UNIVERSIDAD DE CASTILLA-LA MANCHA
ESCUELA SUPERIOR DE INFORMÁTICA

MÁSTER UNIVERSITARIO
EN INGENIERÍA INFORMÁTICA

TRABAJO FIN DE MÁSTER

Trabajo Fin de Máster

Raúl Reguillo Carmona

Octubre, 2017

TRABAJO FIN DE MÁSTER

UNIVERSIDAD DE CASTILLA-LA MANCHA

ESCUELA SUPERIOR DE INFORMÁTICA

TRABAJO FIN DE MÁSTER

Trabajo Fin de Máster

Autor: Raúl Reguillo Carmona

Tutor: Ismael Caballero Muñoz-Reja

Cotutor: Bibiano Rivas García

Octubre, 2017

UNIVERSIDAD DE CASTILLA-LA MANCHA

ESCUELA SUPERIOR DE INFORMÁTICA

TRABAJO FIN DE MÁSTER

Trabajo Fin de Máster

Fdo.: Raúl Reguillo Carmona Fdo.: Ismael Caballero Muñoz-Reja

Octubre, 2017

Raúl Reguillo Carmona

Ciudad Real – Spain

E-mail: raul.reguillo@alu.uclm.es
Teléfono: 638 175 608

TRIBUNAL:

Presidente:

Vocal:

Secretario:

FECHA DE DEFENSA:

CALIFICACIÓN:

PRESIDENTE	VOCAL	SECRETARIO
Fdo.:	Fdo.:	Fdo.:

Resumen

El presente documento es un ejemplo de memoria del Trabajo Fin de Máster según el formato y criterios de la Escuela Superior de Informática de Ciudad Real. La intención es que este texto sirva además como una serie de consejos sobre tipografía, LaTeX, redacción y estructura de la memoria que podrían resultar de ayuda. Por este motivo, se aconseja al lector consultar también el código fuente de este documento.

Este documento utiliza la clase LaTeX *esi-tfm*, disponible como paquete Debian/Ubuntu, consulta:

`https://bitbucket.org/arco_group/esi-tfg`.

Si encuentra cualquier error o tiene alguna sugerencia, por favor, utilice el *issue tracker* del proyecto *esi-tfm* en:

`https://bitbucket.org/arco_group/esi-tfg/issues`

El resumen debería estar formado por dos o tres párrafos resaltando lo más destacable del documento. No es una introducción al problema, es decir, debería incluir los logros más importantes del proyecto. Suele ser más sencillo escribirlo cuando la memoria está prácticamente terminada. Debería caber en esta página (es decir, esta cara).

Abstract

English version of the previous page.

Listado de acrónimos

AAA	*Anyone can say Anything about Any topic*
API	Application Programming Interface
CSV	Comma Separated Values
DQ	Calidad de Datos
DQD	Dimensión de Calidad de Datos
DOAP	Description Of A Project
ER	Entidad-Interrelación
FOAF	Friend Of A Friend
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ISO	International Organization for Standarization
LD	Datos Enlazados
LOD	Datos Enlazados Abiertos
OD	Datos Abiertos
OWL	Lenguaje de Ontología Web
PFC	Proyecto Fin de Carrera
RDF	Framework de Descripción de Recursos
RQL	Relationship Query Language
SGML	Standard Generalized Markup Language
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
TDB	Triple Data Base
URI	Identificador Uniforme de Recurso
URL	Localizador Uniforme de Recurso
W3C	Consorcio para la World Wide Web
XML	Lenguaje de Marcas Extensible

Agradecimientos

Escribe aquí algunos chascarrillos simpáticos. Haz buen uso de todos tus recursos literarios porque probablemente será la única página que lean tus amigos y familiares. Debería caber en esta página (esta cara de la hoja).

Juan[1]

[1] Sí, los agradecimientos se firman

A alguien querido y/o respetado

Índice general

Índice de cuadros

Índice de figuras

Índice de listados

Listado de acrónimos

AAA	*Anyone can say Anything about Any topic*
API	Application Programming Interface
CSV	Comma Separated Values
DQ	Calidad de Datos
DQD	Dimensión de Calidad de Datos
DOAP	Description Of A Project
ER	Entidad-Interrelación
FOAF	Friend Of A Friend
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
ISO	International Organization for Standarization
LD	Datos Enlazados
LOD	Datos Enlazados Abiertos
OD	Datos Abiertos
OWL	Lenguaje de Ontología Web
PFC	Proyecto Fin de Carrera
RDF	Framework de Descripción de Recursos
RQL	Relationship Query Language
SGML	Standard Generalized Markup Language
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
TDB	Triple Data Base
URI	Identificador Uniforme de Recurso
URL	Localizador Uniforme de Recurso
W3C	Consorcio para la World Wide Web
XML	Lenguaje de Marcas Extensible

Capítulo 1

Introducción

ESTO se llama «letra capital» y debería utilizarse únicamente al comienzo de capítulo como artificio decorativo. Para que resulte estéticamente adecuada, este primer párrafo debería tener más del doble de líneas de lo que ocupe verticalmente la letra capital (dos en este caso). El capítulo de introducción debe dar al lector una perspectiva básica —pero completa— del problema que se pretende abordar, pero también de la estrategia y enfoque que el autor propone para su resolución. El lector debería poder determinar si este documento le interesa leyendo únicamente la introducción.

A modo de referencia, este capítulo tendrá una longitud aproximada de entre 3 y 10 páginas.

Título del proyecto

En la portada —y otras páginas de presentación— el nombre o título del proyecto debe aparecer sin comillas, cursiva u otros formatos. Si se cita el título de otra obra, o el nombre de un capítulo sí debe aparecer entre comillas. Por cierto, las comillas que deben usarse en castellano son las «latinas», dejando las "inglesas" para los raros casos en los que aparezca una cita en el cuerpo otra [Mar08].

Contexto del proyecto

Evolución respecto al PFC

Competencias adquiridas

El estudiante deberá enumerar y detallar, en un apartado específico, en qué forma se han alcanzado las competencias específicas del proyecto (en base a las indicadas en la propuesta inicial).

Estructura del documento

Pueden incluirse aquí una sección con algunos consejos para la lectura del documento dependiendo de la motivación o conocimientos del lector. También puede ser útil incluir una lista con el nombre y finalidad de cada uno de los capítulos restantes.

Capítulo 2: Objetivos

Finalidad y justificación (con todo detalle) del presente documento.

Capítulo ??: ??

Explica herramientas y aspectos básicos de edición con LaTeX.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

- First itemtext
- Second itemtext
- Last itemtext
- First itemtext

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{i=n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis

urna dictum turpis accumsan semper.

$$\int_0^\infty e^{-\alpha x^2}\mathrm{d}x = \frac{1}{2}\sqrt{\int_{-\infty}^{\infty} e^{-\alpha x^2}\mathrm{d}x \int_{-\infty}^{\infty} e^{-\alpha y^2}\mathrm{d}y} = \frac{1}{2}\sqrt{\frac{\pi}{\alpha}}$$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$\sum_{k=0}^{\infty} a_0 q^k = \lim_{n\to\infty} \sum_{k=0}^{n} a_0 q^k = \lim_{n\to\infty} a_0 \frac{1-q^{n+1}}{1-q} = \frac{a_0}{1-q}$$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-p \pm \sqrt{p^2 - 4q}}{2}$$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2} = \frac{1}{c^2}\frac{\partial^2 \Phi}{\partial t^2}$$

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Capítulo 2

Objetivos

De acuerdo a la Introducción, el alumno deberá especificar cuál(es) es(son) la(s) hipótesis de trabajo de la(s) que se parte(n), qué se pretende(n) resolver en el presente TFM.

Es importante formular con claridad cuál es el objetivo general y el alcance correspondiente a las hipótesis de trabajo. Del mismo modo, se deberán establecer los objetivos parciales derivados del objetivo general y los resultados esperados.

Como preámbulo a la formulación de los objetivos parciales, el alumno deberá discutir sobre las limitaciones y condicionantes a tener en cuenta en el desarrollo del TFM (lenguaje de desarrollo, equipos, madurez de la tecnología, etcétera).

A modo de referencia, este capítulo tendrá una longitud aproximada de entre 2 y 10 páginas.

Objetivo general

El hito final que se pretende lograr, destacando el dominio concreto estudiado, el problema específico que resuelve y/o la funcionalidad que se aporta en el presente trabajo.

Objetivos específicos

Los objetivos específicos son las partes independientes del proyecto que tienen valor por sí mismas.

Por ejemplo, si el objetivo general fuera destruir una flota enemiga, los objetivos específicos podrían ser: hundir el portaaviones, inutilizar las torretas de los destructores, eliminar los cazas enemigos, etc.

Los objetivos específicos no son tareas; análisis, diseño, etc. no tienen valor intrínseco para el cliente, si por ejemplo el proyecto se cancela en la fase de diseño no se le entrega nada de valor al cliente, luego no se cubre ningún objetivo.

No se deben confundir los objetivos del proyecto con los objetivos del alumno. Indicar como objetivo que el alumno va a aprender o a estudiar determinada disciplina o herramienta no aporta nada al cliente. Deben ser entregables que el cliente puede valorar y por los que estaría dispuesto a pagar. Resumiendo, son **objetivos**, no subjetivos.

Objetivo 1
Objetivo 2
Objetivo 3
Limitaciones y condicionantes
Alcance
Equipo de trabajo
Marco tecnológico

Capítulo 3

Estado del Arte

En el presente capítulo se describen los conceptos teóricos que son necesarios para establecer las bases de la elaboración del PFC. La descripción incluye conceptos como Calidad de Datos (DQ), Datos Abiertos (OD), Datos Enlazados Abiertos (LOD), Web Semántica y Big Data. También se describe un conjunto de tecnologías que se han utilizado para el desarrollo del proyecto.

Web Semántica

En [BLHL01] se define la Web Semántica como una evolución o extensión de la Web tradicional en la que la información es dada mediante significados bien definidos, lo que facilita el procesamiento automático del contenido y permita a las personas y a los ordenadores trabajar en cooperación.

Este concepto ha ido evolucionando y refinándose con el paso del tiempo. En los siguientes apartados se introduce el concepto de Web Semántica y se explicarán sus principales características.

Concepto

El Consorcio para la World Wide Web (W3C) en [W3Ca] define la Web Semántica como:

Una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la Web de más significado y, por lo tanto, de más semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante.

[SBLH06] define a su vez la Web Semántica como sigue:

La Web Semántica es una Web de información procesable - información derivada de los

datos mediante una teoría semántica para la interpretación de símbolos. La teoría semántica proporciona una noción de "significado" en la que la conexión lógica de términos establece la interoperabilidad entre sistemas.

Por otro lado, [HBLM02] definen nuevamente la Web Semántica de la siguiente manera:

La Web Semántica es una extensión de la Web actual, en la que a la información disponible se le otorga un significado bien definido que permita a los ordenadores y a las personas trabajar en cooperación. Está basada en la idea de proporcionar en la Web datos bien definidos y enlazados, permitiendo que aplicaciones heterogéneas localicen, integren, razonen y reutilicen toda la información presente en la Web.

Por lo tanto se puede establecer que la Web Semántica es una extensión de la Web en la que se dota de capacidad de anotar información semántica a los datos de manera que proporcionen un significado. En la última definición aparece el concepto de "datos enlazados" como precursor de lo que posteriormente se considerará Linked Data (LD), con la idea de vincular datos mediante estándares de Web Semántica para alcanzar los siguientes objetivos:

1. Reutilizar entidades ya definidas en el modelo de Web Semántica.
2. Hacer de estas entidades conceptos únicos, evitando redundancia y ambigüedad en los datos.
3. Permitir la interoperabilidad semántica entre aplicaciones heterogéneas.

La reutilización es posible gracias al concepto de Identificador Uniforme de Recurso (URI) utilizado para identificar de forma unívoca, universal y expansible un espacio de nombres de recursos de información.

Respecto de la interoperabilidad semántica, [HT06] explica que la *Web Semántica es la solución al problema de la integración de datos* sin necesidad de llevar a cabo procesos de conversión, gracias a la utilización de un modelo de representación como Framework de Descripción de Recursos (RDF) y utilizando Lenguaje de Marcas Extensible (XML) como fuente sintáctica para su intercambio.

Terminología de Web Semántica

- **Identificador Uniforme de Recurso (URI)**: cadena de caracteres que identifica los recursos de una red de forma unívoca. La diferencia con una Localizador Uniforme de Recurso (URL), es que éstos pueden variar en el tiempo. En el ámbito de la Web Semántica, una URI identificará a un recurso de manera unívoca dentro del conjunto de datos.
- **Recurso**: Se dice que un recurso es cualquier concepto que se pueda identificar.
- **Tripla**: Usando el estándar Framework de Descripción de Recursos (RDF), se define

tripla como una asociación de dos recursos a través de una relación o propiedad. Esta asociación se representa mediante dos nodos conectados por un arco (véase figura 3.1) (sujeto, predicado y objeto), también llamado *sentencia* (statement):

  - **Sujeto**: es el recurso desde el que parte el arco (la propiedad).
  - **Predicado**: es la propiedad que etiqueta el arco.
  - **Objeto**: es el recurso o literal apuntado por el arco.

- **Endpoint**: Interfaz que se ofrece como extremo de una comunicación o como terminal que permite dar un servicio determinado. En el ámbito de la Web Semántica, un Endpoint permitirá tener acceso a las URI de un dataset, así como a las triplas mediante la utilización de protocolo HTTP.
- **Graph/Named Graph**: El término *Graph* se refiere a un conjunto de triplas relacionadas entre sí, de tal forma que tanto predicados como objetos dentro de una tripla hacen referencia a sujetos de otras triplas, enlazándose de esta manera. Un *Named Graph* es un conjunto de triplas que tiene sentido en sí misma (por ejemplo, las triplas que definen a un grupo de música en concreto).
- **Almacenamiento de triplas**: Es una base de datos especializada en almacenar archivos semánticos, es decir, conjuntos de triplas o *graphs*. Su funcionamiento interno dista del modo en que lo hacen las bases de datos convencionales, pues debe permitir inferencia gracias a las propiedades descritas.
- **Servidor de triplas**: Un servidor de triplas es una aplicación que, dado un almacenamiento de triplas, permite que ese contenido sea accesible por otras aplicaciones y mediante protocolos tales como HTTP. Además debe permitir operaciones de consulta, actualización, inserción y borrado sobre los datos almacenados.

Estándares

A continuación, se expondrán los estándares que definen la Web Semántica en el marco de la W3C.

Lenguaje de Marcas Extensible (XML)

En [XML] se define XML como un formato de texto simple, muy flexible, derivado de Standard Generalized Markup Language (SGML) (ISO 8879). Originalmente diseñado para cumplir con los desafíos de la publicación electrónica a gran escala, XML también está desempeñando un papel cada vez más importante en el intercambio de una amplia variedad de datos en la Web y otros sistemas.

XML establece las bases para la elaboración de lenguajes orientados a la representación de información estructurada mediante la descripción de gramáticas, permitiendo diferentes niveles de abstracción.

La principal ventaja de XML es que permite la intercomunicación de aplicaciones y la integración de información, también en el ámbito de las bases de datos.

XML constituye la base de otros lenguajes y en particular, fue precursor de RDF.

Framework de Descripción de Recursos (RDF)

Se define RDF en [RDF] como un modelo estándar para el intercambio de datos en la Web. RDF tiene características que facilitan la fusión de datos incluso si los esquemas subyacentes difieren y soporta específicamente la evolución de esquemas con el tiempo sin necesidad de realizar cambios en los consumidores de los datos.

RDF permite extender la estructura de los enlaces de la Web para hacer uso de las URI como nombre de las relaciones entre conceptos. De esta manera, se establecen lo que se conoce como *triplas* como una relación *(subjeto, predicado, objeto)* (véase figura 3.1 y Sección 3.1.2) en la que todos los componentes son URI.

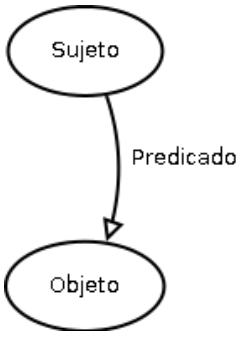

Figura 3.1: Esquema de tripla

Usando este modelo se permite la mezcla de datos estructurados y semi-estructurados a través de diferentes aplicaciones.

Esta estructura de enlaces forma grafos etiquetados y dirigidos donde los arcos representan el tipo de relación entre dos recursos, representados por nodos del grafo (véase figura 3.2).

Lenguaje de Ontología Web (OWL)

Según [OWL], se define OWL como un lenguaje de Web Semántica diseñado para representar conocimiento enriquecido y complejo acerca de conceptos, grupos de conceptos y relaciones entre conceptos. OWL es un lenguaje computacional basado en la lógica de manera que el conocimiento que expresa puede ser explotado por programas de computador.

Los documentos generados según el lenguaje OWL se conocen como **ontologías**. Las ontologías pueden ser publicadas en la W3C o pueden referise o derivarse de otras ontologías.

En el contexto de la computación y ciencias de la información, una ontología define un

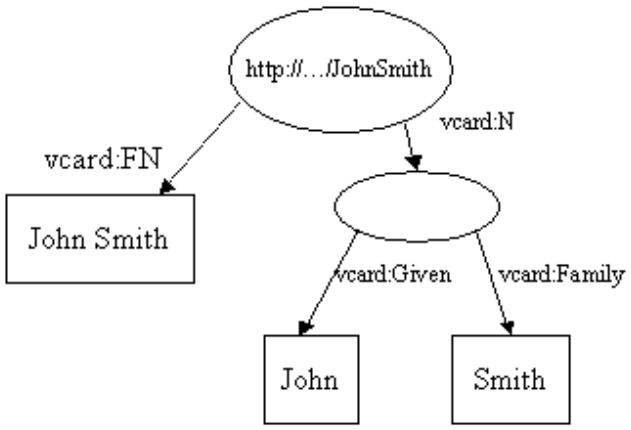

Figura 3.2: Grafo basado en triplas. Extraído de [JEN14]

conjunto de primitivas de representación con la que modelar un dominio de conocimiento [Gru]. La finalidad de las ontologías es ofrecer un modelo formal acerca de un conjunto de conceptos sobre el cual poder aplicar razonamiento automático .

De OWL derivan tres sub-lenguajes basados en la capacidad expresiva, tal y como cita [OWL]:

1. **OWL Lite**: que da soporte a las necesidades básicas del usuario cuando lo que se presente es una representación no tan exhaustiva, como por ejemplo, recursos, clasificaciones jerárquicas y restricciones simples.
2. **OWL DL (Description Logics)**: soportando más expresividad semántica y garantiza que todas las inferencias puedan ser calculadas en un tiempo finito.
3. **OWL Full**: el grado de expresividad es total, pero no asegura que las inferencias puedan ser calculadas en un tiempo finito.

SPARQL Protocol and RDF Query Language (SPARQL)

SPARQL es un lenguaje estándar para la consulta de grafos RDF. En [SPA] se puede encontrar toda la información y especificación de la tecnología.

Muy similar a Structured Query Language (SQL), es un lenguaje declarativo que permite estructurar las consultas como patrones de *triplas* sobre los cuales extraer instancias concretas.

Arquitectura

Tal y como se puede ver en las figuras 3.3 y 3.4, la arquitectura de la Web Semántica se fundamenta sobre los principios de la Web tradicional. Estos principios se pueden resumir:

1. Utilización de Localizador Uniforme de Recurso (URL) para la localización de recursos
2. Uso de HyperText Markup Language (HTML) para la elaboración de documentos, de modo que sea entendible por las personas y procesable por los computadores.
3. Uso de HyperText Transfer Protocol (HTTP) de comunicación cliente-servidor.

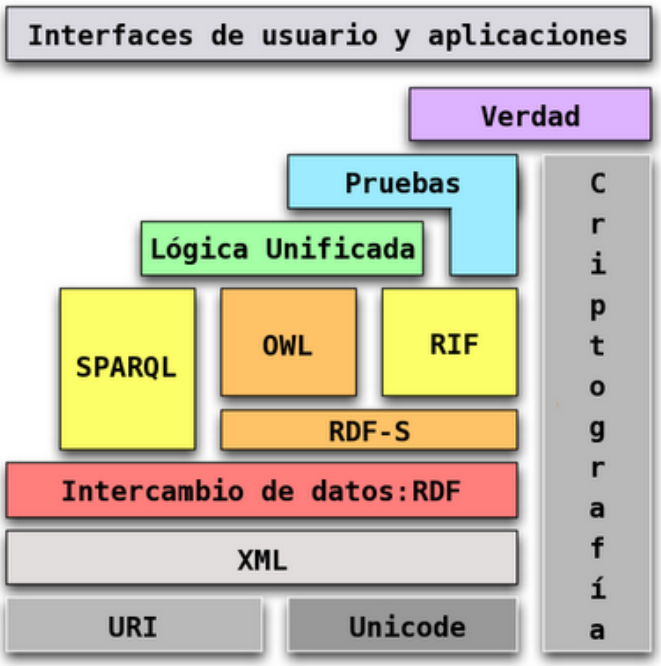

Figura 3.3: Tecnologías de Web Semántica

Las diferentes capas que se muestran en la figura 3.4 [San11] se pueden definir como :

- **Capa de localización y codificación**: el estándar para la codificación de caracteres es UNICODE y para la identificación y localización de recursos se utilizarán las URI o URL.
- **Capa de sintaxis**: estándares necesarios para la representación de información. Cobra especial importancia XML puesto que ofrece un formato de fácil procesamiento con una sintaxis jerárquica. Como derivación de XML surge XML *namespaces* que habilitará la aparición de diferentes vocabularios XML en un mismo documento. Esto permite la reutilización de recursos que previamente ya se hayan definido.
- **Capa de descripción y estructura**: el uso de RDF como estándar de representación de recursos, propiedades y relaciones, es el siguiente gran salto en la Web Semántica. Al tener una representación de la semántica formal, se consigue la interoperabilidad semántica entre aplicaciones heterogéneas.

- **Capa de integración lógica de ontologías y reglas de inferencia**: estrechamente vinculada con la capa de descripción y estructura, en esta capa se amplía la definición de clases, relaciones y propiedades entre recursos, utilizando para ello el lenguaje específico OWL.
- **Capa de consulta**: los recursos ya representados junto con sus relaciones ya conforman un hito. A continuación se requiere establecer unos mecanismos para búsqueda de triplas. SPARQL cumplirá la función de motor de búsqueda declarativo sobre archivos RDF.

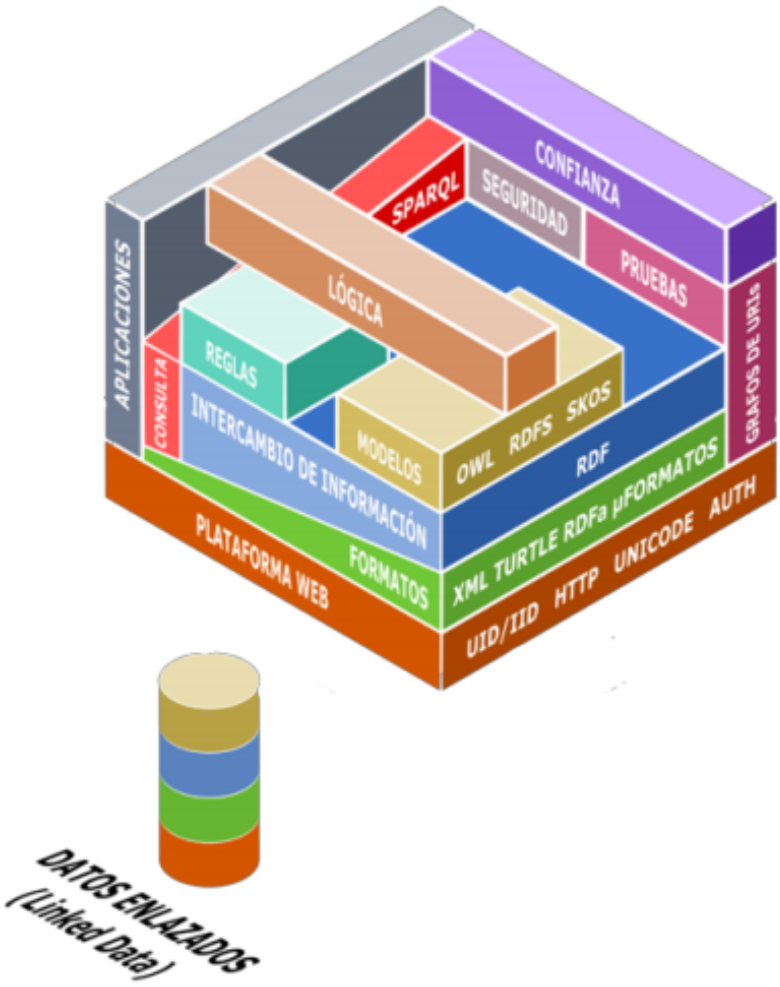

Figura 3.4: Arquitectura de la Web Semántica. Extraído de [San11]

Ontologías

Las ontologías pueden jugar un papel crucial para permitir el procesamiento del conocimiento, el intercambio y la reutilización basada en la Web entre aplicaciones. Las ontologías ofrecen una comprensión común de conceptos, con el fin de que la comunicación entre personas y aplicaciones sea homogénea [DMVH+00].

Definiciones de ontologías

[Gru] define las ontologías como *una especificación explícita de una conceptualización*, refiriéndose este término al modelo abstracto de una realidad concreta.

En [SBF98] se profundiza en los términos que engloba la definición de ontología:

Una ontología es una especificación formal y explícita de una conceptualización compartida, donde:

- el término "conceptualización" se refiere a un modelo abstracto de algún fenómeno en el mundo, identificando los conceptos relevantes del mismo,
- "explícita" porque los tipos de conceptos utilizados así como las constantes en su uso están explícitamente definidas,
- "formal" se refiere al hecho de que la ontología debe ser legible para las computadoras, lo que excluye al lenguaje natural, y
- "compartida" refleja la noción de que una ontología captura el conocimiento consensuado, es decir, no es privativo para ningún individuo, sino que es aceptado por un grupo.

Luego una ontología es una jerarquía que define una serie de clases, atributos y relaciones para describir un dominio sobre un concepto en concreto cuya finalidad es servir de herramienta para la representación del conocimiento.

Componentes de una ontología

En [San11] se enumeran los distintos componentes de las ontologías:

- **Clases**: conceptos generales acerca de un determinado dominio. La idea básica que se pretende formalizar.
- **Relaciones**: enlace entre conceptos (clases) de un mismo dominio.
- **Atributos**: representan la estructura del concepto.
    - **Funciones**: identifican un elemento mediante el cálculo de una función.
- **Instancias**: representa un individuo concreto perteneciente a una clase.
- **Axiomas**: expresiones siempre ciertas sobre relaciones.

Clasificación de las ontologías

A su vez, dependiendo del objetivo de la ontología, se pueden clasificar en distintos ámbitos tal y como propone [Gua98] (véase figura 3.5):

- Ontologías de alto nivel (*Top-level ontologies*): describen conceptos muy generales que son independientes de un problema particular, tales como espacio, tiempo, objeto, evento o acción.

- Ontologías de dominio y tarea (*Domain ontologies and task ontologies*): describen respectivamente, el vocabulario relacionado con un dominio genérico o con una tarea genérica o actividad, especializando los términos de la ontología de alto nivel.
- Ontologías de aplicación (*Application ontologies*): describen conceptos dependiendo tanto del dominio particular como de la tarea, es decir, es una especialización que generalmente particulariza *ambos* tipos de ontología. Generalmente corresponde con *roles* desempeñados por entidades de dominio mientras desarrollan alguna tarea.

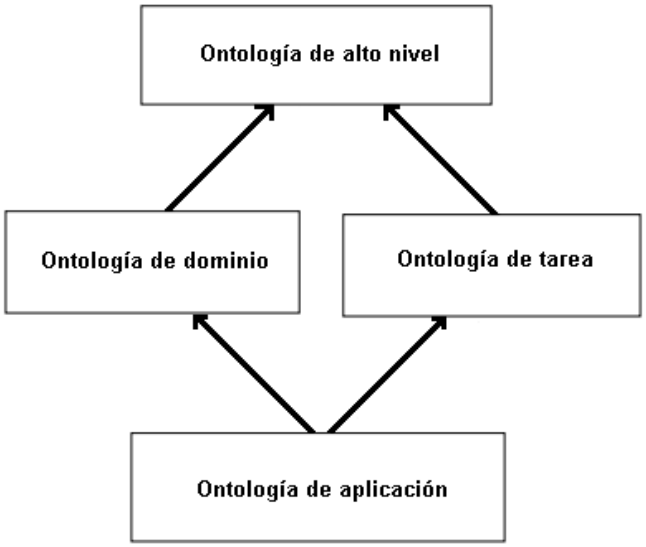

Figura 3.5: Clasificación de Ontologías. Extraído de [Gua98]

Metodología para desarrollo de una ontología en entornos de Web Semántica

Finalmente, [NMo01] propone una secuencia de tareas que todo desarrollo de cualquier ontología debería incluir:

1. Definir las clases en la ontología
2. Organizar dichas clases según una taxonomía
3. Definir propiedades de las clases y los valores que se asocian a esas propiedades
4. Completar los valores de las propiedades para cada una de las instancias.

Vocabularios

Como complemento al concepto de ontología, se describe a continuación el término vocabulario, frecuentemente usado en gran parte de la bibliografía.

Definición

En la Web Semántica, tal y como se explica en [W3Cb], los vocabularios definen los conceptos y relaciones usados para describir y representar un área de conocimiento y así clasificar los términos en una aplicación particular caracterizando relaciones y restricciones. Los vocabularios en la práctica pueden ser enormemente complejos o muy simples (llegando a describir únicamente uno o dos conceptos).

Según esta definición no queda claro en qué se diferencia un vocabulario de una ontología, pues realmente no existe una división clara entre estos dos conceptos. La tendencia es utilizar el término "ontología" para una colección de términos más compleja y formal, mientras que "vocabulario" se usa cuando la flexibilidad en el formalismo no implica necesariamente una pérdida de significado [W3Cb].

Por lo tanto, la función de un vocabulario es similar a la de una ontología en términos de objetivo final: establecer una representación formal de un determinado concepto.

Ejemplos de vocabulario

Existen numerosos vocabularios en uso actualmente. A continuación se citan algunos ejemplos:

Description Of A Project (DOAP)

El objetivo de este vocabulario es la descripción de proyectos software. Para ello, incluye toda la termonología referente al proyecto (licencia, versión de producto, dirección de repositorio, . . .).

Simple Knowledge Organization System (SKOS)

Este vocabulario tiene por objetivo la representación y estructuración de esquemas conceptuales tales como taxonomías, esquemas de clasificación o tesauros.

Friend Of A Friend (FOAF)

Quizá uno de los más extendidos sea Friend Of A Friend (FOAF) [FOA]. Su finalidad es describir personas, relaciones entre ellas así como aspectos de su actividad. Esta tecnología está en alza debido a su extensión en las redes sociales. FOAF a día de hoy es la base de un número considerable de esfuerzo en lo que se conoce como movimiento "open social", que tratan de facilitar al usuario integrar su propia información a través de aplicaciones sociales a través de la Web [AH08].

FOAF trabaja según el principio de *Anyone can say Anything about Any topic* (AAA). En el caso de FOAF los temas usualmente son otras personas [AH08].

En la figura 3.6 y en el listado 3.1 pueden verse ejemplos de uso de este vocabulario. En la figura se aprecian dos recursos (`foaf:ian` y `foaf:mary`), dos propiedades (`foaf:knows` y `foaf:firstName`) y un literal ("`Mary`"). En el listado 3.1 se expone un documento describiendo algunas entidades y relaciones.

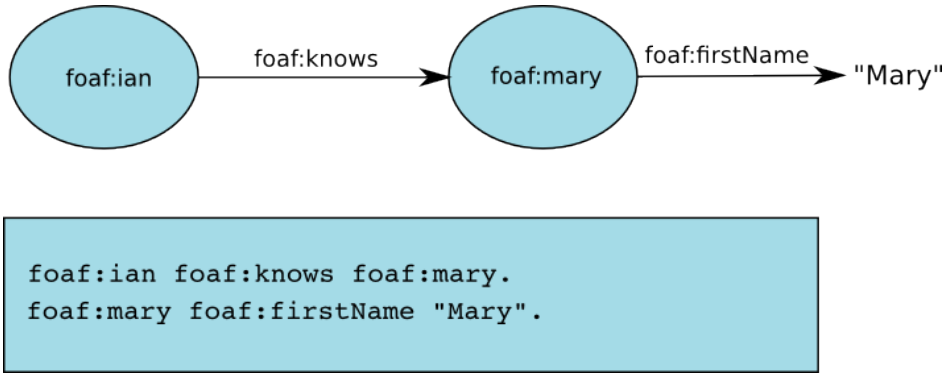

Figura 3.6: Ejemplo de FOAF. Extraído de [JEN14]

```
<rdf:RDF
  xmlns:rdf=''http://www.w3.org/1999/02/22-rdf-syntax-ns#''
  xmlns:rdfs=''http://www.w3.org/200/01/rdf-schema#''
  xmlns:foaf=''http://xmlns.com/foaf/0.1/''>

  <foaf:Person>
    <foaf:name>Raul Reguillo Carmona</foaf:name>
    <foaf:title>Mr</foaf:title>
    <foaf:nick>Radulfr</foaf:nick>
    <foaf:weblog rdf:resource=''http://geeklandtryp.blogspot.com.es />
    <foaf:knows>
      <foaf:Person>
        <foaf:name>Ismael Caballero</foaf:name>
      </foaf:Person>
    </foaf:knows>
  </foaf:Person>
</rdf:RDF>
```

Listado 3.1: Ejemplo de FOAF

Frameworks para el desarrollo de aplicaciónes de Web Semántica

En [W3Cc] se puede consultar una lista con todos los frameworks de Web Semántica disponibles hasta la fecha. En este apartado, se van a nombrar solamente algunos de ellos por ser especialmente significativos.

Apache Jena

Jena es un framework para la construcción de aplicaciones que utilizan tecnologías semánticas y Linked Data [JEN14].

En la arquitectura de Jena (véase figura 3.7) se pueden encontrar tres bloques diferenciados en función de las herramientas que engloban:

- **RDF**
  - Jena ofrece una API para el tratamiento con grafos RDF, serializando las triplas y tratando con ellas de manera ágil y eficiente.

- Paralelamente ofrece un motor SPARQL para las consultas sobre los archivos semánticos.

- **Triple Store**
  - Para hacer persistentes los datos, Jena ofrece Triple Data Base (TDB): una base de datos de triplas nativa de alto rendimiento y alineada con el resto de tecnologías de Jena.
  - Por otra parte, Jena ofrece **Fuseki** como *endpoint*, es decir, como servidor de triplas accesible a través de HTTP, integrándose a la perfección con TDB.

- **OWL**
  - Para trabajar con modelos y OWL, Jena propone una API específica orientada a ontologías.
  - Jena igualmente propone una API para facilitar el razonamiento y comprobar el contenido de los archivos semánticos. Permite especificar distintos razonadores.

A continuación se van a definir algunos de los conceptos que maneja Jena así como se va a introducir brevemente el ámbito de la inferencia en este framework.

Modelo

A la hora de trabajar con triplas, jena utiliza los llamados **modelos**. Un modelo de Jena es un conjunto de triplas RDF con el que se puede trabajar de diversas formas: bien para hacer consultas sobre él o para aplicar inferencia. Estos modelos son la piedra angular de la tecnología puesto que son la fuente del resto de operaciones que se desarrollan en este framework.

Inferencia

La inferencia es un proceso abstracto de derivación de información adicional partiendo de los datos [JEN14]. De esta manera se utilizará la inferencia para hacer explícita información que aparece de forma implícita en los datos.

La inferencia en Jena consta de dos partes fundamentales:

- **Razonadores**: encargados de llevar a cabo la inferencia, pudiendo encontrar en Jena los siguientes [JEN14]:
  1. Razonador transitivo: ofrece soporte para el razonamiento a través de los conceptos de clase y propiedad. Únicamente afecta a propiedades transitivas y reflexivas de `rdfs:subPropertyOf` y `rdfs:subClassOf`.
  2. Razonador de reglas RDFS: Implementando un subconjunto configurable de vínculos RDFS.
  3. Razonadores para OWL, OWL Mini, OWL Micro: Un conjunto no completo para la implementación de inferencia basada en OWL/Lite y OWL/Full.

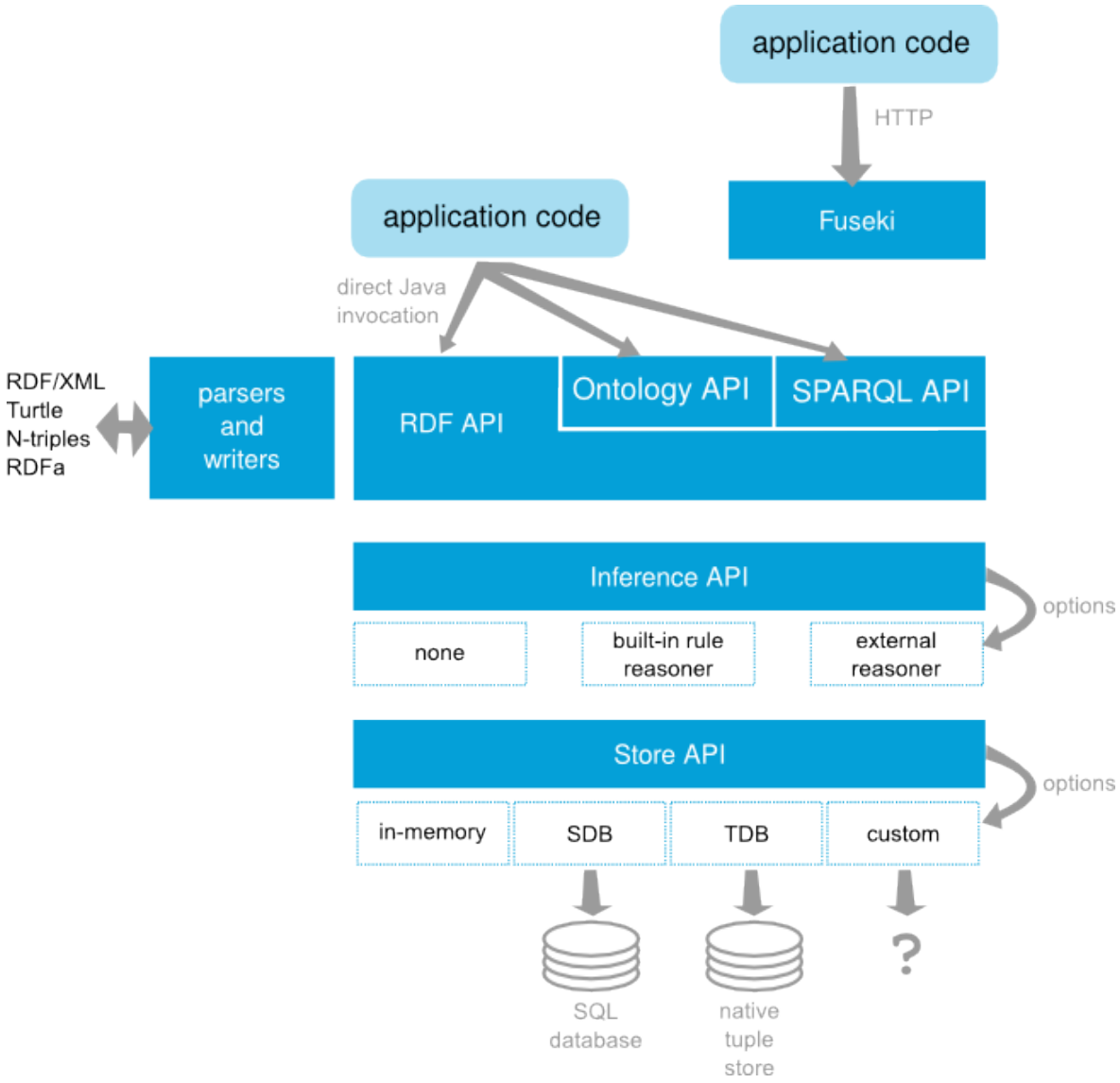

Figura 3.7: Diagrama de la arquitectura de Jena. Extraída de [JEN14]

4. Razonador de reglas genéricas: Se basa en reglas definidas, mediante el encadenamiento hacia adelante, hacia atrás e híbrido.

- **Reglas**: las reglas son sentencias usadas para aplicar inferencia en un determinado modelo. Jena implementa un lenguaje propio para la edición de estas reglas.

Se puede consultar un esquema del razonamiento en Jena en la figura 3.8.

Sesame

OpenRDF Sesame es un framework estándar *de-facto* para el procesado de datos RDF [OPE]. Sesame incluye operaciones para el procesamiento, consulta, almacenamiento e inferencia sobre RDF. Al igual que Jena, Sesame contiene un número considerable de herramientas para la construcción de aplicaciones de tecnologías semánticas.

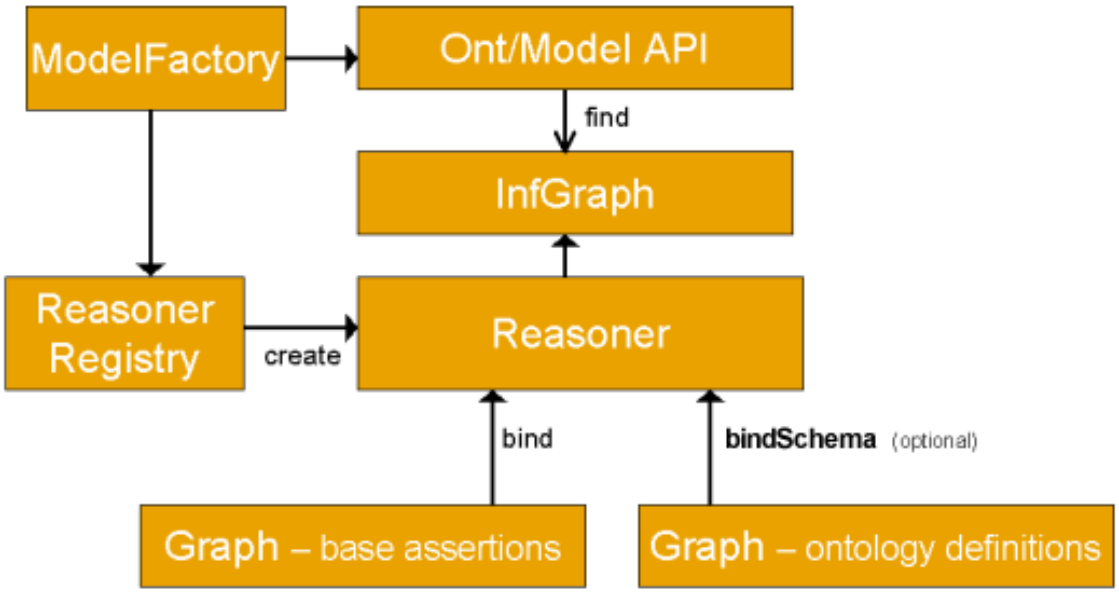

Figura 3.8: Razonamiento en Jena. Extraída de [JEN14]

CubicWeb

CubicWeb [CUB] es otro framework, *open source*, para la realización de aplicaciones semánticas. Está escrito en Python. Sus características engloban:

- Soporta OWL y RDF
- Relationship Query Language (RQL)
- Herramientas de migración para desarrollo ágil
- Una librería de *cubes* como pequeños módulos al estilo de Ruby.

Comparativa

En el cuadro 3.1 se pueden contemplar algunas características de los frameworks anotados anteriormente, pudiendo encontrarse una relación más completa en [W3Cc]:

Framework	Lenguaje	Licencia	Versión
Apache Jena	Java	Apache License 2.0	2.11.0 / September 18, 2013
Sesame	Java	BSD-style license	2.7.11 / March 27, 2014
CubicWeb	Python	Lesser General Public License	3.18.4 / April 7, 2014

Cuadro 3.1: Comparativa de frameworks de Web Semántica. Extraída de [W3Cc]

Datos Enlazados (LD)

En esta sección se incluyen los conceptos clave respecto de los principales movimientos de datos abiertos y datos enlazados.

Definición de LD

Berners-Lee en [BL09] enfatiza la publicación de los datos en la Web no únicamente como exposición de los mismos, sino mediante el establecimiento de enlaces de forma que personas o máquinas puedan explorar una Web de datos. De esta manera se pueden encontrar datos relacionados.

La mayor parte del contenido en la Web está construido con documentos. Estos documentos a su vez tienen enlaces hacia otros documentos cuyo contenido puede estar o no formalizado. El movimiento LD pretende dar un paso más allá, estableciendo relaciones entre los datos de manera global (véase figura 3.9). Para ello, RDF y las URI toman un papel crucial. [BL09] expone sus cuatro reglas para LD:

1. Usar URI como nombres para los conceptos.
2. Usar HTTP para que las personas puedan acceder a esos nombres.
3. Proporcionar información útil cuando la URI sea desreferenciada, usando estándares como RDF o SPARQL.
4. Incluir enlaces a otras URI, de manera que los usuarios puedan descubrir nuevos conceptos.

[BHBL09] establece a su vez una serie de pasos básicos para la publicación de LD:

1. Asignar URI a las entidades descritas por el conjunto de datos y proveer el desreferenciado de las URI a través del protocolo HTTP en representaciones de RDF.
2. Establecer enlaces RDF a otros recursos de datos en la Web, de tal manera que los usuarios puedan navegar a través de la Web de los datos siguiendo enlaces RDF.
3. Facilitar metadatos sobre los datos publicados, de manera que los usuarios puedan evaluar la calidad de los datos publicados y escoger entre diferentes medios de acceso.

Aprovechando LD la información en la Web Semántica puede verse desde diferentes niveles de granularidad, desde el grafo universal formado por todos los documentos RDF en la Web, pasando por documentos individuales hasta sus triplas [DFP+05] (véase figura 3.10).

Datos Abiertos (OD) y Datos Enlazados Abiertos (LOD)

Por otra parte los Datos Abiertos (OD), se presentan en ausencia de formatos privativos, información pública y reutilizables procedentes de organizaciones tales como las gubernamentales. Así, Datos Enlazados Abiertos (LOD) serán todos aquellos datos enlazados que

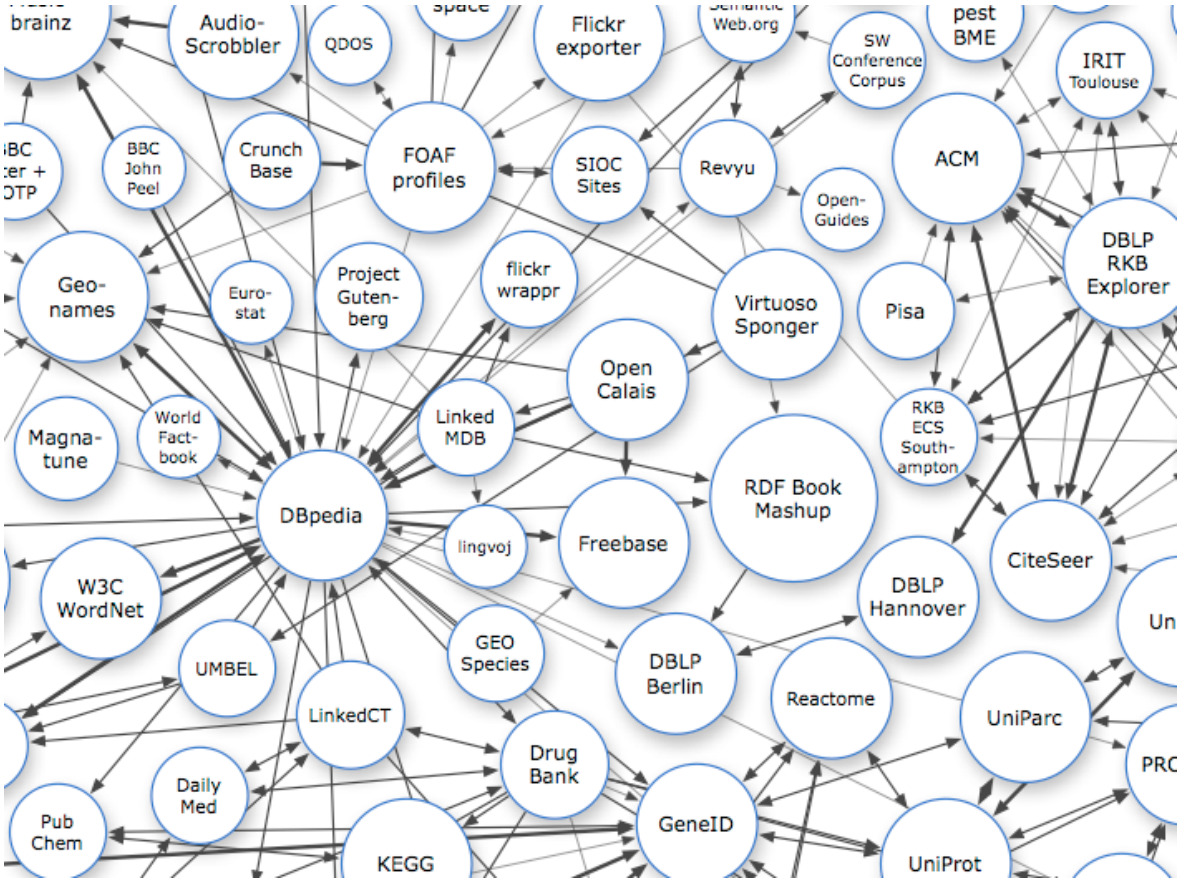

Figura 3.9: Esquema LD DBpedia

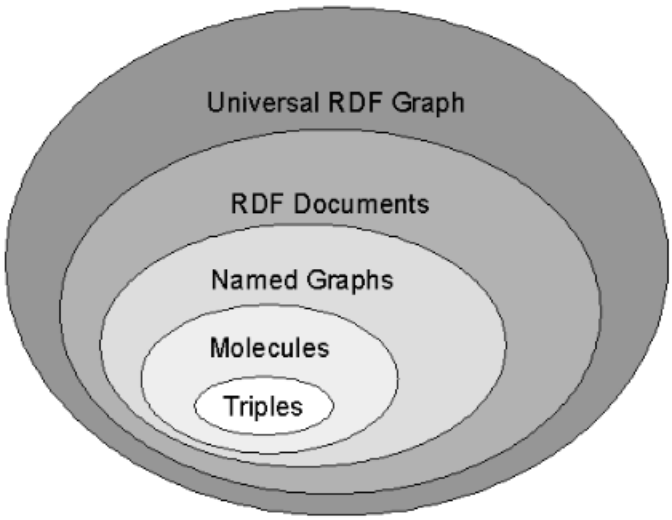

Figura 3.10: Niveles de granularidad en la Web Semántica. Extraído de [DFP+05]

sean abiertos. [BL09] define LOD como LD que es liberado bajo licencia abierta que no impida su reuso de manera gratuita.

Asimismo, define las cinco estrellas de calidad de LOD:

- ⋆ Los datos están disponibles en la Web, independientemente del formato, pero con licencia abierta, para que sean OD.
- ⋆⋆ Disponible para la lectura por parte de máquinas, es decir, datos estructurados. Por ejemplo, utilizar una tabla Excel en lugar de una imagen escaneada de dicha tabla.
- ⋆ ⋆ ⋆ Como el anterior, pero en formato no propietario. En este caso formato Comma Separated Values (CSV) en lugar de Excel.
- ⋆⋆⋆⋆ Todo lo anterior y además, usar estándares abiertos de W3C (RDF y SPARQL) para identificar conceptos, de manera que otros usuarios puedan apuntar a este contenido.
- ⋆ ⋆ ⋆ ⋆ ⋆ Lo anterior y además, enlazar estos datos a los de otros usuarios para proporcionar un contexto.

LOD en la actualidad

Empresas del sector privado y público se han dado cuenta de las ventajas que OD y LOD pueden ofrecerles. Es por ello que están surgiendo muchas aplicaciones tanto a nivel nacional, donde estos movimientos aún no están demasiado extendidos, como a nivel internacional. En [APP] se encuentra una lista con una serie de aplicaciones realizadas utilizando tecnología OD y LOD. A continuación se exponen algunos ejemplos.

Bus Guru

Bus Guru es una aplicación para iOS que monitoriza en tiempo real la situación de los autobuses urbanos así como su hora de llegada estimada para una parada (véase figura 3.12).

Bus Gijón

Igualmente en el ámbito nacional, existen aplicaciones que aprovechan los datos abiertos generados en tiempo real por dispositivos empotrados en autobuses y marquesinas. Un caso homólogo a **Bus Gurú** es **Bus Gijón** (figura 3.11), que recaba información del portal de OD `https://datos.gijon.es` para obtener información de los autobuses.

Calidad de Datos

En [SLW97] se identifican tres roles a la hora de tratar los datos:

1. *Data producers*: aquellas entidades encargadas de **generar** los datos.
2. *Data custodians*: personas responsables del **almacenamiento y procesado** de los datos.
3. *Data consumers*: personas o grupos que **usan** los datos.

Debido a que los diferentes roles tendrán concepciones distintas de lo que los datos significan para ellos según su papel en un sistema, pueden encontrarse distintas perspectivas de lo

Figura 3.11: Bus Gijón App

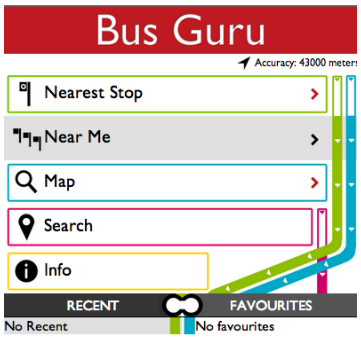

Figura 3.12: Bus Gurú App

que significa calidad de datos. Dos de las principales perspectivas son *Meeting Requirements* y *Fitness for Use*.

- *Meeting requirements* [BH12]: Los datos tienen calidad si satisfacen los requisitos que fueron establecidos. Los **requisitos de datos** que son especificados, por ejemplo, con un modelo Entidad-Interrelación (ER) en el que se subrayan cómo se van a relacionar los datos como conjunto, es decir, un marco arquitectónico para los datos. En [FH10] y [FH11] se pueden consultar aproximaciones a la calidad de datos desde este punto de vista.

- *Fitness for use* [SLW97]: Se dice que los datos tienen calidad si son válidos para el propósito por el que son requeridos, es decir, considerando la calidad de los datos en el **contexto** de uso. En [CVCP08] se adopta esta aproximación.

Dimensiones de Calidad de Datos

La perspectiva *Fitness for Use* se centra en los *Data consumers*, considerando datos de alta calidad aquéllos que son apropiados al usuario final en el contexto de uso. Debido a ello, se deben considerar aspectos tales como *utilidad* o *usabilidad* y en definitiva, todo aquel aspecto que pueda repercutir en la experiencia del usuario.

Puesto que la calidad de los datos dependerá del propósito, se requiere considerar una serie de **categorías**. La calidad de datos es un concepto que es necesario evaluar desde distintos criterios o **dimensiones** (Dimensión de Calidad de Datos (DQD)) [SLW97]. Al conjunto de dimensiones de calidad de datos utilizados para evaluar un conjunto de datos se le conoce como **Modelo de Calidad de Datos**. En el cuadro 3.2 se detallan conjuntos de dimensiones de calidad agrupadas por categorías.

Categoría de Calidad de Datos	Dimensiones de Calidad de Datos
Intrínseca	Precisión, Objetividad, Credibilidad, Reputación
Accesibilidad	Accesibilidad, Acceso seguro
Contextual	Relevancia, Valor añadido, Temporalidad, Completitud, Cantidad de datos adecuada
Representacional	Interpretabilidad, Facilidad de entendimiento, Representación concisa, Representación consistente

Cuadro 3.2: Categorías y dimensiones de DQ. Extraído de [SLW97]

Pese a que existan modelos de calidad de datos consistentes, la calidad de datos no deja de ser un concepto subjetivo, puesto que para un mismo conjunto de datos se pueden obtener niveles de calidad radicalmente distintos dependiendo del usuario o el rol que haga uso de ellos, incluso sobre la misma dimensión de calidad [Wan98].

A continuación se detallarán algunas dimensiones de calidad de datos.

Completeness

Dentro de todas las posibles dimensiones de calidad, existen subconjuntos que cobran especial relevancia. *Completeness* es, por norma general para todos los autores, una dimensión de obligada presencia en todos los trabajos.

Dentro de esta dimensión, en la bibliografía se proponen distintas métricas que contemplan el concepto de *completitud* desde diversas perspectivas. Como se comprobará en adelante (véase Sección **??**), en el presente trabajo se ha optado por una única visión, más compacta, sencilla y acorde a lo que un usuario puede esperar acerca de la presencia de valores no nulos en Linked Data.

Definición

[ZRM+13] define **Completeness** como el grado en el que toda la información requerida está presente en un conjunto de datos particular. En general, esta definición se puede extender sobre otros factores tales como profundidad de los datos, anchura y alcance para la tarea que se quiera realizar.

[PLW02] amplía esta definición. Define Completeness como el grado en que los datos no han desaparecido y poseen la suficiente amplitud y profundidad para la tarea en cuestión.

Por otro lado, la norma ISO 25012 (véase [ISO]) define Completeness como el grado en el que los datos asociados con una entidad tienen valores para todos los atributos esperados e instancias relacionadas en un contexto de uso específico.

Accesibility

En entornos de Web Semántica y Linked Data tiene especial importancia el hecho de que dichos datos sean de fácil acceso. La palabra *accesiblidad* cuando se habla de datos enlazados cobra un valor mayor que en otras dimensiones. Es preciso que los datos enlazados estén, en efecto, adecuadamente enlazados.

Definición

ISO 25012 (véase [ISO]) define *Accessibility* como el grado en el que los datos puedan ser accedidos en un contexto de uso específico, particularmente por gente que necesite tecnología de apoyo o una configuración especial debido a alguna discapacidad.

Por otra parte [PLW02] define Accessibility como el grado en que los datos están disponibles y son accesibles fácil y rápidamente.

Calidad de Datos en LD

Actualmente, existen autores tales como [ZRM+13] o [FH11] que comienzan a aplicar conceptos de LD para algunas dimensiones de calidad. Para ilustrarlo, se muestran algunas métricas propuestas en estos trabajos para las DQD Completeness y Accessibility.

Métricas para Completeness

[ZRM+13] propone una serie de métricas para la evaluación de esta dimensión.

1. *Schema Completeness*: grado en el que las clases y propiedades de una ontología están representadas. También conocido como *ontology completeness*.
2. *Property Completeness*: evaluación sobre los valores perdidos de una propiedad específica.
3. *Population Completeness*: siendo el porcentaje de todos los objetos del mundo real de un determinado tipo que están representados en los conjuntos de datos.

4. *Interlinking completeness*: refiriéndose al grado en el que las instancias en el conjunto de datos están interconectadas. Esta métrica consta de especial importancia en Linked Data. No obstante esta métrica concreta puede entenderse (como así se abordará en este trabajo, en la Sección **??**) como una particularización de otra dimensión de calidad bien distinta: *Accesibility*.

[FH10] ofrece un conjunto de consultas SPARQL que permiten identificar problemas de integración de calidad de datos. Se pueden comprobar dos ejemplos en los listados 3.2 y 3.3 extraídos de ese mismo trabajo.

```
SELECT ?s
WHERE { {
  ?s a <class1> .
  ?s <prop1> "". }
UNION{
  ?s a <class1> .
  NOT EXISTS {
  ?s prop1> ?value}}}
```

Listado 3.2: Consulta SPARQL para identificación de literales perdidos (I)

```
SELECT ?s
WHERE { {
      ?s a <class1> .
      ?s <prop1> <value1>.
   NOT EXISTS{
      ?s <prop2> ?value2 .
   }
   }UNION{
      ?s <prop1> <value1> .
      ?s <prop2> "".
   }}
```

Listado 3.3: Consulta SPARQL para identificación de literales perdidos (y II)

Métricas para Accessibility

[ZRM+13] considera Accessibility como una categoría que a su vez contiene cuatro dimensiones, cada una con sus métricas:

- **Disponibilidad**: grado en el la información está presente y preparada para su uso. Sus métricas propuestas son:
  - *Accessibility of the server*: comprobación sobre el servidor SPARQL ante una consulta.
  - *Accessibility of the* SPARQL *endpoint*: comprobación sobre el servidor SPARQL ante una consulta.

- *Accessibility of the* RDF *dumps*: comprobación sobre la recuperación de datos en un contenedor de datos RDF.
- *Dereferencability issues*: comprobación en el momento que una URI devuelva un error del código de respuesta o enlace roto.
- *No structured data available*: detección de enlace caído o cuando una URI sin metadatos RDF o sin redirección, devuelva el código de error pertinente.
- *Misreported content types*: detección de si el contenido es susceptible a ser consumido y si el contenido puede ser accedido.
- *No dereferenced back-links*: detección de todos los enlaces propios al conjunto de datos: localmente disponibles.

- **Rendimiento**: Eficiencia del sistema vinculado al conjunto de datos de manera que cuanto más eficiente sea la fuente de datos, un sistema puede procesar más eficientemente los datos. Las métricas para esta dimensión son:
  - *No usage of slash*-URIs: uso de URIs abreviadas cuando existen grandes cantidades de datos.
  - *Low latency*: si una petición HTTP es contestada en un tiempo medio de un segundo.
  - *High throughput*: número de peticiones HTTP contestadas por segundo.
  - *Scalability of a data source*: detección de si el tiempo en responder una cantidad de diez peticiones, dividido entre diez, no es mayor que el tiempo en responder una petición.
  - *No use of prolix* RDF *features*: detección del uso de primitivas RDF tales como contenedores y colecciones.
- **Seguridad**: Grado en el que los datos pueden ser restringidos y de esta manera protegidos contra alteraciones ilegales y uso no permitido. Las métricas propuestas son las siguientes:
  - *Access to data is secure*: uso para login de credenciales o uso de protocolos específicos.
  - *Data is of propietary nature*: el propietario de los datos permite el acceso únicamente a ciertos usuarios.
- **Tiempo de respuesta**: Medición del retardo, generalmente en segundos, entre el envío de una consulta por el usuario y la recepción de los resultados. Existe una métrica para esta dimensión:
  - *Delay in response time*: retardo entre el tiempo de envío de una petición por el usuario y la recepción de dicha petición por el sistema.

Big Data

Concepto

Frameworks

Capítulo 4

Método de Trabajo

Para el desarrollo del trabajo, el alumno deberá seguir algún proceso o metodología afín al problema que pretende resolver. Para ello, deberá aportar una pequeña descripción de los procesos o metodologías y justificar su adecuación al problema a resolver.

Se recomienda empezar, antes de hablar de métodos o técnicas específicos, por los aspectos de organización del trabajo (gestión del proyecto, planificación, esfuerzo, calendario, etcétera).

A modo de referencia, este capítulo tendrá una longitud no superior a 15 páginas.

Capítulo 5

Resultados

En esta sección se describirá la aplicación del método de trabajo presentado en el capítulo4 en este TFM concreto, mostrando los resultados (artefactos software y/o hardware, documentos de diversos tipos: modelos, diagramas, especificaciones, diseños, documentos, manuales, etc.) más importantes.

Este capítulo no tiene una extensión recomendada, sino que depende en mayor medida de los contenidos a presentar. No obstante, se recomienda que la longitud total del documento no supere las 120 páginas.

Capítulo 6

Conclusiones

Breve resumen de lo más destacable del trabajo con la solución propuesta. Análisis del logro del objetivo general y objetivos parciales propuestos. Concluir con posibles mejoras, ampliaciones o trabajos relacionados que quedan por hacer y que tienen interés para el tema tratado.

A modo de referencia, este capítulo tendrá una longitud aproximada de entre 2 y 10 páginas.

Conclusiones
Propuestas de trabajos futuros
Publicaciones
Opinión personal

ANEXOS

Anexo A

Ejemplo de anexo

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Referencias

[AH08] Dean Allemang y James Hendler. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2008.

[APP] Your City Needs These 7 Open Data Apps. url: http://mashable.com/2012/11/07/open-data-city-apps/.

[BH12] Peter Benson y Melissa Hildebrand. *Managing Blind: A Data Quality and Data Governance Vade Mecum*. ECCMA, Bethlehem (Pensylvania), 2012.

[BHBL09] Christian Bizer, Tom Heath, y Tim Berners-Lee. Linked Data - The Story So Far:. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009. url: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/jswis.2009081901.

[BL09] Tim Berners-Lee. Linked-data design issues. W3C design issue document, June 2009. http://www.w3.org/DesignIssue/LinkedData.html. url: http://www.w3.org/DesignIssues/LinkedData.html.

[BLHL01] Tim Berners-Lee, James Hendler, y Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, Mayo 2001. url: http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21.

[CUB] CubicWeb Semantic Web Framework. url: http://www.cubicweb.org/.

[CVCP08] Ismael Caballero, Eugenio Verbo, Coral Calero, y Mario Piattini. DQRDFS - Towards a Semantic Web Enhanced with Data Quality. En José Cordeiro, Joaquim Filipe, y Slimane Hammoudi, editors, *WEBIST (1)*, páginas 178–183. INSTICC Press, 2008. url: http://dblp.uni-trier.de/db/conf/webist/webist2008-1.html#CaballeroVCP08.

[DFP+05] Li Ding, Tim Finin, Yun Peng, Paulo Pinheiro Da Silva, y Deborah L. McGuinness. Tracking rdf graph provenance using rdf molecules. En *Proc. of the 4th International Semantic Web Conference (Poster)*, página 42, 2005. url: ftp://www.ksl.stanford.edu/local/pub/KSL_Reports/KSL-05-06.pdf.

[DMVH+00] Stefan Decker, Sergey Melnik, Frank Van Harmelen, Dieter Fensel, Michel Klein, Jeen Broekstra, Michael Erdmann, y Ian Horrocks. The semantic web: The roles of XML and RDF. *Internet Computing, IEEE*, 4(5):63–73, 2000. url: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=877487.

[FH10] Christian Fürber y Martin Hepp. Using Semantic Web Resources for Data Quality Management. En Philipp Cimiano y Helena Sofia Pinto, editors, *EKAW*, volume 6317 of *Lecture Notes in Computer Science*, páginas 211–225. Springer, 2010. url: http://dblp.uni-trier.de/db/conf/ekaw/ekaw2010.html#FurberH10.

[FH11] Christian Fürber y Martin Hepp. Towards a Vocabulary for Data Quality Management in Semantic Web Architectures. En *Proceedings of the 1st International Workshop on Linked Web Data Management*, LWDM '11, páginas 1–8, New York, NY, USA, 2011. ACM. url: http://doi.acm.org/10.1145/1966901.1966903.

[FOA] The Friend of a Friend (FOAF) project | FOAF project. url: http://www.foaf-project.org/.

[Gru] Tom Gruber. Ontology (Computer Science) - definition in Encyclopedia of Database Systems. url: http://tomgruber.org/writing/ontology-definition-2007.htm.

[Gua98] Nicola Guarino. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*, volume 46. IOS press, 1998.

[HBLM02] James Hendler, Tim Berners-Le, y Eric Miller. Integrating Applications on the Semantic Web, 2002. url: http://www.w3.org/2002/07/swint.

[HT06] Harry Halpin y Henry S. Thompson. One document to bind them: combining XML, web services, and the semantic web. página 679. ACM Press, 2006. url: http://portal.acm.org/citation.cfm?doid=1135777.1135877.

[ISO] ISO25012. ISO/IEC 25012 x ISO/IEC 25012 Software-Engineering - Qualitätskriterien und Bewertung von Softwareprodukten (SQuaRE) - Modell der Datenqualität. Technical report.

[JEN14] Apache Jena - Home, 2014. url: http://jena.apache.org/.

[Mar08] J. Martínez de Sousa. *Ortografía y ortotipografía del español actual*. Trea, 2008.

[NMo01] Natalya F. Noy, Deborah L. McGuinness, y others. *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, 2001. url: http://liris.cnrs.fr/~amille/enseignements/Ecole_Centrale/What%20is%20an%20ontology%20and%20why%20we%20need%20it.htm.

[OPE] openRDF.org: Home. url: http://www.openrdf.org/.

[OWL] OWL - Semantic Web Standards. url: http://www.w3.org/OWL/.

[PLW02] Leo L. Pipino, Yang W. Lee, y Richard Y. Wang. Data quality assessment. *Communications of the ACM*, 45(4):211–218, 2002. url: http://dl.acm.org/citation.cfm?id=506010.

[RDF] RDF - Semantic Web Standards. url: http://www.w3.org/RDF/.

[San11] Juan Antonio Pastor Sanchez. Tecnologias de Web Semantica, 2011.

[SBF98] Rudi Studer, V.Richard Benjamins, y Dieter Fensel. Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1-2):161–197, Marzo 1998. url: http://linkinghub.elsevier.com/retrieve/pii/S0169023X97000566.

[SBLH06] Nigel Shadbolt, Tim Berners-Lee, y Wendy Hall. The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3):96–101, Mayo 2006. url: http://dx.doi.org/10.1109/MIS.2006.62.

[SLW97] Diane M. Strong, Yang W. Lee, y Richard Y. Wang. Data Quality in Context. *Commun. ACM*, 40(5):103–110, Mayo 1997. url: http://doi.acm.org/10.1145/253769.253804.

[SPA] SPARQL Query Language for RDF. url: http://www.w3.org/TR/rdf-sparql-query/.

[W3Ca] Semantic Web - W3C. url: http://www.w3.es/Divulgacion/GuiasBreves/WebSemantica.

[W3Cb] Semantic Web - W3C. url: http://www.w3.org/standards/semanticweb/.

[W3Cc] Tools - Semantic Web Standards. url: http://www.w3.org/2001/sw/wiki/Tools.

[Wan98] Richard Y. Wang. A Product Perspective on Total Data Quality Management. *Commun. ACM*, 41(2):58–65, Febrero 1998. url: http://doi.acm.org/10.1145/269012.269022.

[XML] Extensible Markup Language (XML). url: http://www.w3.org/XML/.

[ZRM+13] Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, Soren Auer, y Pascal Hitzler. Quality assessment methodologies for linked open data. *Submitted to Semantic Web Journal*, 2013. url: http://semantic-web-journal.net/system/files/swj414.pdf.

Este documento fue editado y tipografiado con LaTeX empleando
la clase **esi-tfm** (versión 0.20170902) que se puede encontrar en:
`https://bitbucket.org/arco_group/esi-tfg`