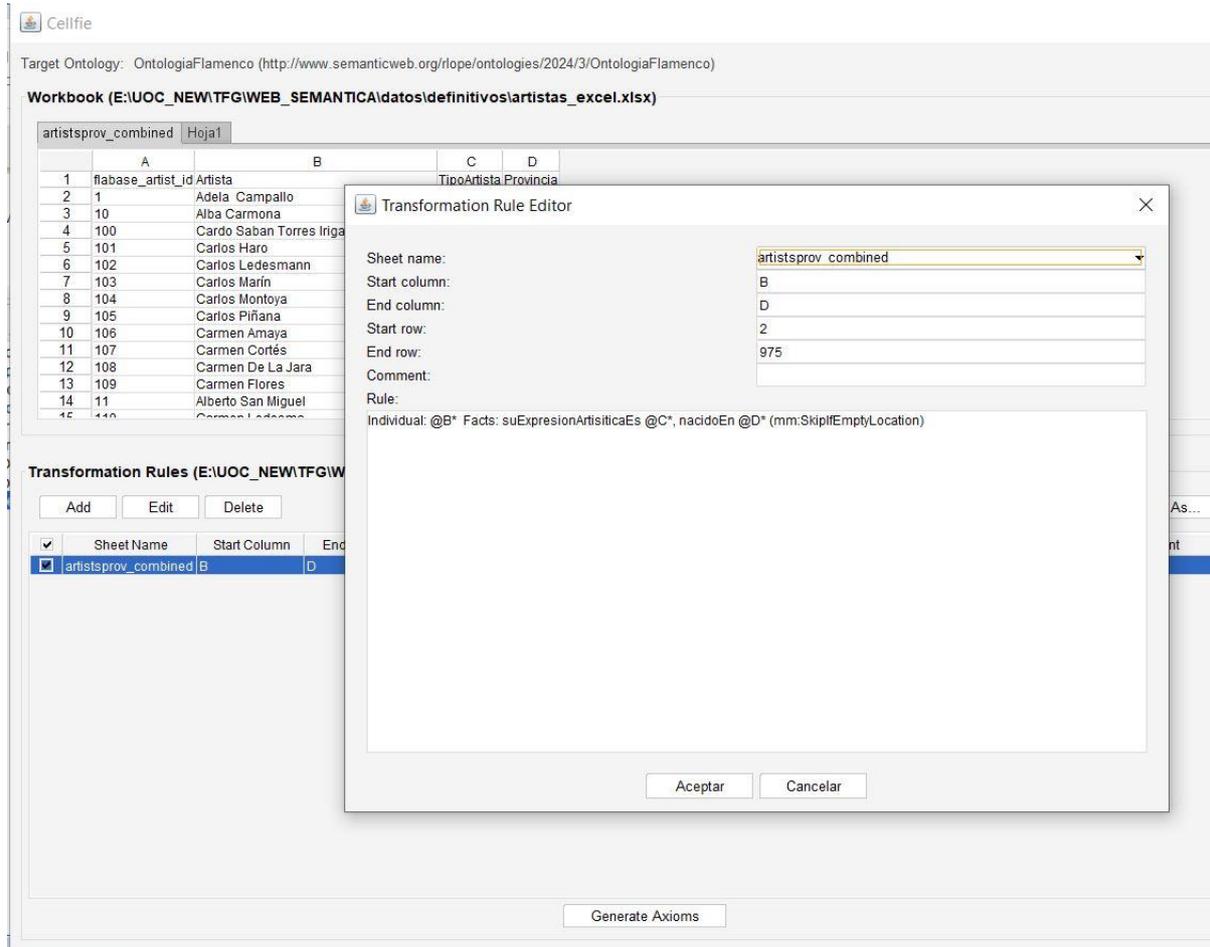


Figura 40. Cellfie Inserción Artistas

Una vez ejecutados estos scripts quedaría poblada la ontología y se podría comenzar a realizar consultas y razonamientos inferidos relacionados con el mundo del flamenco representado en la misma.

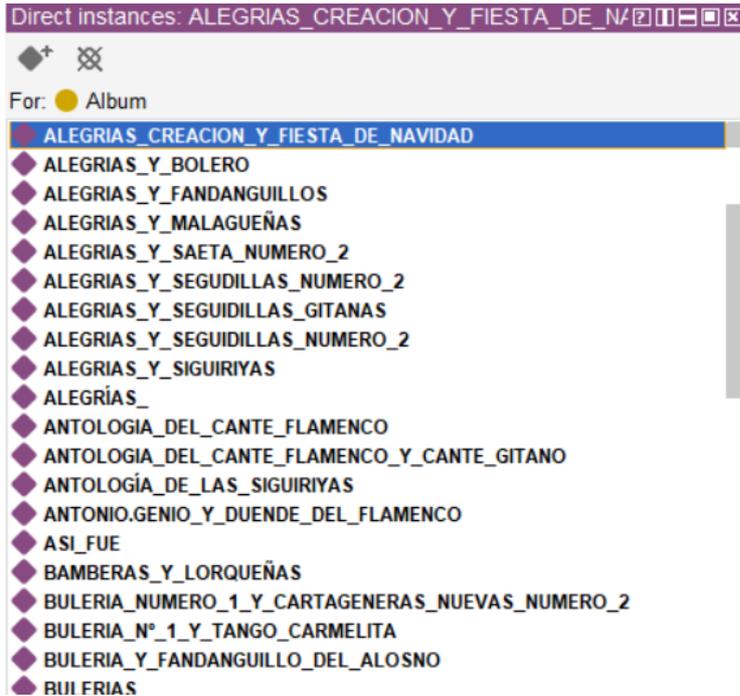


Figura 41. Instancias/individuals de los albums

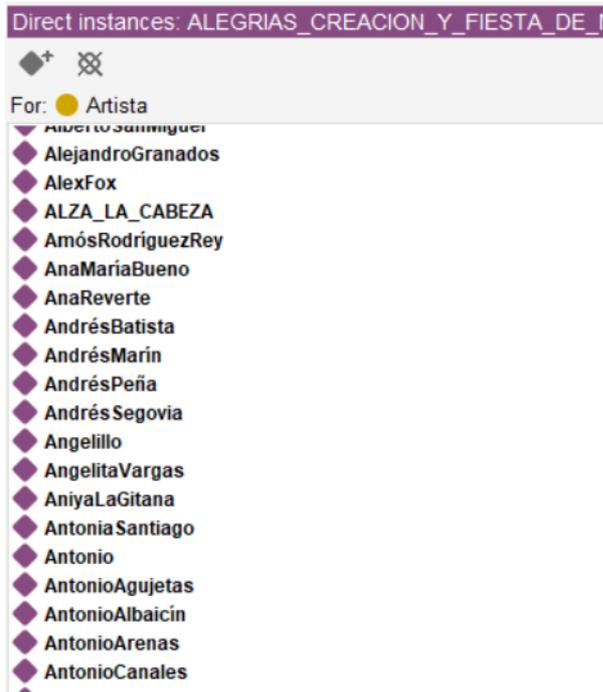


Figura 42. Instancias/individuals de los albums

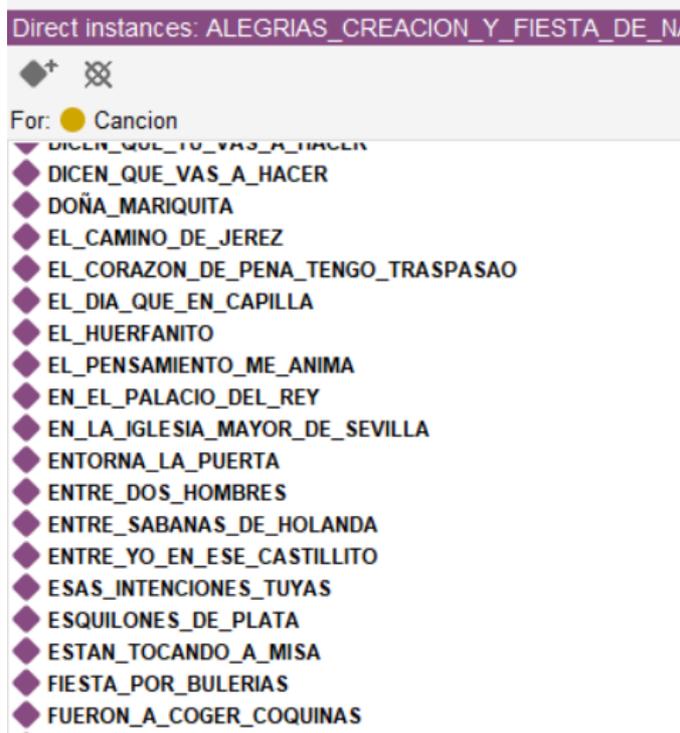


Figura 43. Instancias/individuals de los canciones

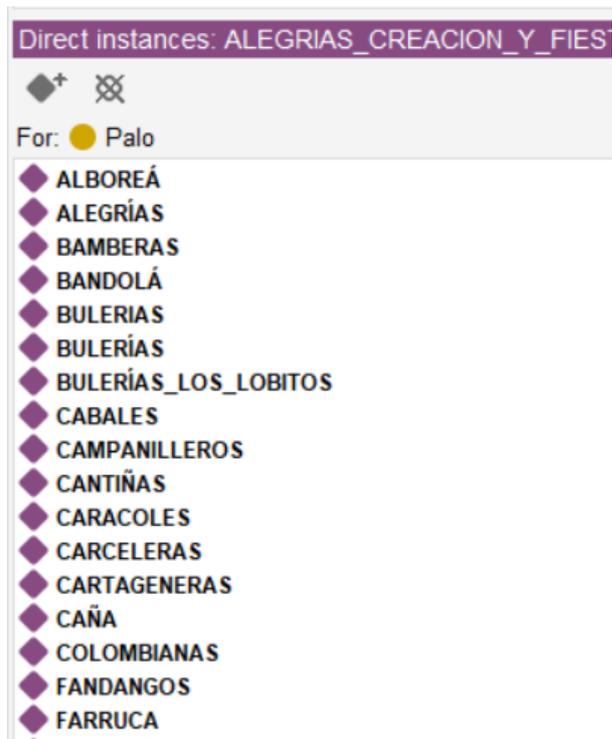


Figura 44. Instancias/individuals de los palos



Figura 45. Instancias/individuals de los provincias

4.5.7. Diseño final del sistema

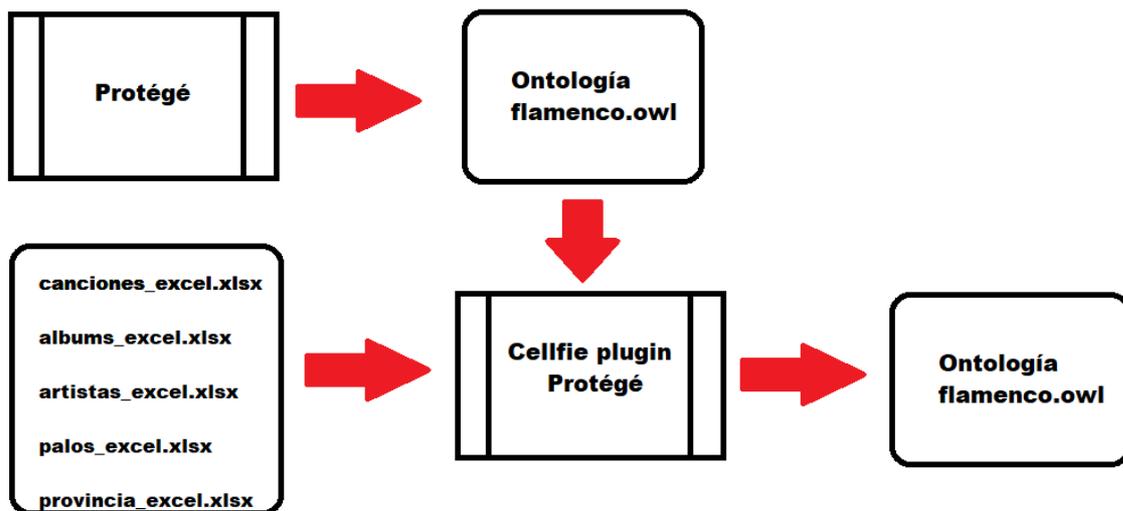


Figura 46. Diseño final del sistema

La ontología se crea en Protégé donde, a través del plugin Cellfie, también se puebla con los datos previamente tratados. Una vez realizadas las importaciones y una vez definidas las clases, sus propiedades, sus restricciones y sus reglas inferidas se puede considerar que la ontología está preparada para poder ser utilizada. Ontología_Flamenco.owl queda disponible para su futuro aprovechamiento en alguna aplicación, alguna Wiki o algún portal web ya sea cuestionando o infiriendo razonamientos relativos al flamenco.

5. Implementación

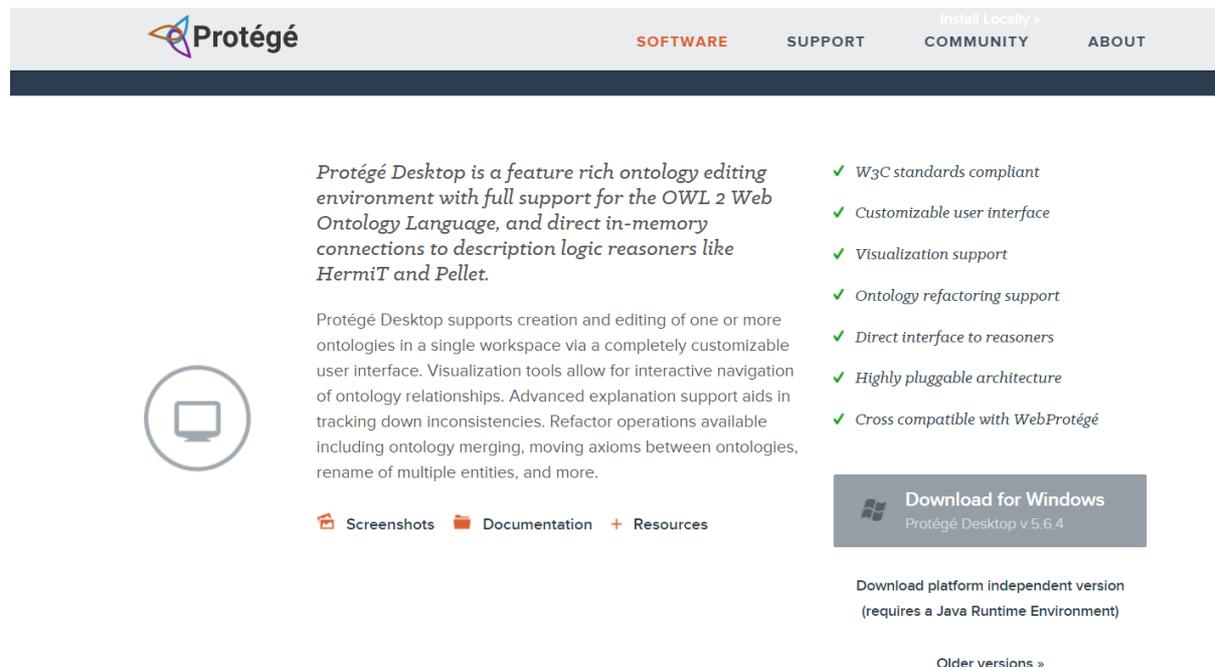
Para ejecutar consultas a la ontología se podrá utilizar cualquier sistema que soporte consultas SPARQL o SWRL. Al ser un primer acercamiento al flamenco a través del mundo de la web semántica se decide dejar la ontología disponible para poder ser utilizada en cualquier tipo de sistema que pueda soportar los razonamientos y los interrogantes a la misma.

Por tanto, la implementación se realizará íntegra en el programa Protégé en su versión de escritorio.

5.1. Instrucciones de instalación

Los pasos a seguir para la instalación son bastante sencillos. La descarga del software de la versión de escritorio de Protégé hay que realizarla a través de la siguiente dirección:

<https://protege.stanford.edu/software.php>



The screenshot shows the Protégé website's software page. At the top, there is a navigation bar with the Protégé logo and links for SOFTWARE, SUPPORT, COMMUNITY, and ABOUT. The main content area features a description of Protégé Desktop, a list of features, and download options. The features list includes: W3C standards compliant, Customizable user interface, Visualization support, Ontology refactoring support, Direct interface to reasoners, Highly pluggable architecture, and Cross compatible with WebProtégé. There are also links for Screenshots, Documentation, and Resources. A prominent button says 'Download for Windows Protégé Desktop v.5.6.4'. Below it, there is a link for 'Download platform independent version (requires a Java Runtime Environment)' and a link for 'Older versions »'.

Figura 47. Descarga/Instalación de Protégé.

Una vez descargado se procederá a ejecutar el run.bat (si se trabaja con Windows) que se encuentra en el mismo directorio de instalación.

Para ejecutar Protégé 5.6.4., conviene tener instalada la versión de Java 8. Protégé recomienda explícitamente Java 8, ya que versiones anteriores o posteriores pueden no ser compatibles o no ofrecer la misma estabilidad.

Si fuera necesaria la instalación de Java 8 se puede realizar desde aquí:

<https://www.java.com/es/download/>

Una vez instalado todo conviene actualizar los siguientes plugins que se van a utilizar posteriormente:

- Cellfie Protege 5.0+ Plugin (sólo en el caso de querer poblar la ontología con más datos de forma automática a través de ficheros con extensión .xlsx)
- Snap SPARQL Query
- SWRL Tab Plugin

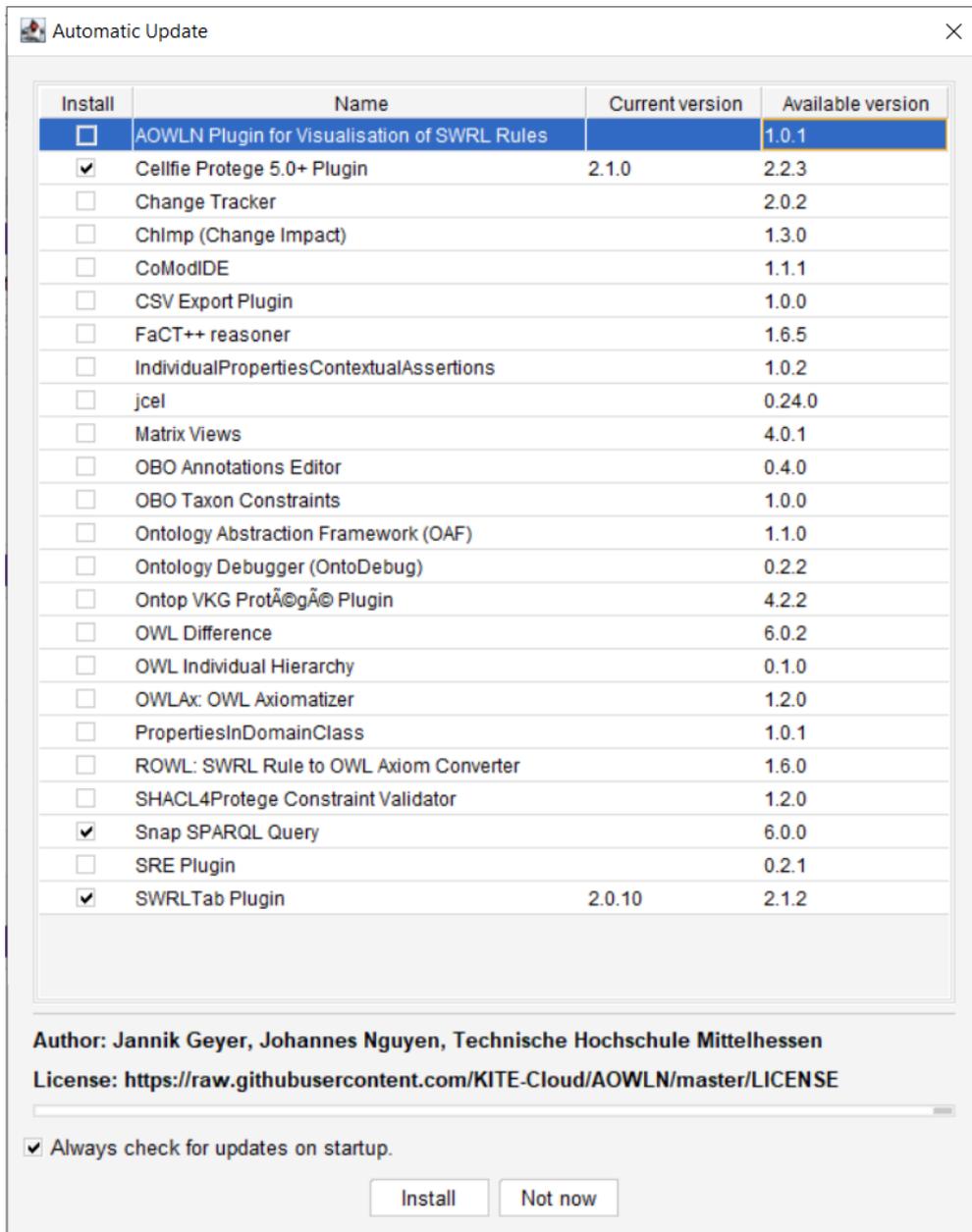


Figura 48. Instalación de plugins Protégé

5.2. Consultas y reglas de inferencia

Con la ontología poblada y con la verificación de que no existen inconsistencias, es el momento de extraer el conocimiento representado y asegurarse de que la ontología responde adecuadamente a las preguntas que se le formulen. Esta tarea se lleva a cabo utilizando dos lenguajes con los que poder realizar consultas: SWRL/SQWRL y SPARQL. Estas funcionalidades están disponibles en Protégé 5.6.4. mediante las pestañas “SPARQL Query”, “SWRLTab” y “SQWRLTab”.

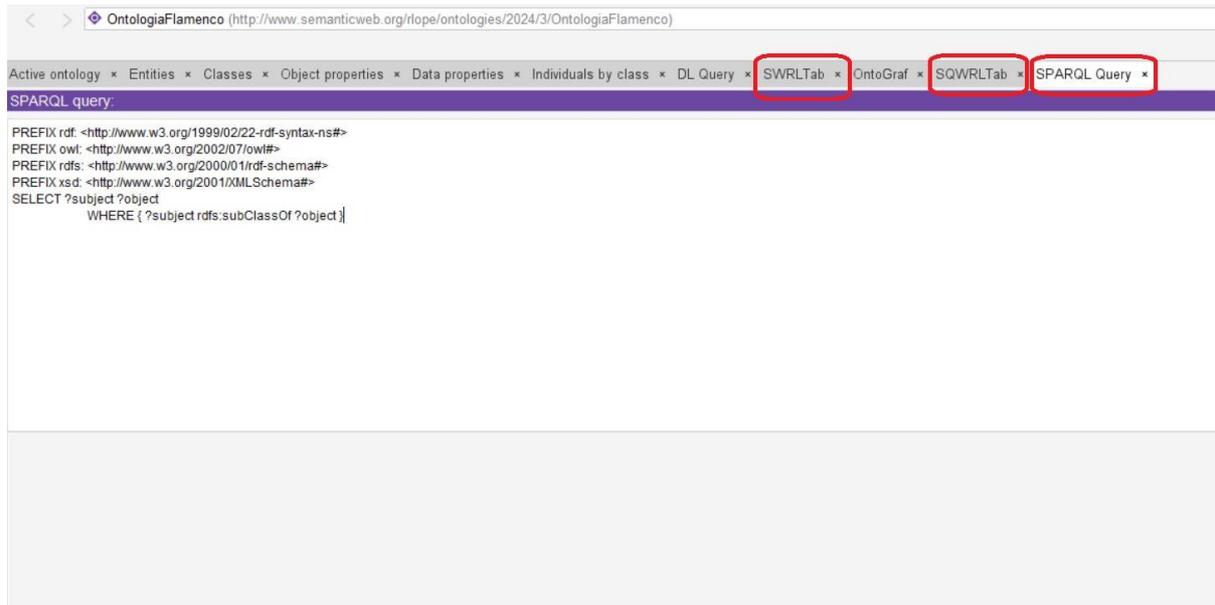


Figura 49. Pestañas Protégé

Las consultas de SPARQL se realizan sobre grafos RDF, utilizando patrones de tripletas para buscar y recuperar datos. Sin embargo, SQWRL permite realizar consultas que involucran razonamiento basado en reglas.

5.2.1. Consultas y reglas de inferencia con SWRL y SQWRL

Se han incluido las reglas de inferencia en este apartado debido a la estrecha dependencia que tiene SWRL de SQWRL.

A través de la funcionalidad que plantea Protégé una serie de cuestiones SQWRL

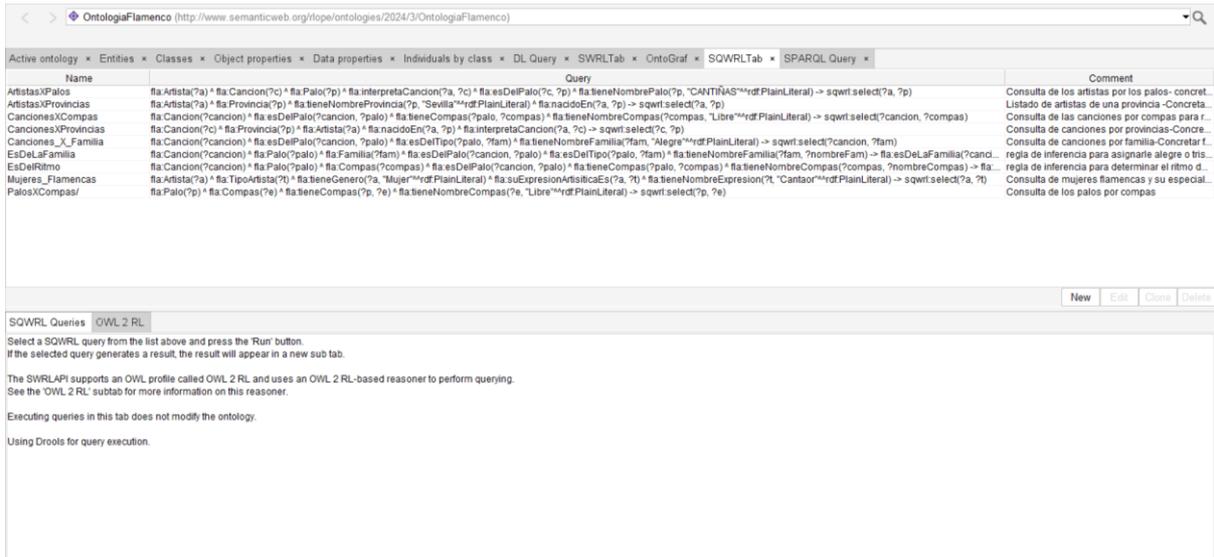


Figura 50. Cuestiones SQWRL

Cada una de ellas se realiza individualmente y se ejecutan en el mismo protégé pulsando el botón “RUN”.

A continuación se muestran algunos ejemplos:

Consulta 1: En esta consulta se solicitan los artistas que han interpretado alguna canción con el palo que se determine.

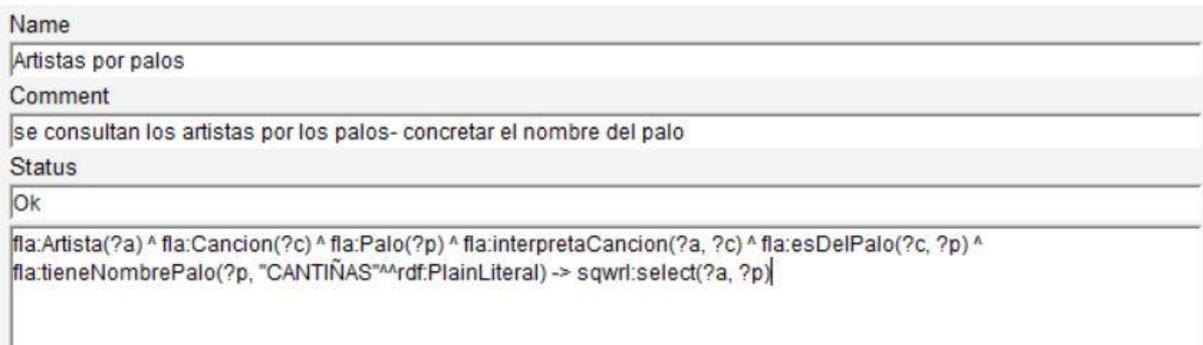


Figura 51. SQWRL Artistas por palo

Cuyo resultado será:

SQWRL Queries	OWL 2 RL	Artistas por palos
		a
fla:AntonioMairena		fla:CANTIÑAS
fla:ChanoLobato		fla:CANTIÑAS

Figura 52. SQWRL Resultado Artistas por palo

Consulta 2: En esta consulta se solicitan los artistas que han nacido en la provincia que se especifique.

Name	Artistas por provincias
Comment	Listado de artistas de una provincia -Concretar el nombre de la provincia
Status	Ok
	fla:Artista(?a) ^ fla:Provincia(?p) ^ fla:tieneNombreProvincia(?p, "Sevilla"^^rdf:PlainLiteral) ^ fla:nacidoEn(?a, ?p) -> sqwrl:select(?a, ?p)

Figura 53. SQWRL Artistas por provincia

Cuyo resultado será:

SQWRL Queries	OWL 2 RL	Artistas por provincias
		a
fla:PacoEIDePará		fla:Sevilla
fla:MaríaFernándezGranados		fla:Sevilla
fla:ManuelCuevas		fla:Sevilla
fla:EIVarilla		fla:Sevilla
fla:JuanaLaFeonga		fla:Sevilla
fla:EnriqueMontoya		fla:Sevilla
fla:ManuelPavón		fla:Sevilla
fla:IsraelGalván		fla:Sevilla
fla:JuanHabichuela		fla:Sevilla
fla:FranciscoCarrascoSoto		fla:Sevilla
fla:PastoraGalván		fla:Sevilla
fla:GasparDeUtrera		fla:Sevilla
fla:ManuelDíazFernández		fla:Sevilla
fla:EITupé		fla:Sevilla
fla:PastoraSoler		fla:Sevilla
fla:PacoJarana		fla:Sevilla
fla:JuanCarmona		fla:Sevilla

Figura 54. SQWRL Resultado Artistas por provincia

Consulta 3: En esta consulta se solicitan las canciones interpretadas por artistas que han nacido en la provincia que se especifique. .

Name	Canciones por provincias
Comment	Consulta de canciones por provincias-Concretar nombre provincia
Status	Ok
	fla:Cancion(?c) ^ fla:Provincia(?p) ^ fla:Artista(?a) ^ fla:nacidoEn(?a, ?p) ^ fla:interpretaCancion(?a, ?c) ^ -> sqwrl:select(?c, ?p)

Figura 55. SQWRL Canciones por provincia

Cuyo resultado será:

SQWRL Queries	OWL 2 RL	Canciones por provincias
		c
fla:A_PURO_GOLPE_Y_FUEGO		fla:Sevilla
autogen:p26		fla:Sevilla
fla:MADURALO		fla:Sevilla
fla:EL_HUERFANITO		fla:Sevilla
fla:TE_HE_QUERIO_TANTO		fla:Cádiz
fla:YO_HE_VISTO_VARIOS_PINTORES		fla:Sevilla
fla:Y_A_DIOS_CON_FE_YO_LE_PIDO		fla:Sevilla
fla:CON_SAN_ANTONIO		fla:Sevilla
fla:QUE_ALEGRIA_MAS_GRANDE		fla:Sevilla
fla:HASTA_EL_ALMA_ME_DUELE		fla:Cádiz
fla:ME_DA_IGUAL_LA_LUZ_DEL_DIA		fla:Sevilla
fla:PRIMA_DE_MI_ALMA		fla:Huelva
fla:DEJA_QUE_LA_GENTE_HABLE		fla:Sevilla
fla:MARIA_DOLORES		fla:Sevilla
fla:MENTIRA_ES_EL_MUNDO		fla:Sevilla
fla:MORENA_ERA_SU_CARA		fla:Cádiz
fla:PILATOS_POR_NO_DEJAR		fla:Sevilla

Figura 56. SQWRL Resultado Canciones por provincia

Consulta 4: En esta consulta se solicitan las mujeres que ejecutan la especialidad artística que se especifica.

Name	Mujeres Flamencas
Comment	Consulta de mujeres flamencas y su especialidad-Concretar Expresión artística
Status	Ok
	<pre>fla:Artista(?a) ^ fla:TipoArtista(?t) ^ fla:tieneGenero(?a, "Mujer"^^rdf:PlainLiteral) ^ fla:suExpresionArtisiticaEs(?a, ?t) ^ fla:tieneNombreExpresion(?t, "Cantaor") -> sqwrl:select(?a, ?t)</pre>

Figura 57. SQWRL Mujeres flamencas

Cuyo resultado será:

SQWRL Queries	OWL 2 RL	Mujeres Flamencas
a		
fla:LaPompi		fla:Cantaor
fla:LolitaValderrama		fla:Cantaor
fla:DoloresAgujeta		fla:Cantaor
fla:MaríaLaMica		fla:Cantaor
fla:MaríaToledo		fla:Cantaor
fla:PepaDeBenito		fla:Cantaor
fla:EstherMerino		fla:Cantaor
fla:LaRempompa		fla:Cantaor
fla:PauleraDeJerez		fla:Cantaor
fla:DoloresDeLaHuerta		fla:Cantaor
fla:LaTobala		fla:Cantaor
fla:DoloresVargas		fla:Cantaor
fla:NiñaPastori		fla:Cantaor
fla:LaPerla		fla:Cantaor
fla:PastoraSoler		fla:Cantaor
fla:CancanillaDeMarbella		fla:Cantaor
fla:Guadiana		fla:Cantaor
Save as CSV...		

Figura 58. SQWRL Resultado Mujeres flamencas

Consulta 5: En esta consulta se solicitan los palos que tienen el compás que se especifica.

Name	palos por compas
Comment	consulta de los palos por compas
Status	Ok
	<pre>fla:Palo(?p) ^ fla:Compas(?e) ^ fla:tieneCompas(?p, ?e) ^ fla:tieneNombreCompas(?e, "Libre"^^rdf:PlainLiteral) -> sqwrl:select(?p, ?e)</pre>

Figura 59. SQWRL Palos por compás

Cuyo resultado será:

SQWRL Queries	OWL 2 RL	palos por compas
p		
fla:LEVANTICA		fla:Libre
fla:GRANAINA		fla:Libre
fla:ZAMBRA		fla:Libre
fla:MEDIAGRANAINA		fla:Libre
fla:MIRABRAS		fla:Libre
fla:NANA		fla:Libre
fla:SAETAS		fla:Libre
fla:GILIANA		fla:Libre
fla:MALAGUEÑAS		fla:Libre
fla:MINERAS		fla:Libre
fla:CARTAGENERAS		fla:Libre
fla:TONÁ		fla:Libre
fla:TARANTAS		fla:Libre
fla:ROMANCE		fla:Libre
fla:ZAPATEADO		fla:Libre
fla:FANDANGOS		fla:Libre
fla:MURCIANAS		fla:Libre

[Save as CSV...](#)

Figura 60. SQWRL Resultado Palos por compás

A continuación se especifican dos reglas de inferencia (SWRL) que están previamente comprobadas como consultas de SQWRL.

Regla 1: Con esta regla se consigue que las canciones tengan su familia (alegre o triste) como propiedad en función del palo a la que pertenezcan.

Edit
×

Name

EsDeLaFamilia

Comment

regla de inferencia para asignarle alegre o triste a las canciones

Status

Ok

`fla:Cancion(?cancion) ^ fla:Palo(?palo) ^ fla:Familia(?fam) ^ fla:esDelPalo(?cancion, ?palo) ^ fla:esDelTipo(?palo, ?fam) ^ fla:tieneNombreFamilia(?fam, ?nombreFam) -> fla:esDeLaFamilia(?cancion, ?nombreFam)`

Cancel
Ok

Figura 61. Regla SWRL EsDeLaFamilia

El resultado, al ejecutarla siguiendo el orden de los botones “OWL+SWRL-> drools” “Run Drools” y “Drools->OWL”, se verá plasmado afectando a todos los elementos de la ontología como sigue:

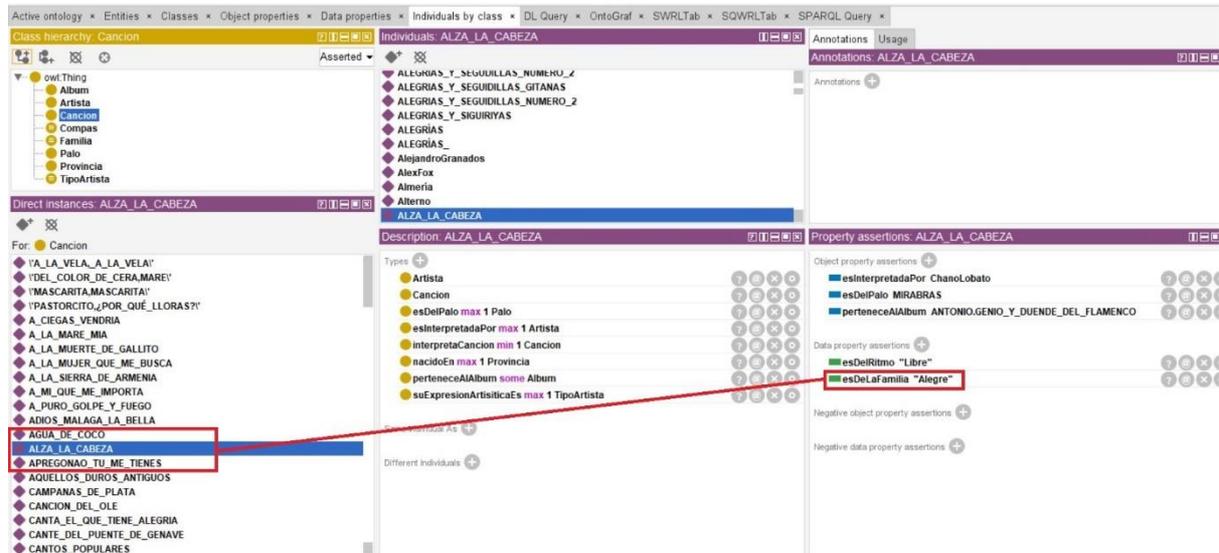


Figura 62. Propiedad esDeLaFamilia

Regla 2: Con esta regla se consigue que las canciones tengan su compás como propiedad en función del palo a la que pertenezcan.

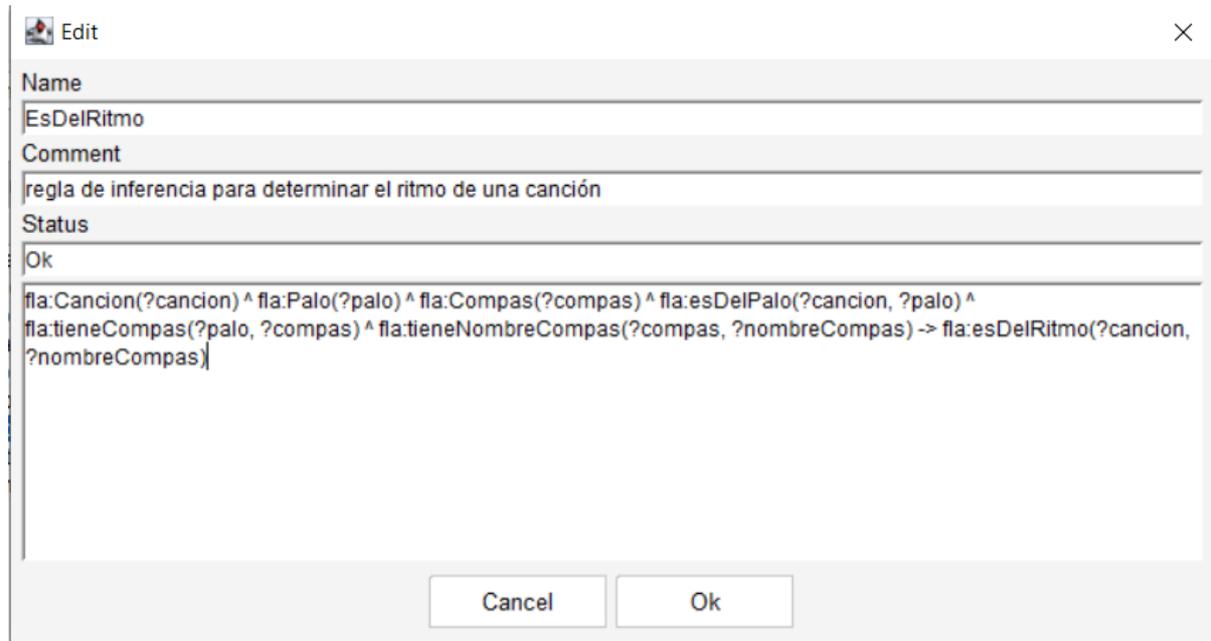


Figura 63. Regla SWRL esDelRitmo

El resultado, al ejecutarla siguiendo el orden de los botones “OWL+SWRL-> drools” “Run Drools” y “Drools->OWL”, se verá plasmado afectando a todos los elementos de la ontología como sigue:

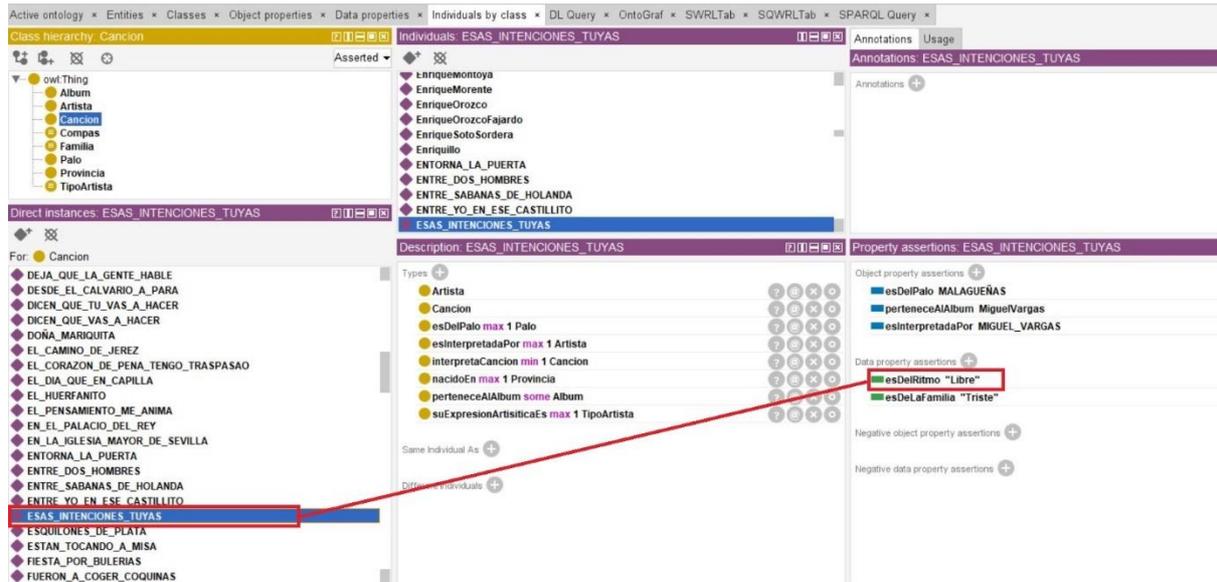


Figura 64. Propiedad esDelRitmo

5.2.2. Consultas SPARQL

Una vez ya se ha observado que las consultas en SQWRL y que las reglas de inferencia SWRL funcionan correctamente se procede a probar con un par de ejemplos en SPARQL:

Consulta 1: Se consultan los palos por sus compases.

```

SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX F: <http://www.semanticweb.org/riope/ontologies/2024/3/OntologiaFlamenco#>
SELECT ?Palo ?Compas
  WHERE{?Palo F:tieneCompas ?Compas}
ORDER BY ?Palo
    
```

Figura 65. SPARQL Palos por compases.

Dando como resultado:

Palo	Compas
ALBOREÁ	Binario
ALEGRÍAS	Alterno
BAMBERAS	Ternario
BANDOLÁ	Ternario
BULERÍAS	Alterno
CABALES	Binario
CAMPANILLEROS	Ternario
CANTIÑAS	Ternario
CARACOLES	Ternario
CARCELERAS	Alterno
CARTAGENERAS	Libre
CAÑA	Alterno
COLOMBIANAS	Binario
FANDANGOS	Libre
FARRUJICA	Rinario

Figura 66. SPARQL Resultado Palos por compases.

Consulta 2: Se consultan canciones por artistas.

```
SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX F: <http://www.semanticweb.org/rlope/ontologies/2024/3/OntologiaFlamenco#>
SELECT ?Cancion ?Artista
  WHERE{?Cancion F:esInterpretadaPor ?Artista.
}
ORDER BY ?Artista
```

Figura 67. SPARQL Canciones por artistas.

Dando como resultado:

Cancion	Artista
HASTA_EL_ALMA_ME_DUELE	AntoniaSantiago
LA_MANO_EN_EL_EVANGELIO	AntoniaSantiago
EL_DIA_QUE_EN_CAPILLA	AntonioMairena
EL_CAMINO_DE_JEREZ	AntonioMairena
A_LA_MUJER_QUE_ME_BUSCA	AntonioMairena
ME_HUELE_A_FLOR_DE_ROMERO	AntonioMairena
MAL_FIN_TENGA_ESTO_CIVIL	AntonioMairena
YO_NO_TE_OBLIGO_GITANA	AntonioMairena
ENTRE_DOS_HOMBRES	AntonioMairena
POR_TU_CAUSA_YO_ME_VEO	AntonioMairena
DE_QUE_SIEMPRE_TE_QUERRIA	AntonioMairena
QUE_TU_NO_ME_QUIERAS	AntonioMairena
VIVIR_CON_UNA_SERRANA	AntonioMairena
SI_ESTA_PENA_MIA	AntonioMairena
ME DA MIFDO DE LA LLUNA	AntonioMairena

Figura 68. SPARQL Resultado Canciones por artistas.

6. Conclusiones y líneas de futuro

Este trabajo de fin de grado ha sido el resultado de una amplia labor de investigación y de aplicar posteriormente los conceptos analizados.

Al comienzo del mismo mis conocimientos respecto a la Web Semántica eran bastante escasos. Por ello, me he visto obligado a invertir una buena cantidad de tiempo probando y entendiendo el uso de diferentes herramientas, al igual que repensando y rehaciendo el modelo una y otra vez.

La idea de adaptar algo tan libre y con un origen tan popular como es el Flamenco a una posible clasificación semántica en el mundo digital me ha provocado una incertidumbre permanente de si transitaba por el camino correcto o no.

La ausencia de información referente a esta temática en el mundo de la Web semántica ha supuesto una inversión de muchas horas de investigación para conseguir dar forma a un modelo que pudiera encajar. A pesar de que mi tutor me advirtiera de su posible complejidad, decidí optar por seguir adelante con la idea.

Finalmente se ha desarrollado una ontología que podría considerarse pionera, o al menos como un posible punto de partida para poder ser ampliada o reutilizada por otras futuras que quieran modelar el mundo flamenco .

A nivel técnico ha supuesto un reto debido a la falta de costumbre a la hora de manejar entidades en la forma que lo hace Protégé. La importación de datos ha sido la tarea que más complicaciones me ha generado debido a que se ha planteado e intentado el uso de diferentes tecnologías.

El primer planteamiento de la planificación no ha podido cumplirse al 100% debido a esta serie de dificultades que me he ido encontrando por el camino. He aprendido que, a la hora de establecer una distribución de tareas en el tiempo, es conveniente dar cierto margen a algunas de ellas. Aprender a detectar cuáles podrían suponer un problema o cuales requerirán más tiempo del estimado previamente es todo un arte. Estoy convencido que con experiencia y conocimiento del área se termina adquiriendo.

En cuanto al resultado obtenido, he de reconocer que mi idea inicial era la de crear una ontología un tanto más completa y, por tanto, más compleja. Pero los datos disponibles desde los que se ha partido y las publicaciones a las que he acudido no daban más margen a la ampliación del modelo.

6.1. Líneas de futuro

En cuanto al futuro del uso de la web semántica es posible que su evolución esté pasando por el momento en el que se encuentra actualmente la Inteligencia Artificial. Una posible adaptación podría ser la combinación de un modelo de Generación Aumentada de Recuperación (RAG, por sus siglas en inglés) con un Modelo de Lenguaje Grande (LLM) como pueda ser ChatGPT. Esto implica que

tras un entrenamiento previo y un buen conjunto de datos se puedan autogenerar respuestas de una calidad extraordinaria.

Bibliografía

b1: T. Berners-Lee and M. Fischetti, The Semantic Web. American Scientific. San Francisco: HarperSanFrancisco, 1999. <https://www.scientificamerican.com/article/the-semantic-web/>

b2: desarrollo explicativo de las ontologías y la web semántica.

<https://www.carlosgonzalo.es/ontologias-y-la-web-semantica/> (Marzo 2024)

b3: Ontologías y Web Semántica. Jordi Durán Cals, Jordi Conesa i Caralt y Robert Clarisó Viladrosa. UOC. (PDF)

b4: Pulpitarrita con pasaporte flamenco. Silvia Marin. Editorial FlamencoVive.

b5: Investigación y Flamenco J.M. Díaz-Báñez, F.J. Escobar-Borrego (Eds.) , Signatura Ediciones, ISBN 978-8496210-90-5, 2011.

b6: Palos y el estilo Flamenco. Felipe Gertrudix. <https://www.bubok.es/libros/9562/palos-y-estilos-del-flamenco>

b7: Modelo web semántico para contactar grupos musicales en la ciudad de Bogotá

<http://portal.amelica.org/ameli/journal/536/5362548005/html/> (Abril 2024)

b8: Performed Music Ontology <http://performedmusicontology.org/> (Abril 2024)

b9: Music Representation Ontology. <https://github.com/polifonia-project/music-representation-ontology> (Abril 2024)

b10: Flamencopolis.(Web). <https://flamenco.plus/flamencopolis/index.php?> (Abril 2024)

b11: Protégé foro. <https://protege-user.stanford.narkive.com/OV96Tj4b/csv-files-in-jena-ontology-api> (Mayo 2024)

Anexo A Glosario

GLO1: WWW: World Wide Web

GLO2: W3C: World Wide Web Consortium

GLO3: HTML: Lenguaje de Marcas de Hipertexto

GLO4: IA: Inteligencia Artificial

GLO5: UNICODE: Unicode: Conjunto de caracteres estándar que permite expresar texto en cualquier alfabeto, incluidos alfabetos no latinos, como por ejemplo el árabe, el cirílico o el chino.

GLO6: XML: Lenguaje de Marcado Extendido

GLO7: UTF: Formato de transformación de UNICODE. (8, 16 o 32)

GLO8: WS: Web Semántica

GLO9: XML Schema: Esquema XML

GLO10: URI: Identificador de Recursos Unificado

GLO11: IRI: Identificador de Recursos Internacionalizado

GLO12: URL: Localizador Uniforme de Recursos

GLO13: URN: Nombre Uniforme de Recurso

GLO14: RDF: Marco de Descripción de Recursos

GLO15: RDF-S: Esquema RDF

GLO16: COFLA: COFLA II: análisis Computacional de la música Flamenca

GLO17: DataSet: Conjunto de datos.

GLO18: FLABASE: Dataset de conocimiento flamenco

GLO19: JSON: formato ligero de intercambio de datos

GLO20: BIBFRAME: Modelo para expresar y conectar datos bibliográficos.

GLO21: SWRL: Semantic Web Rule Language

GLO22: SQLWRL: Extensión de SWRL que permite realizar consultas con inferencia lógica

GLO23: Protégé: Aplicación para desarrollar ontologías

GLO24: Wikipedia: Portal enciclopedia web.

GLO25: CSV: archivo de texto plano separado por comas

GLO26: Script: conjunto de instrucciones en lenguaje de programación o scripting.

GLO27: Python: Lenguaje de programación de alto nivel.

GLO28: Excel: Programa de office de Microsoft para hojas de cálculo

GLO29: XLSX: Extensión de los archivos de Microsoft Excel

GLO30: Plugin: Software adicional a una aplicación

GLO31: Cellfie: Plugin de Protégé para importación de datos.

GLO32: Manhattan: Lenguaje enfocado al análisis de datos.

GLO33: Ontograf: Herramienta visual de esquemas de ontologías de Protégé

GLO34: SWRL: Semantic Web Rule Language. Extensión del lenguaje OWL para crear reglas de inferencia

GLO35: Jena: Librería de Java para procesado de datos en formato RDF.

GLO36: Eclipse: IDE para Java

GLO37: Java: Lenguaje de programación.

GLO38: ROBOT: Aplicación de comandos para ontologías

GLO39: SPARQL: Lenguaje de consultas de RDF

GLO40: SQSWRL: Extensión del lenguaje OWL para realizar consultas

GLO41: Panda: biblioteca de Python para leer archivos csv.

GLO42: rdflib: biblioteca de Python para trabajar con datos en formato RDF

GLO43: RAG: Generación Aumentada de Recuperación

GLO44: LLM: Modelo de Lenguaje Grande

GLO45: ChatGPT: Aplicación de IA